

Note technique

Schéma Stratégique du Numérique

OSU-Réunion

Référence :

Auteurs : G. Payen, F. Gabarrot, V. K/Bidi, G. Desprairies, E. Cordier

Version du document : a4

Première édition : 10/12/2019

Dernière révision : 08/06/2020

Nombre de pages : 32

Historique des évolutions

10/12/2019	Création.
17/02/2020	Version va0 stable, présentation directeur OSU-Réunion
26/02/2020	Version va1 : corrections mineures
24/04/2020	Version va2 : modifications plan, présentation CS OSU-Réunion
05/06/2020	Version va3 : version pour présentation conseil OSU-R
08/06/2020	Version va4 : corrections mineures

Abréviations

API	Interface de programmation d'application
DMP	Data Management Plan
SSN	Schéma Stratégique du Numérique
DOI	Digital Object Identifier

Table des matières

Table des matières

Table des matières	4
1. Documentation	6
1.1. Documents de référence	6
1.2. Documents associés	6
2. Introduction	6
2.1. Le SSN	6
2.2. L'OSU-Réunion	6
2.3. Rôle de l'équipe informatique de l'UMS3365	7
2.4. Les principes FAIR et le SSN	7
2.5. Les différents acteurs de la gestion de la donnée	9
3. Gestion de la donnée	10
3.1. Données et métadonnées	12
3.2. Collecte de la donnée	12
3.2.1. Tramontane : serveur FTP	12
3.2.2. Eobs : gestionnaire des métadonnées	13
3.2.3. L'intégration des plans de gestion de la donnée	13
3.3. Stockage de la donnée	14
3.3.1. Zarlou, la base de données de l'OSU-Réunion	14
3.3.2. Classement des données	14
3.4. Accès à la donnée	15
3.4.1. Les différents accès possibles	15
3.4.2. GeOsuR-catalogue : service de catalogage de la donnée	15
3.4.3. GeOsuR-Data : diffusion de la donnée en « open-data »	16
3.2.4. Gestion des DOI (Digital Object Identifier)	17
3.5. Flux et traitements de la donnée	17
3.5.1. Flux de données	17
3.5.2. Airflow : service de gestion des flux	17
3.5.3. Licorne : traitements de la donnée	18
3.6. Conclusion sur la gestion de la donnée	Erreur ! Signet non défini.
4. Infrastructure	19
4.1. L'infrastructure de gestion de la donnée	19
4.1.1. Community : serveurs et espace disque	20

4.1.2.	Infrastructure des sites distants.....	22
4.2.	Outils de gestion de l'infrastructure.....	22
4.2.1.	Identity : gestion centralisée des accès aux services	23
4.2.2.	Hotei : supervision de l'infrastructure.....	23
4.2.3.	Ansible : automatisation des déploiements dans l'infrastructure.....	23
4.2.4.	Sauvegarde de l'infrastructure.....	24
4.2.5.	Gestion de parc informatique	24
5.	Supports informatiques aux unités de l'osu-r.....	25
5.1.	Support à l'instrumentation.....	25
5.1.1.	Support à l'installation des instruments (site de l'OSU-R et hors site).....	25
5.1.2.	Accès et surveillance du fonctionnement des instruments	25
5.1.3.	Filosur : gestionnaire de documentation	26
5.1.4.	Maintenance PC acquisition	26
5.2.	Aide au développement.....	26
5.2.1.	Algorithmie & traitements & opérationnalisation.....	26
5.2.2.	GitLab : pérennisation des codes sources et qualité logicielle	26
5.3.	Environnements virtuels de recherche	27
5.4.	Services complémentaires proposés par l'université de La Réunion.....	27
5.4.1.	Centre de calcul	27
5.4.2.	Salle de visioconférence	27
5.4.3.	Informatique de proximité.....	28
5.5.	Formations des utilisateurs.....	28
6.	Évaluation des moyens	28
6.1.	D'hier à aujourd'hui, évolution des moyens humains	28
6.2.	Stratégie à 5 ans	29
6.3.	Moyens humains et financiers.....	30
7.	Conclusion.....	31

1. DOCUMENTATION

1.1. Documents de référence

Code	Informations
DR01	Référence : osur-documentation/ctai/service/nt_ums3365_Conventions-standards-pour-stockage-diffusion-donnees-OSU-R_vb0.odt Classement : osur-documentation\ctai\service\G - Gestion des données Titre : Conventions d'organisation et standards pour le stockage et la diffusion des données de l'OSU-R Auteurs : Franck Gabarrot Date : 20/09/2019
DR02	Référence : note_OSU_SchemaStrategiqueNumerique Classement : osur-documentation\ctai\gestion\fonctionnement\schema_strategique_numerique\2020 Titre : Schéma stratégique du numérique des OSU Auteurs : Date :
DR03	Référence : nt_ums3365_Catalogue-services-info-a2 Classement : osur-documentation\ctai\service Titre : Catalogue de services proposés par le service informatique de l'OSU-Réunion Auteurs : Remy Decoupes Date : 09/10/2018

1.2. Documents associés

Code	Référence	Titre	Auteurs	Date
DAxxx				

2. INTRODUCTION

2.1. Le SSN

Afin de clarifier le rôle des différents acteurs dans l'écosystème local et de déterminer le niveau et le lieu d'affectation des moyens, il a été demandé par l'INSU de définir un schéma stratégique du numérique (SSN) à l'échelle de chaque OSU (DR02). Cette note rappelle les missions des OSU et décrit comment l'OSU répond à ces missions.

2.2. L'OSU-Réunion

L'OSU-Réunion, créé en 2010, couvre l'ensemble des disciplines relevant des sciences de l'univers au sens large. Il regroupe, à l'Université de la Réunion, des unités de recherche, des laboratoires ou équipes associés, et des stations d'observation, l'ensemble représentant environ 80 personnes (permanents ou temporaires).

Les tutelles principales de l'OSU-Réunion sont l'Université de la Réunion, l'INSU-CNRS et Météo-France. Sont membres du Conseil de l'OSU-R les représentants des organismes suivants : IPGP, IRD, Région Réunion, TAAF, D2RT, CIRAD, BRGM et Ahence NEXA. Le Conseil de l'OSU est

présidé par une personnalité extérieure choisie pour ses compétences. Le Comité Scientifique a une composition permettant de couvrir tous les domaines de spécialité de l'OSU, avec des représentants de tous les laboratoires et stations d'observation, les directeurs des unités de recherche membres de l'OSU ainsi qu'au moins quatre personnalités extérieures cooptées et choisies pour leurs compétences.

Outre l'Unité Mixte de Service UMS-3365 (CNRS-INSU, Université de La Réunion et Météo-France) qui fait partie intégrante de l'OSU-Réunion, les unités mixtes de recherche (UMR) de l'Université de La Réunion et les laboratoires rattachés à l'OSU sont :

- Le laboratoire ENTROPIE à La Réunion (Ecologie Marine Tropicale des Océans Pacifique et Indien, UMR 250, Université de La Réunion, IRD, CNRS-INEE)
- Le laboratoire Espace-Dev à La Réunion (IRD / Université de la Réunion / Université Montpellier 2 / Université des Antilles / Université de la Guyane).
- Le laboratoire LACy (Laboratoire de l'Atmosphère et des Cyclones, UMR 8105, Université de la Réunion, CNRS-INSU, Météo France)
- Une équipe du laboratoire universitaire LE2P (Laboratoire d'Energétique, d'Electronique, et Procédés, EA 4079, Université de La Réunion)
- Le laboratoire LGSR (Laboratoire Géosciences Réunion, représentation à l'Université de La Réunion de l'UMR 7154, Institut de Physique du Globe de Paris)
- Une équipe du laboratoire universitaire LIM (Informatique et Mathématiques, EA 2525, Université de La Réunion)
- Une équipe du laboratoire universitaire PIMENT (Laboratoire de Physique et Ingénierie Mathématique pour l'Energie et l'environnement, EA Université de La Réunion)
- Une équipe du Laboratoire PVBMT (Peuplement Végétaux et Bioagresseurs en Milieux Tropical, UMR CIRAD - Université de la Réunion)

Les stations d'observation rattachées à l'OSU sont :

- L'Observatoire de Physique de l'Atmosphère de la Réunion (OPAR)
- La station forestière (STAFOR)
- La station côtière et marine (STACOT)
- La station hydrologique (ERORUN)

L'Observatoire Volcanologique du Piton de la Fournaise (OVPF) de l'IPGP (Institut de Physique du Globe de Paris) et la station de télédétection spatiale SEAS-OI (associée à la représentation réunionnaise de l'UMR Espace-Dev) sont associés scientifiquement à l'OSU-Réunion.

2.3. Rôle de l'équipe informatique de l'UMS3365

L'unité mixte de service (UMS3365) est composée de 19 personnes (15 agents permanents et 4 agents en CDD). L'équipe informatique est représentée par 5 personnels représentant 4,5 ETP : 3 permanents (1 IGR, 2 IGE) et 2 CDD (2 IGR). Elle est divisée en 3 pôles métier : le pôle « système réseau », le pôle « développement » et le pôle « calcul scientifique ».

2.4. Les principes FAIR et le SSN

Notre stratégie de gestion de la donnée est basée sur les principes FAIR : Findable, Accessible, Interoperable, Reusable (<https://www.go-fair.org/fair-principles>).

Le tableau ci-après relie les principes FAIR avec les outils et les services pour la donnée qui sont

développés et exploités par le service informatique de l'OSU-R.

<p>Findable (trouvable)</p>	<p>Un humain ou une machine doit pouvoir trouver nos données. Des métadonnées machine-exploitable (donc standardisées) sont essentielles pour une découverte automatisable des jeux de données et des services. Le service Geosur Catalogue assure ce premier principe de « fairisation » pour tous nos jeux de données à vocation à être diffusés.</p> <p>Les experts des observatoires pourront accéder à un niveau plus fin de recherche grâce au service Eobs qui va leur permettre notamment d'avoir une recherche par date et par produits disponibles.</p>
<p>Accessible (accessible)</p>	<p>Une fois que les utilisateurs ont trouvé la donnée, ils ont besoin de savoir comment y accéder (incluant possiblement des aspects d'authentification). Encore une fois Geosur Catalogue assure ce principe de « fairisation » en associant à chaque fiche de métadonnées : les liens de téléchargement (liens vers nos services de diffusion et/ou ceux des centres de données extérieurs) et les contacts pour l'accès (demande de compte, etc.) et l'exploitation de la donnée (responsable scientifique, etc). Les conditions d'accès doivent être clairement établies (licence, droits de réutilisation, etc.). Même quand les données sont inaccessibles, les métadonnées demeurent accessibles afin que le jeu de données reste découvrable. L'auteur du jeu de données pourra alors être contacté pour une demande d'information, ou une demande d'accès aux données.</p> <p>Eobs fournit également pour les experts des observatoires un lien de téléchargement HTTP simple des fichiers de données ainsi qu'un affichage des fichiers graphiques au travers de son interface web afin de faciliter le contrôle qualité.</p>
<p>Interoperable (Interopérable)</p>	<p>L'interopérabilité rend possible l'intégration de différents jeux de données sur un portail unique, par exemple au sein de la même application web. Concrètement il s'agit pour nous de fournir des données dans un format standard interopérable (ex. NetCDF) et avec des métadonnées qui suivent une convention standard du domaine (ex. Climate Forecast). Le premier axe est donc de favoriser la standardisation des données, c'est l'objectif de notre service de transformation et de traitement de la donnée Licorne (cf 3.5.3). Le second axe est l'exploitation de services permettant un accès interopérable à la donnée suivant différents protocoles et formats, notamment NetCDF et OpenDap : c'est l'objectif du service Geosur Data (cf 3.4.3).</p>
<p>Reusable (Réutilisable)</p>	<p>C'est l'objectif ultime des principes FAIR que celui d'optimiser la réutilisation de la donnée. Il s'agit d'optimiser les étapes précédentes et d'assurer à l'utilisateur final, en plus de la capacité à accéder à la donnée, la capacité à réutiliser la donnée : pouvoir lire le(s) fichier(s), description standardisée ET riche, détails sur la provenance et la qualité. Le service de transformation et de traitement de la donnée Licorne suit cette démarche en complément du service Geosur Catalogue et du service Eobs pour le suivi de production et le suivi instrument.</p>

2.5. Les différents acteurs de la gestion de la donnée

L'infrastructure de recherche Data Terra (www.data-terra.org) a pour but de « développer un dispositif global d'accès à des données, produits et services permettant d'observer, comprendre et prévoir de manière intégrée l'histoire, le fonctionnement et l'évolution du système Terre soumis aux changements globaux ». Pour faire cela, elle s'appuie sur 4 pôles de données : AERIS pour les données atmosphériques, FORMATER pour les données de terre solide, THEIA pour les données de surface continentale et enfin ODATIS pour les données océaniques et littorales. Le rôle de ces pôles est de centraliser les informations sur les données et de les mettre à disposition à travers des portails de diffusion de la donnée. Ces pôles peuvent également contenir des centres d'expertises. Chaque pôle s'appuie sur des CDS (Centre de Données et Services), qui ont pour rôle la gestion de la donnée.

Tout ce système est récent, et chaque pôle et chaque CDS a sa propre façon de fonctionner. Prenons trois exemples concrets. Pour les données issues du SNO NDACC, qui est géré par le pôle ACTRIS, il est demandé à ce que les données soient simplement nettoyées du bruit en local, puis envoyées directement aux différents pôles sous 30 jours, où les données seront ensuite traitées et diffusées. Pour les données issues de l'IR OZCAR, qui fait partie du pôle THEIA, le traitement doit être réalisé complètement en local et les fichiers de sorties doivent être envoyés selon un format spécifique. Enfin, pour les données du SNO Dynalit, qui fait partie de l'IR Ilico, qui fait partie du pôle THEIA, le traitement complet doit être réalisé en local et hébergé en local aussi : les données et métadonnées doivent être disponibles à travers un portail interopérable.

L'OSU-Réunion doit répondre aux demandes des différents pôles. Il a donc fallu construire un système qui soit suffisamment souple et robuste, afin de répondre aux demandes d'aujourd'hui et de commencer à répondre aux demandent de demain, le tout en étant interconnectable avec les différents systèmes : du système d'acquisition jusqu'à la diffusion dans les portails de données gérés par les pôles.

Les missions de l'OSU-Réunion sont donc la collecte des données (Figure 1), que ce soit à partir d'instruments automatiques ou bien de collectes manuelles, la sauvegarde de ces données en local, le traitement de ces données afin de passer d'une donnée physique à une donnée géophysique, et enfin, l'accès à ces données, que ce soit par le PI (Principal Investigator, c.a.d le responsable scientifique) qui souhaite valider ces données, mais aussi par les CDS qui valorisent ces données par thématique.

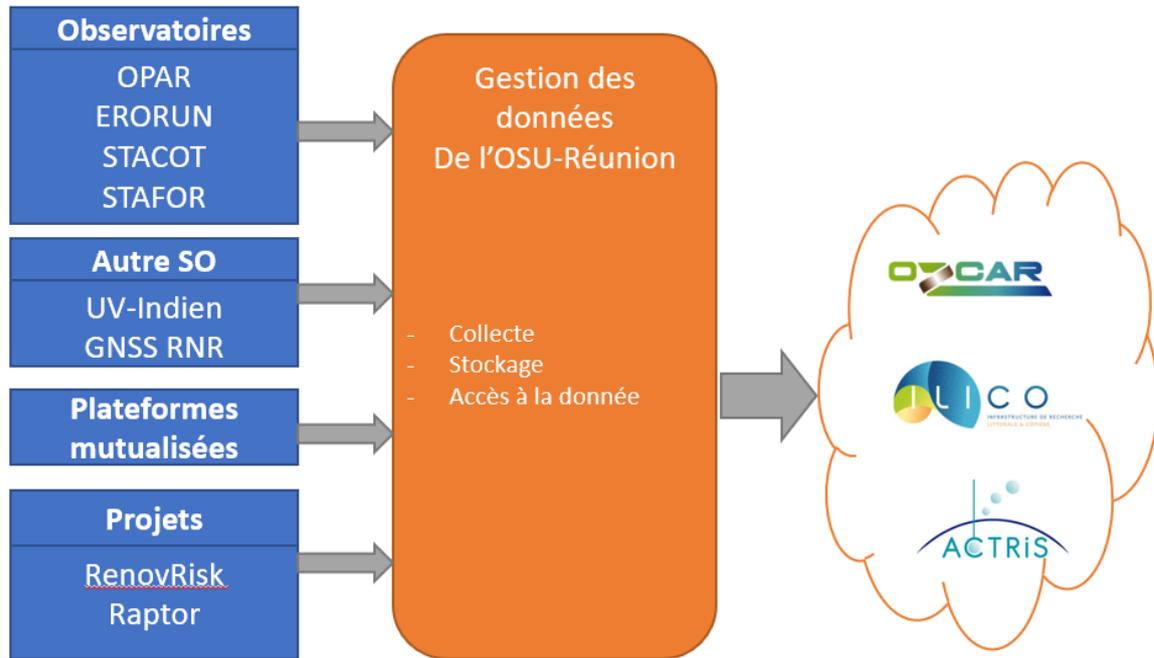


Figure 1: rôle de l'OSU-Réunion dans la gestion des données

2.6. Schéma présentant la gestion de la donnée par l'OSU-Réunion

Toute la gestion de la donnée par l'OSU-Réunion peut se résumer par la figure 2. D'un côté, il y a les producteurs de la donnée. Cette donnée rentre dans notre gestion de la donnée par deux moyens : soit par FTP pour la donnée en elle-même, soit par notre interface web Eobs pour les métadonnées. La donnée sera stockée et sauvegardée à travers notre service de stockage Zarlou. Elle pourra être classée à travers notre système de gestion de flux de données Airflow, et traitée par notre ferme de serveur et notre librairie de traitement Licorne. Enfin, cette donnée pourra être récupérée soit par FTP par des utilisateurs avertis, soit par notre interface web GeOsur avec une présentation détaillée du jeu de données. Toute cette description correspond à la partie jaune de la figure 2 et correspond à la partie 3 de ce rapport. Pour réussir cette gestion des données, nous avons mis en place différents services, permettant de nous aider à gérer l'infrastructure autour de la donnée (en rose, Fig. 2, partie 4 de ce rapport) et des services que l'on propose aux différents chercheurs (en bleu ciel, Fig. 2, partie 5 de ce rapport).

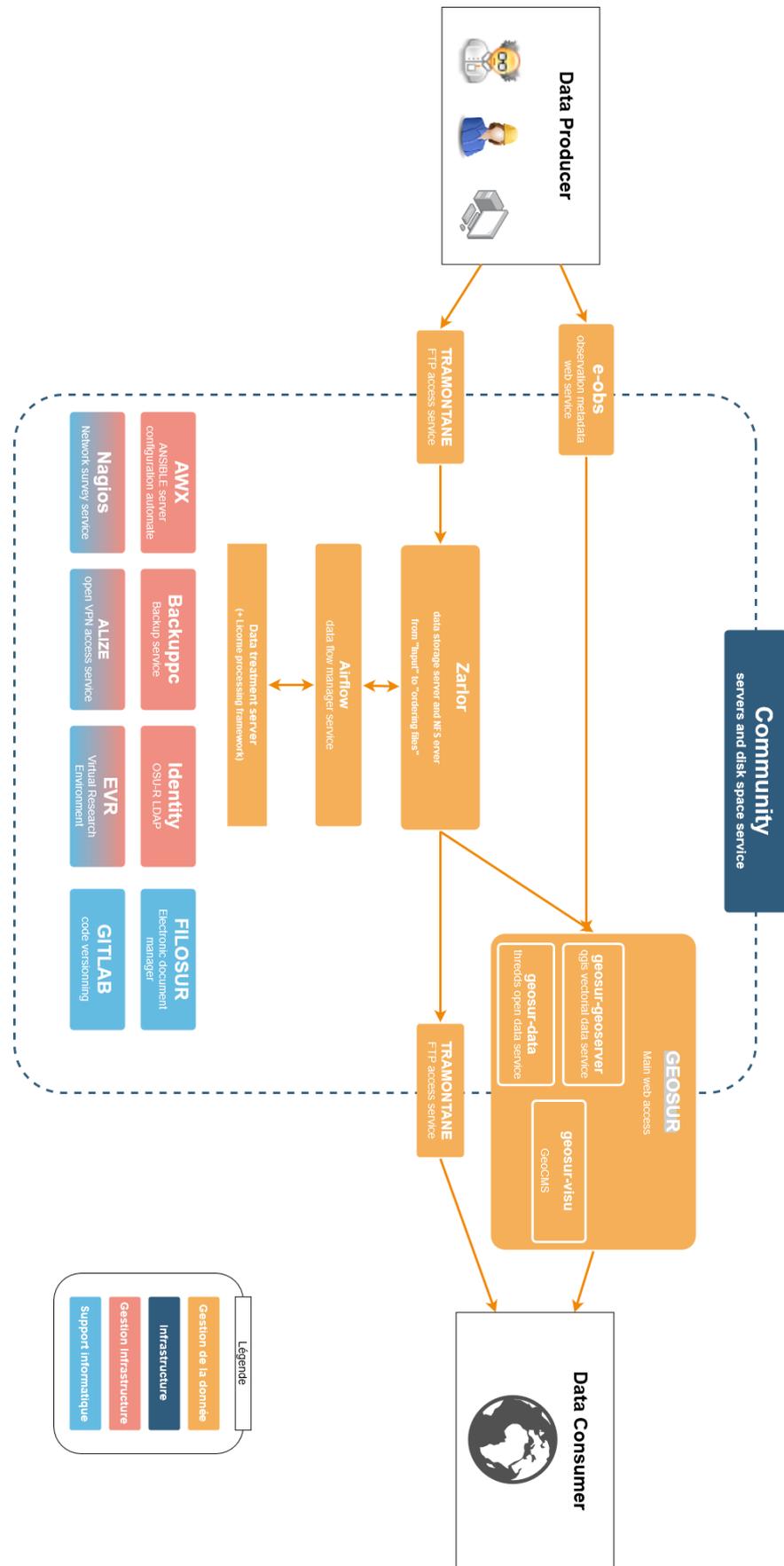


Figure 2: Schéma de gestion de la donnée au niveau du service informatique de l'OSU-Réunion

3. GESTION DE LA DONNÉE

3.1. Données et métadonnées

Quand on parle de « gestion de la donnée », on parle de la donnée au sens général, c'est-à-dire la donnée physique ou géophysique et la métadonnée associée. Cette dernière correspond à tout ce qui décrit le contour de la donnée. Les deux ne peuvent être dissociées, car savoir ce que nous avons produit (métadonnées d'opération et de production), dans quel contexte (métadonnées de découverte et de suivi du site d'observation, de l'instrument, du logiciel, etc.) et avec quel niveau de qualité (métadonnées de contrôle qualité) est crucial pour la pérennité de la donnée, son interprétation et sa réutilisation.

Cependant, la donnée et la métadonnée ne sont pas toujours enregistrées de la même façon. Dans la plupart des cas que nous avons, la donnée est enregistrée dans un fichier. En général, une partie de la métadonnée se trouve également dans ce fichier, que ce soit dans l'entête, ou bien dans le nom du fichier. Cependant, une partie de la métadonnée peut être enregistrée de façon différente, dans une base de données dissociée de la donnée elle-même. Cette solution a été mise en place, afin de distinguer la métadonnée associée au programme qui crée la donnée (le système d'acquisition, le programme de traitement), et la métadonnée qui est ajoutée en plus et qui n'est pas gérée par ces programmes.

3.2. Collecte de la donnée

Il y a deux façons de collecter la donnée pour la rentrer dans notre système d'information : soit par FTP pour les données physique ou géophysique, soit par un serveur de gestion de métadonnée pour la métadonnée.

3.2.1. Tramontane : serveur FTP

Service de collecte et de diffusion par FTP

Nom : Tramontane FTP

Gestionnaire : OSU-R – service informatique

Site : Moufia

Système : Serveur Debian installé sur community hébergeant un serveur VsFTP;

Description : Le Protocole FTP est le protocole de base de partage de donnée de l'OSU-R. Le serveur Tramontane permet d'une part de récupérer automatiquement les données envoyées par les instruments, et permet d'autre part aux personnels internes, mais aussi aux collaborateurs externes et centre de données, de pouvoir accéder aux données des observatoires dès leur enregistrement sur notre serveur de stockage (accès proche du temps réel).

Stratégie d'exploitation : -

Axes d'évolution : Le serveur Tramontane propose uniquement l'accès et l'écriture des données via le protocole FTP qui bien qu'il soit robuste et éprouvé ne répond plus désormais aux nouveaux standards de sécurité. Il est donc prévu de migrer le serveur sur des protocoles plus sécurisés comme le SFTP ou le FTPS. La gestion des utilisateurs et des accès se fait pour l'instant de manière locale, il est prévu pour faciliter la gestion de relier tramontane à notre service d'authentification centralisée Identity (cf 4.2.1).

3.2.2. Eobs : gestionnaire des métadonnées

Service de gestion des métadonnées des observatoires

Nom : Eobs

Gestionnaire : OSU-R – service informatique

Site : Moufia

Système : Serveur Debian avec Python 3, Django REST Framework pour la partie Back-end, npm et Vue.js pour le Front-end et PostgreSQL, PostGIS et Elasticsearch pour la partie base de données.

Description : Eobs est une application web à destination des observatoires afin de permettre la saisie, la conservation, la consultation et l'extraction des métadonnées de production et de suivi instrumental. Eobs est prioritairement un outil de saisie et de conservation de l'information observatoire pour l'équipe observatoire et pour le suivi long terme des mesures.

Eobs est relié à notre service d'authentification centralisée Identity (cf 4.2.1) permettant à tout utilisateur de l'OSU-R de pouvoir se connecter et un module d'ajout d'utilisateurs externes est aussi proposé. Les rôles sont ensuite définis par l'administrateur. Chaque utilisateur a, en fonction de son rôle par rapport à une ressource donnée, les autorisations d'accès aux interfaces et aux fichiers de données. De plus, il aura une vue adaptée à son rôle dans son tableau de bord.

L'outil permet aux PI de définir en profondeur leurs instruments, leurs logiciels, ... de limiter les accès aux collections de données, et d'attribuer des fiches de contrôle de qualité sur ces dernières. Ils peuvent également créer des événements et les attribuer aux opérateurs. Les opérateurs de leur côté saisissent des fiches d'opérations où ils définissent les actions qu'ils accomplissent.

Enfin, Eobs offre des capacités complémentaires pour le contrôle qualité : téléchargement des fichiers, visualisation des quick-looks, gestion de la validation, statistiques d'exploitation, export des métadonnées vers les centres de données nationaux ou vers d'autres outils.

Stratégie d'exploitation : Le service Eobs est installable sur n'importe quelle machine ou serveur. Ce qui permet à d'autres observatoires de l'installer chez eux. Les codes sont accessibles sur notre service Gitlab (cf 5.2.2) avec une procédure d'installation. Eobs permet aux PI de définir les métadonnées des instruments. Cependant, dans un premier temps l'équipe informatique va accompagner ces derniers pour définir ensemble les modèles de saisie pour la découverte de ces instruments pour, en parallèle, définir au mieux un vocabulaire contrôlé en s'appuyant sur les standards des communautés.

Axes d'évolution : Une première version est actuellement en production et la seconde est en cours de développement. Cette dernière a été profondément revue et se veut plus flexible afin d'être exploitable par tous les observatoires.

Eobs a pour ambition d'être l'outil central de l'OSUR, où les autres applications comme Geosur (cf 3.4.2) pourront inter communiquer avec l'interface de programmation (API) de Eobs. Pour ce faire, nous sommes partis sur une architecture REST (*representational state transfer*) pour la nouvelle version permettant ainsi à toute application tierce de communiquer avec l'API via des requêtes HTTP. Ceci, pourra nous permettre par la suite de développer des modules spécifiques en lien avec Eobs (par exemple, des ordinateurs de contrôle devant les LIDARs, des applications mobiles, ...).

5.2.3. L'intégration des plans de gestion de la donnée

Nous travaillons sur l'utilisation de plans de gestion de la donnée (Data Management Plan, DMP) depuis plusieurs années afin, en premier lieu, d'améliorer notre préparation d'accueil de nouveaux instruments et la conservation des métadonnées de contexte. Après avoir suivi la formation de l'INIST du CNRS, nous avons demandé, pour chaque filière instrumentale, l'élaboration par le PI d'un DMP selon un modèle mis à disposition à cette adresse <https://opar.univ-reunion.fr/instrumentinstall>, avec une première version à remplir avant installation afin de permettre à l'équipe d'accueil de préparer

l'installation et de prévoir les ressources nécessaires. L'intérêt est indéniable mais l'effort demandé n'est pas toujours compris et ce document n'est pas exploité ou mis en avant ensuite, et donc il n'est pas mis à jour. S'agissant de métadonnées de découverte, nous avons pris le parti d'inclure la saisie des informations contenues dans le DMP dans les évolutions du logiciel Eobs afin de mieux les conserver et de mieux les mettre en valeur en permettant de les re-exploiter. Cette stratégie devrait nous permettre de proposer à moyen terme un meilleur accompagnement sur cet aspect incontournable de tous les projets incluant la création de jeux des données d'observation.

3.3. Stockage de la donnée

Les données d'observation ont un coût de production souvent important et ne sont pas reproductibles, il est donc indispensable d'avoir un système performant de sauvegarde de la donnée. C'est pour cela que nous avons créé le service Zarlor : il assure cette fonction de stockage avec en plus une sauvegarde journalière et une sauvegarde délocalisée de toutes les données. Zarlor est un service central exploité par tous les autres services de l'OSU-R dédiés à la donnée.

3.3.1. Zarlor, la base de données de l'OSU-Réunion

Service de stockage de la donnée

Nom : Zarlor

Gestionnaire : OSU-R – service informatique / DSI Service Informatique Centrale

Site : Moufia (Data Center Université de la Réunion)

Système Serveur NFS associé à un espace disque important. Basé sur l'infrastructure community de l'OSU-R (3 serveurs ESXI avec Vcenter) et sur l'infrastructure de stockage de l'Université de la Réunion.

Description : Zarlor est le stockage principal de l'OSU-R. Il a été mis en place pour centraliser et conserver à un seul endroit l'ensemble des données issues des quatre observatoires/stations, mais aussi des programmes de recherches menés par les différents laboratoires affiliés. Le serveur Zarlor est hébergé dans un espace de stockage hautement disponible (SAN) géré par l'université de La Réunion. Il exploite actuellement 4 To sur le datastore, mais devrait augmenter rapidement dans les années qui viennent. Les données de Zarlor sont partagées aux différents serveurs exploitant les données (Tramontane, Filosur, ...) via le protocole NFS. Cela permet de créer des accès en mode lecture ou lecture et écriture. Le fait de passer par le système de l'université permet d'avoir une sauvegarde avec une rétention sur 45 jours. Et afin de protéger encore plus nos données, nous effectuons une sauvegarde de ces données sur un autre site de l'université.

Stratégie d'exploitation : Mise en place d'une convention de stockage et de diffusion de la donnée qui a vocation à être utilisée par l'ensemble des acteurs de l'OSU-R.

Axe d'évolution : La taille et le volume des données enregistrées et envoyées sur Zarlor ne cessent d'augmenter. En effet les usages actuels de caméra et d'appareils photos hautes résolutions font que les besoins en volumétries de stockage évoluent rapidement, pour être plus flexible et pouvoir répondre positivement à l'ensemble des sollicitations sur le stockage nous allons prochainement mettre en place une solution de stockage tampon. La solution basée sur un NAS nous permettra de gagner en souplesse et de pouvoir augmenter notre capacité de stockage rapidement.

3.3.2. Classement des données

Les données stockées sur Zarlor proviennent de sources hétérogènes et sont stockées sous des formats très variés. Pour permettre à des services associés à la gestion et à la diffusion des données

de fonctionner sur un modèle stable, les données sont classées et stockées selon une arborescence structurée et stable qui est définie dans le document “Convention d’organisation et standards pour le stockage et la diffusion des données de l’OSU-R” (DR001). Ce document décrit également les limitations d’accès aux fichiers de données en fonction des règles définies par le PI (privé, restreint, public). Ce classement est nécessaire notamment afin de fournir des chemins d’accès (ou URI-URL) stables pour la diffusion des données et des métadonnées.

Les fichiers de données sont stockés par observatoire, par collection de données, par niveau de traitement, par type de produit, par version de production et par date. La granulométrie la plus fine possible pour la date est le jour.

Les données stockées dans le répertoire “Observatoires” de Zarlou transitent d’abord par une zone de dépôt : le dossier “input”. Il est le seul dossier accessible en écriture par les fournisseurs de données (opérateurs, technicien, instrument, ...). Les données sont ensuite classées et rangées en fonction de la convention de stockage par le service Airflow (cf 3.5.1 Flux de la donnée).

3.4. Accès à la donnée

3.4.1. Les différents accès possibles

Par quel biais une personne extérieure ou encore une infrastructure de données (IDS¹) extérieure peut-elle accéder à la donnée, c.a.d connaître : l’existence de la donnée, les liens et informations d’accès ? On va retrouver plusieurs cas de figure :

- La personne ou l’IDS ne connaît pas notre organisation et/ou ne connaît pas l’existence de la donnée : le service Geosur Catalogue, accessible via internet, lui permet d’identifier les jeux de données disponibles et l’accès à la donnée. **C’est l’accès large public.**
- La personne ou l’IDS nous connaît et connaît l’organisation de notre stockage, est référencée par nos services (login/password), connaît le jeu de données qu’elle veut récupérer : le service Tramontane FTP est l’accès le plus simple et fournit un accès en temps proche du temps réel sur notre production. **C’est l’accès expert temps quasi réel.** Il est utilisé par le pôle de données AERIS pour les données de l’IR Actris-Fr et les données des SNO NDACC-Fr.
- L’IDS a un fonctionnement distribué, elle va principalement utiliser des protocoles interopérables pour moissonner les catalogues des producteurs : le service Geosur Catalogue est compatible avec le protocole CSW (protocole standard d’interopérabilité des métadonnées de l’Open Geospatial Consortium). **C’est l’accès interopérable.** Il est utilisé par le SNO Dynalit.

Nous disposons également d’un autre mode d’accès à la donnée, plutôt dédié aux personnels et experts des observatoires, qui est le service Eobs (cf 3.2.2).

3.4.2. GeOsuR-catalogue : service de catalogue de la donnée

Service de catalogage des données

Nom : Geosur Catalogue

Gestionnaire : OSU-R – service informatique

¹ Infrastructure de données spatialisées

Description : Geosur a pour objectif de documenter, diffuser et promouvoir un catalogue de données relevant des activités d'observations pérennes de l'OSU-Réunion et également ponctuelles sur projets de recherche. Ces données peuvent être des données géographiques ou des séries temporelles, à destination des laboratoires et des partenaires de l'OSU-Réunion, ainsi que des acteurs nationaux et européens œuvrant dans les divers domaines des Sciences de l'Univers.

Le catalogue Geosur fournit des outils favorisant et facilitant l'édition, la consultation et la mise à disposition de fiches de métadonnées (associées à ces données), avec un accès libre (open access via Thredds) ou restreint (FTP ou sur demande) au contenu de chaque fiche.

Les fiches de métadonnées du catalogue Geosur peuvent être moissonnées par des Geoservices extérieurs comme Sextant (IFREMER) ou Dynalit (CNRS, INSU) pour une plus large diffusion et visibilité des données produites au sein de l'OSU-Réunion.

Le catalogue Geosur permet également de répondre à certains critères de labellisation de SNO (Système National d'Observation) ou d'IR (Infrastructure de Recherche), par exemple le SNO Dynalit qui impose une distribution open-access à fréquence bi-annuelle de Modèle Numériques de Terrain (MNT) des littoraux suivis et labellisés Dynalit. La fiche de métadonnées des MNT littoraux produits à l'OSU-Réunion est accessible sur GEOSUR avec un lien d'accès direct à la donnée associée qui est stockée et archivée sur Zarlor.

Stratégie d'exploitation : Geosur est basé sur le catalogue de partage de données géographiques GEONETWORK qui est une application open source. La version utilisée est la 3.2.1 (2016) et nécessite une mise à jour vers la version la plus récente. Bien qu'ancien, le service est opérationnel. Le mode de fonctionnement actuel est basé sur l'intervention d'un administrateur qui crée une fiche de métadonnée sur la demande du producteur de la donnée (chercheur, scientifique, responsable de stations d'observation) et permet la diffusion de la donnée associée selon les critères du chercheur.

Axes d'évolution : mise à jour de l'application GEONETWORK vers la version 3.10 la plus récente. Ajout et intégration de thesaurus pour le renseignement de métadonnées selon les directives inspire. Ajout d'un Map Viewer pour l'affichage des données sur une carte.

3.4.3. GeOsuR-Data : diffusion de la donnée en « open-data »

De plus en plus, il est demandé par les financeurs à ce que les données produites avec de l'argent public soit accessible sans identification, c'est-à-dire en « open-data ». Pour cela, elle doit être téléchargeable, dans des formats et protocoles interopérables (NetCDF, OpenDap, WMS, etc). Les métadonnées doivent être renseignées et standardisées. Enfin, les données en accès libre doivent être de qualité, c'est-à-dire avoir passé un certain processus de validation, qui soit automatique ou manuel.

Service de diffusion Open Data

Nom : Geosur Data

Gestionnaire : OSU-R – service informatique

Site : Moufia

Système : Serveur Fedora installé sur community hébergeant un serveur Thredds Data Server

Description : Geosur-data est un serveur web qui fournit des métadonnées et des accès aux données scientifiques en utilisant des protocoles d'accès interopérable comme HTTP, opendap, ogc wms. L'objectif de ce service est de pouvoir fournir un lien d'accès HTTP stable et ouvert (sans identification) afin de pouvoir télécharger certains produits identifiés « open data ».

Stratégie d'exploitation :

Axes d'évolution : Le service OpenData (Thredds) est actuellement en phase de préproduction. La mise en production est prévue pour fin 2020 - début 2021, elle nécessitera des actions d'optimisation,

de mise à jour et de raccordement au service Zarlor. Ces actions sont supportées via le projet Renovrisk-transfert qui nous permettra d'illustrer l'inter communication en métadonnées et données entre notre infrastructure et l'infrastructure régionale Péigeo (<http://peigeo.re/>) et entre l'OSUR et l'observatoire des risques naturels à la Réunion (<http://www.risquesnaturels.re/>).

5.2.4. Gestion des DOI (Digital Object Identifier)

Le DOI est un attribut unique qui permet d'identifier un jeu de données spécifique, en pointant vers une page web qui regroupe toutes les informations liées au jeu de donnée, et l'endroit où se trouvent ces données. L'utilisation de DOI est de plus en plus demandée lors de la soumission d'articles scientifiques. Il y a donc une très forte demande de la part des chercheurs.

Il y a beaucoup de discussions concernant l'attribution de DOI, et en particulier qui doit les gérer. Les pôles de données sont en train de mettre en place un système de création de DOI pour tous les jeux de données qui rentreront dans leur giron. Cependant, ce système national et commun n'est pas encore prêt, il faudra encore attendre plusieurs années avant qu'il ne soit pleinement opérationnel. En attendant, il existe des solutions avec certains centres de données, ou bien des solutions clé en main, avec dépôt de fichier sur des bases de données externes.

L'OSU-Réunion, en partenariat avec l'université de La Réunion, peut proposer des DOI, qui pointeront notre catalogue de service GeOsUR. En effet, notre catalogue contient toutes les informations nécessaires à la description du jeu de données, ainsi que l'endroit où se trouvent les données.

3.5. Flux et traitements de la donnée

3.5.1. Flux de données

Le flux de la donnée représente le chemin de vie de la donnée qui part d'un instrument ou de la collecte d'un expérimentateur pour être déposé de façon automatique ou manuelle dans notre dépôt d'entrée ("zarlor-input") (étape 1). Depuis ce dépôt, la donnée entre dans un flux de classement/stockage et de traitement (ce dernier pouvant être un contrôle qualité - basiquement un tracé graphique -, un traitement afin d'extraire des paramètres géophysiques ou tout autre traitement imaginable, ou rien) (étape 2). À partir de là, l'outil qui référence l'avancement de la production, c.a.d Eobs (3.2.2), pourra venir scanner l'espace de stockage et y détecter les nouvelles données.

À l'étape (1), c'est-à-dire au dépôt des fichiers de données dans notre dépôt d'entrée "zarlor-input", le choix des outils est laissé libre pour l'utilisateur. Nous nous positionnons en assistance en proposant des clients FTP que nous avons testés afin de développer un script de transfert automatique.

L'étape (2) est réalisée à l'aide de notre service Airflow. Il nous permet d'automatiser et de contrôler un ensemble hétérogène de flux de données.

L'automatisation de l'ensemble du flux de la donnée nous donne la capacité de proposer un accès à la donnée proche du temps réel.

3.5.2. Airflow : service de gestion des flux

Service de gestion des flux

Nom : Service Airflow

Gestionnaire : OSU-R – service informatique

Site : Moufia

Systèmes : ce service nécessite un serveur de gestion principal (qui inclut la base de données des flux exécutés, le programmeur des tâches, l'interface graphique de pilotage et de contrôle, etc.) et des serveurs d'exécution des tâches de classement-stockage-traitement (le nombre de serveurs est défini en fonction du nombre de flux et de la nécessité de serveurs spécialisés, entre 2 et 5). En plus du service Community nécessaire pour créer des serveurs (cf 4.1.1). Le déploiement et la configuration des serveurs sont automatisés à l'aide du service Ansible (cf 5.2.2). La gestion et l'automatisation du déploiement des codes de flux sont assurées par le service Gitlab (cf 4.2.3).

Description : ce service s'appuie sur le logiciel Apache Airflow (<https://airflow.apache.org>), qui est devenu une référence ces dernières années pour la gestion des flux associés à une notion de date (gamme temporelle d'éligibilité du flux, capacité de reprises, gestion par date d'exécution que l'on fait coïncider avec la date d'acquisition). Un flux est constitué par un script Python qui sera exécuté par Airflow selon la programmation configurée. Ce script peut exploiter des composants de traitement disponibles avec l'API Airflow (transfert de fichiers, envoi d'email, exécution d'un code de traitement tiers, etc.). Ceci nous permet d'écrire des scripts spécialisés avec des composants standards avec une infrastructure d'exécution partagée entre tous les flux et sans trop se soucier de la charge de calcul (gérée par Airflow).

Stratégie d'exploitation : les développeurs du service informatique assurent le développement et la mise en production des flux. L'un d'entre eux assure la maintenance et la gestion globale du service. Ce service nécessite un effort important de mise en production et de maintenance, principalement du fait des montées en compétence nécessaires pour l'exploitation des logiciels associés et leurs interactions (Airflow, RabbitMQ, PostgreSQL, et en option utile Gitlab et Ansible). La définition préalable de l'organisation du service demande également un effort important d'intégration et de tests. Fort de ce constat, et le besoin d'automatisation des flux étant présents dans de nombreux observatoires et laboratoires, il apparaît essentiel de développer le partage des connaissances pour la définition, le déploiement et l'exploitation de ce service. Dans ce sens, nous avons présenté ce service à plusieurs reprises dans les réseaux et observatoires de notre communauté (ACTRIS-Fr, SIST, OVPF), et nous avons mis en place une plateforme projet² dans laquelle chacun pour trouver la documentation, les travaux et tickets d'anomalies que les autres ont déposés.

Axes d'évolution : le service est à ce jour en cours de consolidation. L'objectif étant de le rendre plus fiable et plus simple d'exploitation, de mieux le documenter et de proposer des formations afin de l'ouvrir à d'autres développeurs de l'OSU Réunion au-delà du service informatique.

3.5.3. *Licorne : traitements de la donnée*

Afin de favoriser la standardisation de la donnée, l'automatisation des traitements, le transfert en production des chaînes de traitement, et la réutilisation des codes de transformation ou de traitement des données, le service informatique développe un framework logiciel dédié, Licorne, à vocation à être exploité pour toutes les données de l'OSU Réunion.

Framework de transformation et de traitement de la donnée

Nom : **Licorne**

Gestionnaire : OSU-R – service informatique

² Outils Taïga, projet <https://projets.univ-reunion.fr/project/gabarrot-systeme-de-gestion-des-flux-airflow>

Description : Licorne³ est un logiciel de type framework écrit en Python qui offre en premier lieu un système de lecture et de transformation de format unique pour tous les types de données et de séries temporelles de données de l'OSU-R (il se base sur un format pivot compatible NetCDF et un système de plugins de lecture/écriture que l'on peut enrichir au fil du temps). En même temps qu'il importe une donnée ou une série temporelle, il peut enrichir les métadonnées à l'aide d'un système d'ajout et/ou de forçage de métadonnées. Ceci nous permet à la fois de conserver les plugins de lecture/écriture des données et de transformer le format initial en un format standard. Le format pivot dans lequel sont chargées les données est compatible avec l'univers de développement Scientifique Python et intègre les conventions de métadonnées ACDD et CF. À partir de ce format pivot, la généralisation des tracés graphiques et du contrôle qualité deviennent envisageables lorsqu'on a une importante hétérogénéité de formats et de types de données à gérer. Licorne permet également d'intégrer des algorithmes de traitement unitaires (avec une gestion de l'historique des algorithmes) et de les exécuter à la carte dans une chaîne de traitement. La définition et les paramètres d'une chaîne de traitement sont entièrement inclus dans un seul fichier de configuration au format XML (l'archive de ce fichier et du logiciel permette d'assurer la possibilité de retraitement à l'identique).

Stratégie d'exploitation : Licorne est conçu pour faciliter la standardisation des données et l'automatisation des traitements. Il a été conçu afin de fonctionner aussi bien sur l'ordinateur du chercheur que dans un système de production automatisé (il est utilisé actuellement à l'OSU Réunion par le service informatique et au pôle de AERIS par le CDS AERIS/ESPRI sur deux filières de traitement). Il s'intègre très bien dans le service Airflow (cf 3.5.2), ce dernier peut utiliser le système d'exécution de chaîne de traitement de Licorne ou l'utiliser comme une librairie en ciblant quelques fonctions. La gestion des codes et l'intégration continue utilisent le service Gitlab (cf 5.2.3).

Axes d'évolution : malgré qu'il soit déjà en production pour des chaînes de traitement spécifique, ce logiciel nécessite encore des efforts de développement afin de le stabiliser, de consolider et d'étendre ses fonctionnalités, et afin d'intégrer plus de plugins de lecture/écriture. Afin d'assurer sa pérennité et sa robustesse, le développement de Licorne suit une stratégie d'industrialisation.

4. INFRASTRUCTURE

4.1. L'infrastructure de gestion de la donnée

L'infrastructure informatique de l'OSU-R (Figure 3) a été construite pour répondre aux contraintes et aux besoins hétérogènes des observatoires et des stations qui la composent. Ainsi elle est composée de plusieurs sites qui sont reliés au réseau principal de l'université par la technologie la plus pertinente en fonction de l'éloignement, du coût et du besoin en bande passante (Liaison FH, fibre, VPN).

L'infrastructure de l'OSU-R a, en grande partie, été mise en place en collaboration avec les équipes de la direction des services informatiques de l'Université de la Réunion. Ainsi l'ensemble de nos serveurs physiques du campus de Saint-Denis sont hébergés dans le data center de l'UR, géré à l'heure actuelle à l'aide de la suite logicielle VMWARE.

³ <http://licorne.univ-reunion.fr/>

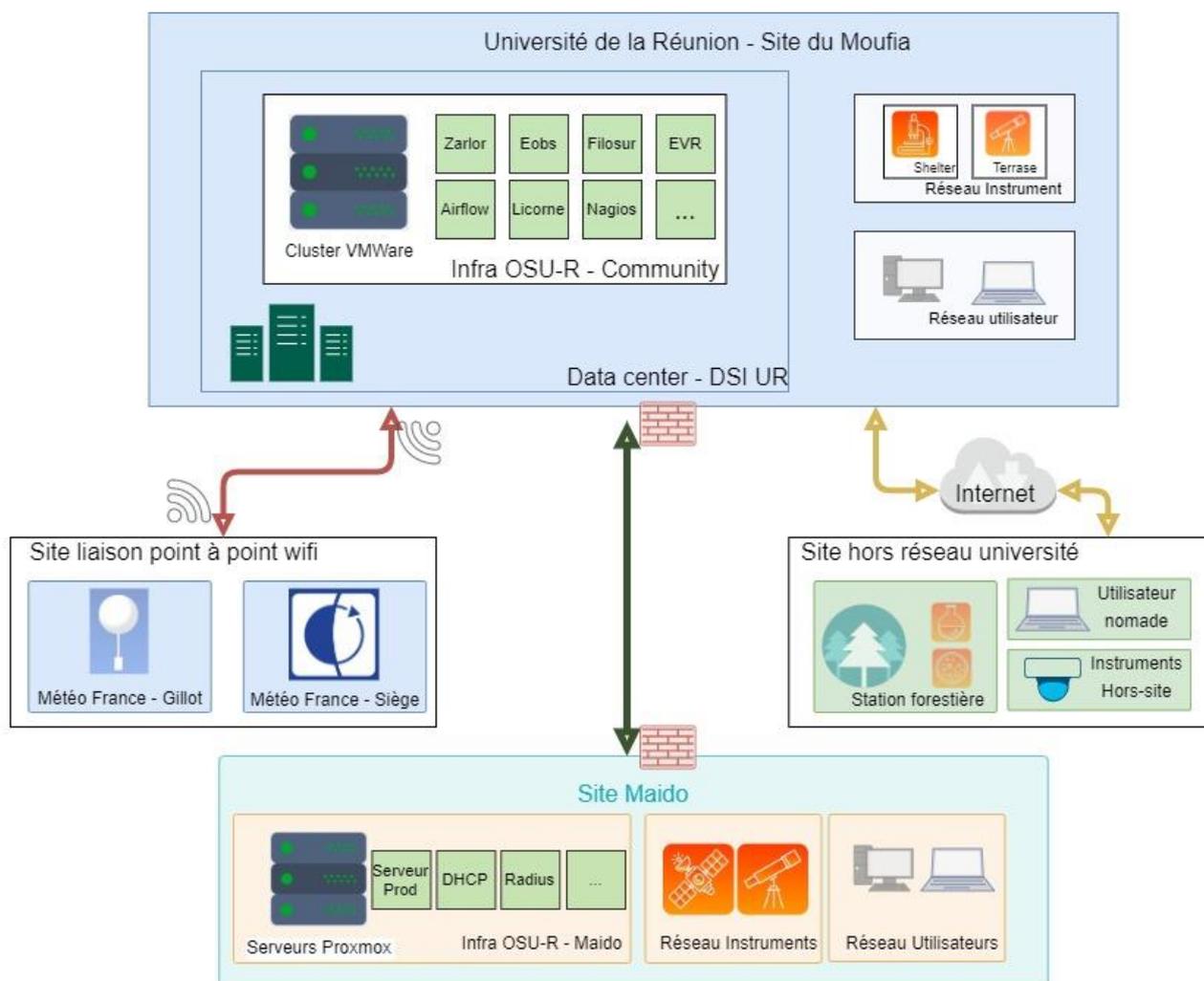


Figure 3: Schéma de l'infrastructure informatique de l'OSU-Réunion

4.1.1. Community : serveurs et espace disque

L'infrastructure-serveur principale de l'OSU-R est aujourd'hui entièrement virtualisée et intégrée dans l'infrastructure-serveur de la DSI. L'avantage de cette intégration est de permettre un redémarrage de nos machines virtuelles sur l'un de leurs hyperviseurs en cas de défaillance d'un de nos serveurs et vice versa. L'autre intérêt est de pouvoir mutualiser le financement des extensions de mémoires (SAN) et le coût des licences.

Service Community

Site : Moufia (Data Center Université de la Réunion)

Gestionnaire : OSU-R – service informatique / DSI Service Informatique Centrale

Système : 3 x Serveur IBM x3650 sous VMWare ESXI 6.5 :

Infrastructure de virtualisation de l'OSU-R basé sur les solutions VMware, pour répondre aux besoins d'hébergement et de stockage de l'OSU-R et de la fédération OMNCG. Composé de trois serveurs ESXi connectés à un SAN IBM.

Description : Community a pour objectifs premiers de répondre aux besoins de calcul (hors besoins fléchés vers les centres de calcul) et de stockage de l'OSU-R. En effet l'ensemble des serveurs de l'OSU-R sont désormais virtualisés : stockage, partage de fichiers, gestion d'identité, VPN, serveurs web. Le deuxième objectif de Community est de proposer aux laboratoires fédérés un service d'hébergement de serveurs à la demande de différents types en fonctions de leurs besoins :

- Environnement de développement (comprenant différentes versions de compilateurs)
- Serveurs web et base de données
- Serveurs de stockage, ...

Le stockage de Community se base sur une convention avec la DSI de l'Université de la Réunion. L'OSU-R en fonction des besoins achète de l'espace disque à la DSI dans leur infrastructure de stockage qui est basée sur une solution SAN (Storage Area Network) et le protocole Fiber Channel.

Les 3 hyperviseurs de l'OSU-R sont intégrés dans l'infrastructure VMWare vCenter de la DSI, ce qui permet au service informatique de l'OSU-R de se baser sur les compétences des équipes de la DSI pour le support et la mise à jour des hyperviseurs. Néanmoins l'équipe conserve une complète autonomie concernant la gestion des machines virtuelles.

Pour répondre aux différents besoins, l'infrastructure a été dimensionnée sur 3 ESXi avec un plan de jouvence sur 6 ans (jouvence d'un serveur tous les 2 ans).

Stratégie d'exploitation : la contrainte principale sur l'exploitation de ce service est la gestion de la volumétrie disque. Le nombre de serveurs physiques nécessaires par rapport aux nombres de machines virtuelles déployées est analysé chaque année afin, si nécessaire, de revoir le dimensionnement du service en termes de serveurs physiques. Concernant la gestion de la volumétrie, on différencie 2 cas d'exploitation :

1. Exploitation pour les besoins du service informatique de l'OSU Réunion :

C'est le service support essentiel qui nous permet de déployer des serveurs pour construire tous nos autres services. La volumétrie disque est principalement conditionnée par les besoins de stockage des données des observatoires et des données des projets de recherche à pérenniser. Nous évaluons annuellement l'augmentation de volumétrie à anticiper pour l'année suivante et le financement est intégré dans le budget annuel du service informatique de l'OSU Réunion.

2. Exploitation pour les besoins des équipes de recherche :

Un chercheur ou une équipe de recherche peut disposer, sans coût, sur demande au service informatique de l'OSU Réunion, d'un serveur configuré selon ses besoins et d'une volumétrie disque maximale de 250Go. Au-delà de cette volumétrie, l'utilisateur devra financer l'extension de sa volumétrie auprès de la DSI (facturation interne annuelle). Nous assistons l'utilisateur dans cette démarche. En cas de délai de mise à disposition trop important, nous déployons en ce moment un service intermédiaire d'espace de stockage sur un serveur NAS que nous administrons et qui pourra être utilisé comme solution d'attente (accès disque moins performant mais mis à disposition rapide).

Après exploitation du serveur, l'utilisateur pourra libérer les ressources qu'il occupait et, le cas échéant, faire une demande à la direction de l'OSU Réunion afin que les données du projet soient pérennisées et prises en charge par le service informatique (cas 1 / projet de recherche à pérenniser).

Axes de développement : Les serveurs actuellement en service ont été mis en service en 2014, selon le plan de jouvence sur 6 ans mis en place ils sont à renouveler à court terme (2020/2021).

4.1.2. Infrastructure des sites distants

Gillot

Le site de Gillot étant relativement peu éloigné ($\leq 5\text{km}$), il est directement connecté au réseau de l'Université de la Réunion via une liaison wifi point à point très haut débit par antenne parabolique.

Météo France - Moufia

Pour assurer l'accès au réseau de l'université par le personnel OSU/LACY depuis les locaux de Météo France une liaison wifi point à point très haut débit par antenne parabolique a été mise en place.

Maïdo

La situation géographique du Maïdo et le manque d'infrastructure des fournisseurs d'accès internet (FAI) ne permettent pas au site du Maïdo de se connecter au réseau de l'Université via un lien haut débit. De ce fait le site du Maïdo dispose d'une infrastructure système et réseau semi-autonome.

L'infrastructure système est basée sur trois hyperviseurs réunis en cluster utilisant le système d'exploitation Proxmox VE.

Pour permettre l'accès en réseau des instruments les services de base tels que DHCP et RADIUS ont été installés sur site.

Un serveur de stockage intermédiaire (MaidoProd) a également été mis en place. Cette solution a pour avantage premier que, lors de coupure d'accès réseau (fréquente), le site puisse fonctionner de manière autonome et le second avantage est de pouvoir répartir sur toute la journée le flux d'envoi de donnée vers le serveur de stockage principal (Zarlor) et ainsi d'éviter les problèmes de congestion de la bande passante du site.

Fonctionnement : les instruments envoient leurs mesures directement sur le serveur de stockage MaidoProd via le protocole FTP. Le serveur MaidoProd envoie les données via RSYNC sur le serveur Zarlor.

L'infrastructure réseau est composée d'un routeur et de 4 switches. Le réseau est découpé en cinq sous réseau virtuel (VLAN) afin d'augmenter la sécurité en séparant les flux utilisateurs des flux serveur et instruments.

Axes d'évolution : la plus grande problématique du site du Maïdo se situe au niveau de la faible bande passante. Pour pallier à ce point, dans un premier temps nous allons mettre en place des outils pour gérer la qualité de service, c'est-à-dire limiter la bande passante en fonction de la priorité de l'équipement et du besoin. Dans un second temps il a été demandé lors de l'appel d'offre concernant le marché de liaison inter-sites de l'UR (2020-2023) d'augmenter la bande passante à 40 Mbits/s.

4.2. Outils de gestion de l'infrastructure

4.2.1. Identity : gestion centralisée des accès aux services

Service de gestion d'identité (identity)

Gestionnaire : OSU-R – service informatique

Site : Moufia

Système : serveur LDAP pour l'authentification et la gestion des droits basé sur Fusiondirectory.

Description : Le serveur Identity a été mis en place en premier lieu pour permettre la gestion des accès aux services et ressources de l'OSU-R par des collaborateurs extérieurs. Il a donc pour objectif de centraliser tous les collaborateurs internes ou externes dans une seule base de données utilisateur (serveur LDAP). Ceci nous permet de gérer plus efficacement et à travers une seule interface web, l'ensemble des droits et permissions de toutes les applications de l'OSU-R.

En effet, grâce à ce serveur il est possible de regrouper les utilisateurs de manières fines et de leur donner des accès en fonction de leur besoin que ce soit par laboratoire, par projet ou par domaine scientifique. Ainsi la présence dans un groupe donne certains privilèges. Exemple : les personnes dans le groupe OPAR auront accès aux données de l'OPAR sur toutes nos plateformes (FTP : tramontane, Web : Eobs ou geosur).

Stratégie d'exploitation : Les utilisateurs personnels de l'université de la Réunion sont directement synchronisés depuis l'annuaire LDAP de l'Université grâce à un script. Les collaborateurs extérieurs doivent être intégrés à l'annuaire de l'OSU-R manuellement.

Axes d'évolution : Automatisation de l'import des utilisateurs personnels de l'UR via un script Ansible : suite à la mise à jour du produit en octobre 2019 l'import via Talend n'est plus fonctionnel.

4.2.2. Hotei : supervision de l'infrastructure

Serveur Nagios (Hotei)

Gestionnaire : OSU-R – service informatique

Site : Moufia

Système : Serveur Debian sur Community.

Description : Nagios est la solution de supervision qui a été retenue par l'OSU-R pour surveiller le bon fonctionnement du système et du réseau informatique. Il offre la possibilité de surveiller l'état de santé de l'infrastructure informatique de l'OSU-R via une interface web présentant différents tableaux de bord. En cas de dysfonctionnement, le système de supervision nous permet de recevoir des alertes et ainsi de pouvoir intervenir rapidement pour limiter la perte de donnée.

Il permet aussi de vérifier la disponibilité et le bon fonctionnement de certains des instruments déployés dans les observatoires ou par les laboratoires dans le cadre des projets de recherches.

Stratégie d'exploitation : Les instruments sont supervisés à la demande des chercheurs ou des responsables d'instrument.

Axes d'évolution : mis à jour vers la dernière version de la solution (actuellement 4.3.1 (2017)). Ajout de fonctionnalité de graphe et d'historisation via les plugins grafana et pnp4nagios. Ouvrir des accès restreints aux chercheurs et laboratoires afin qu'il puisse monitorer leur équipement de manière autonome. Une possible mutualisation avec la solution de supervision de l'Université de la Réunion sera étudiée dans l'année à venir.

4.2.3. Ansible : automatisation des déploiements dans l'infrastructure

Service Ansible / Awx-Tower

Gestionnaire : OSU-R – service informatique

Site : Moufia

Système : Serveur Debian installer sur Community

Description : En vue de faciliter et d'améliorer la gestion et la configuration de nos serveurs, l'OSU-R a mis en place un serveur ANSIBLE / AWX-TOWER.

Il permet d'automatiser un certain nombre de tâches d'administration (mise à jour des serveurs, ajout des instruments dans la supervision, création des accès VPN, ...) et ainsi d'améliorer la réactivité du service informatique.

Stratégie d'exploitation :

Axes d'évolution : Continuer l'automatisation des tâches d'administration qui le permettent.

4.2.4. Sauvegarde de l'infrastructure

Les données associées à la plupart des instruments de L'OSU-R n'étant pas reproductibles, il est primordial d'assurer la pérennité de ses données dans le temps via un système de sauvegarde robuste.

Serveur de sauvegarde OSU-R (Backuppc)

Gestionnaire : OSU-R – service informatique

Site : Moufia

Système : Serveur Debian et serveur NAS Qnap.

Description : Le logiciel Backup PC assure une sauvegarde des données des observatoires mais aussi des documents administratifs et techniques de l'OSU-R. Il est entièrement géré par le service informatique et permet ainsi une plus grande réactivité lors des besoins de restauration, mais cela nous permet aussi de nous assurer de la cohérence et de la qualité des sauvegardes.

Stratégie d'exploitation : Les données sont sauvegardées avec une rétention d'une semaine par le serveur.

Axes d'évolution : Pour le moment la rétention maximale des données de sauvegarde par l'OSU-R ou la DSI est de 45 jours, il serait intéressant de discuter avec la DSI d'une solution de sauvegarde avec des durées de rétention plus longue. Pour améliorer la sécurité des données du Maïdo, il est planifié d'installer une instance de BackupPc sur l'infrastructure Proxmox.

Service IBM TSM

Gestionnaire : DSI Université de la réunion

Site : Moufia

Système: Infrastructure IBM TSM.

Description : Service entièrement géré par la DSI : il offre la possibilité de sauvegarder entièrement les environnements VMWare sans interruption et de restaurer les machines virtuelles complètes ou juste quelques fichiers à l'intérieur des machines virtuelles.

4.2.5. Gestion de parc informatique

Service GLPI (soubic)

Gestionnaire : DSI Université de la réunion

Site : Moufia

Description : GLPI est la solution de gestion parc proposé par l'université de la Réunion. Le service a dans un premier temps été mis en place par l'OSU-R pour ses propres besoins, puis le service a été repris et mutualisé par le service informatique de l'Université de la Réunion pour gérer les ordinateurs et périphériques à l'échelle de l'université. Il nous permet de faire l'inventaire (par site, utilisateurs, projets, ...) du matériel informatique déployé dans le cadre des missions de l'OSU-R.

5. SUPPORTS INFORMATIQUES AUX UNITES DE L'OSU-R

Le service informatique de l'OSU-Réunion développe et apporte son soutien aux activités d'observation et de modélisation. Nous assistons les chercheurs et les équipes techniques lors de la mise en place de nouveaux instruments, mais aussi tout au long de la vie de l'instrument.

Nous proposons différents services ou applications pour suivre le cycle de vie des données d'observation : de l'acquisition à la diffusion en passant par le traitement et l'enrichissement de métadonnées. Pour un listing exhaustif des services de support, voir le catalogue de service "*nt_ums3365_Catalogue-services-info-a2.odt*" (DR03).

5.1. Support à l'instrumentation

5.1.1. Support à l'installation des instruments (sites de l'OSU-R et hors sites)

L'OSU-R gère le réseau de plusieurs sites distants pour les besoins des observatoires (Maido, Gillot, Météo France, Mare Longue).

Lors de l'installation d'instruments sur ses réseaux, nous pouvons accompagner les équipes en attribuant les adresses IP aux instruments et en communiquant la configuration réseau à appliquer afin que les PC d'acquisition fonctionnent correctement. Sur demande nous réalisons aussi plusieurs actions visant à la mise en production de l'instrument : configuration de logiciel d'acquisition, du serveur de temps, sauvegarde des données et de la configuration.

Nous pouvons aussi accompagner l'installation de capteurs hors locaux de l'OSU-Réunion. Grâce à la base de connaissance élaborée dans le projet **ITI**, Informatique de Terrain pour l'Instrumentation (OMNCG2016).

5.1.2. Accès et surveillance du fonctionnement des instruments

Nous proposons un service d'accès (prise en main à distance) sécurisé aux instruments pour les personnels en télétravail, en déplacement ou extérieur à l'université via une solution VPN (Virtual Private Network). Le VPN permet de connecter virtuellement un poste de travail distant au réseau intranet de l'OSU-Réunion.

Via l'outil de supervision déployé pour surveiller l'état de l'infrastructure système et réseau Nagios nous proposons aux équipes instrumentales de pouvoir surveiller le bon fonctionnement de leurs équipements et de recevoir des alertes en cas de problème. Plusieurs services ou ressources peuvent être surveillés via Nagios : vérification de la connectivité réseau, vérification de l'état d'un service ou d'un processus, utilisation des ressources système (RAM, CPU, disque)

5.1.3. *Filosur : gestionnaire de documentation*

Service de centralisation et de diffusion de la documentation

Nom : Filosur

Gestionnaire : OSU-R – service informatique

Site : Moufia

Systèmes : Serveur Debian installé sur community hébergeant un serveur Nextcloud accessible via HTTPS. Le serveur accède aux données hébergées sur Zarlou via NFS. Gestion des accès via le service identity.

Description : Nextcloud offre une plateforme d'accès aux fichiers et de synchronisation sur site avec des capacités de collaboration. Les données sont ainsi accessibles via des interfaces de bureau, mobiles et Web. Il a été mis en place à l'OSU-R pour répondre aux besoins en matière de gestion de la documentation scientifique mais aussi de la gestion des documents administratifs. En plus de la partie GED (Gestion Electronique Documentaire) Nextcloud offre des fonctionnalités de travail collaboratif via le plugin ONLY Office.

L'utilisation d'un serveur Nextcloud a aussi l'avantage d'apporter une gestion simplifiée et fine via LDAP des droits d'accès aux différentes ressources et aux documents de l'OSU-R.

Stratégie d'exploitation : Mettre en place une solution d'accès à la donnée et de travail collaboratif auto-hébergée et décorrélée des GAFAM.

Axes d'évolution : Le projet de mise en place du serveur Nextcloud est toujours en phase de test, la mise en production est prévue dans les prochains mois.

5.1.4. *Maintenance PC acquisition*

Nous assurons un service de dépannage et d'intervention sur les systèmes informatiques d'acquisition. Afin de remettre en état le plus rapidement possible les PCs d'acquisition nous pouvons remplacer les composants défectueux ou dans des cas plus critiques mettre à disposition et configurer un matériel de prêt.

5.2. Aide au développement

5.2.1. *Algorithmie & traitements & opérationnalisation*

Le pôle « Calcul Scientifique » de l'équipe informatique est composé de deux ingénieurs permettant d'aider les chercheurs à mettre en place des algorithmes de traitements, et de les opérationnaliser.

5.2.2. *GitLab : pérennisation des codes sources et qualité logicielle*

Osus-devspot : le serveur gitlab

Gestionnaire : OSU-R – service informatique

Site : Moufia

Système : Serveur Debian avec Gitlab CE

Description : Le serveur osus-devspot permet aux utilisateurs de stocker le code source de leurs développements, leurs papiers, ..., de façon sécurisée, versionnée et permettant le travail en équipe sur ces derniers.

L'outil, basé sur git, propose en sus les fonctionnalités de wiki, un système de suivi des bugs,

l'intégration continue et la livraison continue pour la qualité logicielle.

5.3. Environnements virtuels de recherche

Afin de permettre aux chercheurs et aux équipes des laboratoires de collaborer, l'OSU-R offre un service d'environnement virtuel de recherche.

Virtual Research Environment

Gestionnaire : OSU-R – service informatique

Site : Moufia

Système : Utilisation des ressources de Community : 3 hyperviseurs VmWare ESXi

Description : Les environnements virtuels de recherches ou VRE peuvent prendre plusieurs formes en fonction des besoins :

- **Serveur de calcul** : ce type de serveur permet d'automatiser le lancement d'un programme de traitement ou bien d'un modèle numérique. Cela permet à l'utilisateur de faire fonctionner ses algorithmes sur une machine autre que son propre poste et ainsi de le faire tourner plusieurs jours sans risque de coupure sur une infrastructure redondée. Ce service se situe entre l'utilisation de sa propre machine et l'utilisation d'un supercalculateur.
- **Serveur d'application partagée** : ce type de serveur permet d'installer une application sur un serveur virtuel afin de la rendre disponible à plusieurs personnes, le but étant de réduire le coût des licences sur les applications ou l'utilisation peut être mutualisée.
- **Serveur de développement** : ce type de serveur permet de mettre en place un environnement de développement contrôlé pour les travaux sur du code informatique collaboratif.
- **Serveur Web** : il est possible au travers d'un serveur virtuel d'héberger un site ou une application WEB qui sera disponible à l'ensemble des utilisateurs internes ou externes via le VPN. Si besoin le site peut aussi être mis à disposition sur internet mais avec des contraintes de sécurité plus forte.
- **Serveur d'espace disque partagé** : ce type de serveur permet de mettre à disposition un espace de stockage qui peut être utilisé soit par d'autres serveurs soit par des postes personnels. Limite : 250Go.

5.4. Services complémentaires proposés par l'université de La Réunion

5.4.1. Centre de calcul

Une des missions de la Direction des Services Informatiques de l'université de La Réunion est de proposer et mettre en œuvre la politique du système d'information dans le domaine du traitement numérique de l'information relative à la recherche. Dans ce périmètre d'action, la DSI a mis en production un Nouveau Centre de Calcul, disponible et accessible au sein de l'établissement par l'ensemble de la communauté scientifique.

L'UMS 3365 utilise occasionnellement ces ressources informatiques pour la mise en opérationnel de certains calculs numériques conséquents (par exemple pour la modélisation numérique de la propagation des vagues, des courants marins, modélisation du climat ...).

5.4.2. Salle de visioconférence

Une salle de réunion est spécifiquement équipée par un système de visioconférence de type

Polycom. L'université propose une aide pour la mise en route de visioconférence.

5.4.3. Informatique de proximité

Un service de proximité est proposé par l'université de La Réunion. Il est chargé de :

- Être le point d'entrée du support aux utilisateurs,
- Délivrer un support de proximité aux utilisateurs,
- Gérer les relations avec les services et les composantes dans le cadre de contrats de service,
- Définir et faire évoluer les caractéristiques des postes de travail : matériel, systèmes, logiciels, anti-virus, ...
- Assurer l'administration et le suivi des infrastructures : câblage Cf, commutateur réseau, salle serveur, stockage et sauvegarde des données client, serveur physique et virtuel, surveillance, téléphonie IP ...

L'équipe informatique de l'OSU-Réunion offre un service de proximité au personnel de l'UMS, en particulier pour la gestion du parc informatique utilisateur et à l'achat de petit matériel informatique. Cependant, nous ne sommes pas en mesure de fournir ce service à l'ensemble de l'OSU-Réunion.

5.5. Formations des utilisateurs

En 2019, l'équipe informatique de l'OSU-Réunion a mis en place une formation sur GitLab pour tous les personnels de l'OSU-Réunion. Le retour sur cette formation a été très bon, et nous avons décidé de faire cette formation tous les ans. Nous avons également l'idée de proposer d'autres formations, en fonction des demandes des utilisateurs.

6. ÉVALUATION DES MOYENS

6.1. D'hier à aujourd'hui, évolution des moyens humains

Depuis le début de l'OSU-Réunion, créé en 2010, l'équipe informatique a su se construire et mettre en place un schéma de fonctionnement afin de répondre aux missions demandées par l'OSU-Réunion. En 2010, deux personnes étaient présentes, répondant aux deux principaux pôles métier de l'équipe informatique : le pôle système réseau et le pôle développement. En 2013, un troisième pôle métier a été créé, le pôle Calcul Scientifique.

Le pôle système réseau est le pilier de notre organisation, car sans l'infrastructure, tous les outils tombent. Ce poste est actuellement occupé par un CDD, suite à un détachement en septembre 2019. Il sera nécessaire de récupérer le poste à la fin du détachement, qui devrait avoir lieu fin 2020.

La gestion des données passe forcément par de nombreux outils, qui ont été soit développés en interne (Eobs, Geosur) ou bien intégrés (AirFlow, nextcloud, etc). C'est le pôle développement qui est en charge de cette partie. Le nombre d'outils et services mis à disposition a nécessité depuis plusieurs années deux personnes à temps plein dans cette équipe : un permanent et un CDD. Ce deuxième poste a été pérennisé pour la fin de cette année (poste IGE de développeur sur la Campagne Emploi 2020), et permettra la maintenance et l'évolution de ces outils.

Enfin, le pôle calcul scientifique offre des services de traitement du signal aux laboratoires qui le demandent, principalement dans le traitement numérique des données ainsi que dans l'utilisation de modèles océanographiques.

Afin d'aider les producteurs des données à s'intégrer à notre système un nouveau pôle a été créé, le pôle « Coordinateurs Données », qui va permettre de solliciter les différents pôles métier en fonction des besoins des producteurs de données.

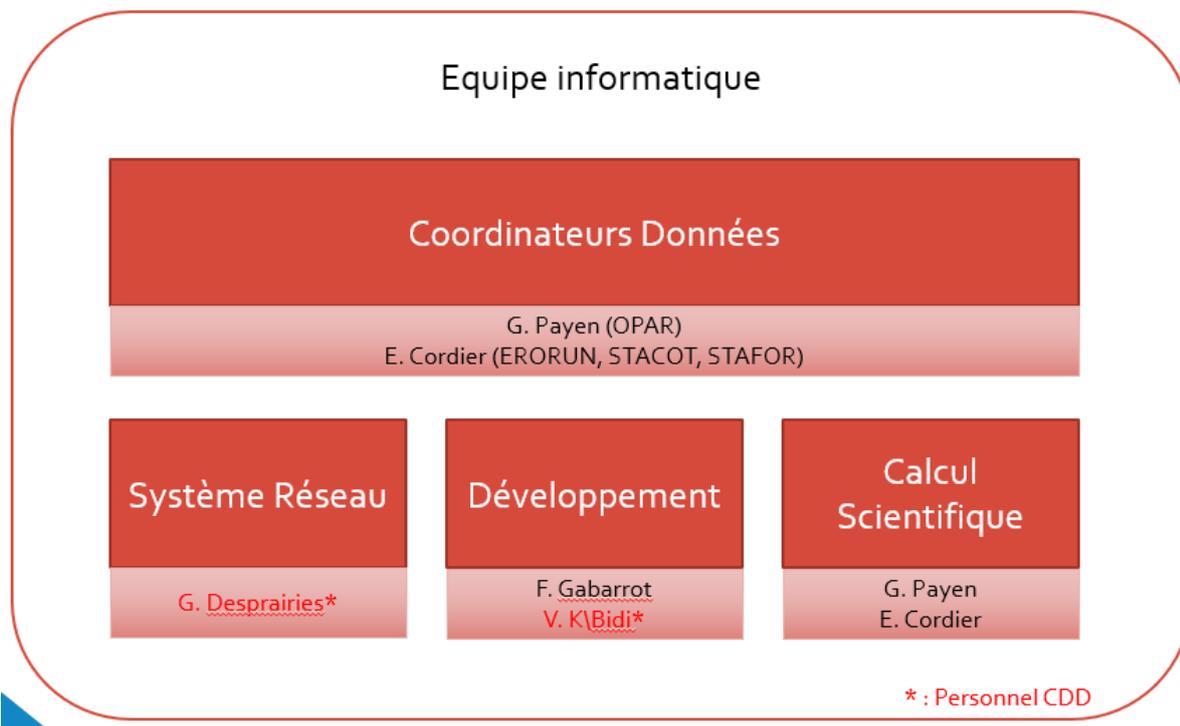


Figure 4: organisation de l'équipe informatique de l'OSU-Réunion

6.2. Stratégie à 5 ans

Dans la stratégie à 5 ans (Figure 5), nous sommes actuellement en phase de consolidation de tous nos outils. En effet, la plupart des outils sont en phase de test et ne sont pas encore totalement opérationnels. Les deux prochaines années vont nous servir à consolider toute notre infrastructure et notre offre de service. Évidemment, on continuera en parallèle l'exploitation, en particulier sur des projets spécifiques, ce qui nous permettra de tester nos outils en conditions réelles. Grâce à cela, la mise en production permettra de travailler sur la consolidation de l'infrastructure et de ces services, que ce soit au niveau sécurité ou robustesse (résistance à la panne). Plusieurs mesures ont déjà été prises en ce sens (sauvegarde et redondance des systèmes), mais la rédaction de ces documents permettra d'avoir une idée claire des éléments à améliorer pour répondre aux exigences des observatoires et des centres de données.

Dans la deuxième partie de cette stratégie, il s'agira de maintenir notre système d'information, mais également de l'exploiter sur toutes les données de l'OSU-Réunion. Cette exploitation va amener à un nouveau travail pour l'équipe, qui sera l'ajout systématique de tout nouvel instrument, ainsi qu'à la vérification et la maintenance en temps réel de tous nos services.

Nous allons également mettre l'accent sur la mise en place d'indicateurs pour chacun de nos services. Le but est de pouvoir suivre l'évolution de l'utilisation de ces services dans le temps, ce qui nous aidera à planifier les maintenances et les évolutions de nos outils.



Figure 5: stratégie à 5 ans du travail de l'équipe informatique de l'OSU-Réunion

6.3. Moyens humains et financiers

Le budget de l'équipe informatique est stable depuis plusieurs années, à 33k€ par an. Nous avons réussi à fonctionner à budget constant grâce à la mise en place d'un plan de renouvellement de nos équipements, et par un partenariat fort avec la DSI de l'université. Globalement, la répartition est la suivante : 45% d'équipement, 35% de fonctionnement, et 20% de missions. Cependant, en plus des nouveaux instruments qui arrivent tous les ans, nous commençons à voir arriver des instruments très gourmands en volumétrie de stockage (comme des caméras haute définition). Cette augmentation très forte risque d'avoir un impact sur les demandes financières de l'équipe informatique au cours des prochaines années.

Au niveau des moyens humains, durant la phase de consolidation (2020-2022), nous aurons besoin de stabiliser notre équipe actuelle : création d'un poste de développeur à la Campagne Emploi 2020 de l'UR (actuellement assuré par un agent en CDD), et remettre au concours en 2021 le poste système réseau (sur lequel il y a actuellement un agent en CDD). Ajouté à cela, il sera nécessaire d'avoir une augmentation temporaire de ressources humaines, pour finaliser tous nos outils. Pour cela, nous sommes intégrés dans plusieurs projets :

- RenovRisk Transfert : 1 CDD IR pour un an, mise en production de la gestion des données du projet. Projet commencé, début du CDD prévu pour septembre.
- Label IR - INFRA : 3 prestations par des partenaires privés : 1 prestation d'un an sur le pôle système réseau, 1 prestation d'un an sur le développement de Licorne, et une prestation de 6 mois sur geosur. Projet en cours d'instruction. Début prévu pour fin 2020.
- Label IR - RECHERCHE : 1 CDD IR pour deux ans, afin d'aider à consolider notre réponse aux besoins prioritaires des PIs (poursuivre l'automatisation des flux des observatoires et généraliser la création de quick-looks sur les jeux de données disponibles), Projet en cours d'instruction. CDD prévu pour fin 2020.
- ADAGE : mise en production de la gestion des données du projet. 1 CDD IR pour 1 an. Projet en cours d'instruction. Pas de date de début pour l'instant.

La deuxième partie de la stratégie, qui concerne l'exploitation optimale de notre système de gestion des données, nécessitera un travail différent et accés sur la vérification et la maintenance en temps réel de tous nos outils. Cette augmentation de travail ne sera pas ponctuelle mais pérenne, car le

nombre d'instruments devrait augmenter encore dans les années qui arrivent. Pour ce qui est de la gestion de la donnée, nous aurons besoin d'un assistant-ingénieur pour l'équipe développement, afin d'aider l'intégration de nouveaux instruments de notre système et de maintenir la chaîne de gestion de la donnée au moindre changement. Il sera également nécessaire d'avoir un support au sein du pôle infra-réseau, car les outils mis en place pour la gestion de la donnée sont de plus en plus complexes, et nécessitent une vigilance permanente. C'est pour cela que nous demanderons également un assistant-ingénieur en système réseau.

Nous voulons construire la logique suivante : les ingénieurs sont présents pour maintenir et créer les outils nécessaires à la gestion des données de l'OSU-Réunion, et les assistants-ingénieurs sont là pour utiliser ces outils et vérifier que les processus fonctionnent bien. Il est important d'avoir un assistant-ingénieur dans chacun des deux pôles (système réseau et développement), afin de pouvoir créer à chaque fois un binôme ingénieur – assistant-ingénieur, qui permettrait de relier l'exploitation et le développement au niveau des outils.

7. CONCLUSION

Le premier rôle de l'OSU-Réunion est d'aider à la gestion des données issues des observatoires dont elle a la charge, mais également des autres systèmes d'observation gérés par les chercheurs des laboratoires associés. Les OSU sont le lien entre les producteurs de données et les consommateurs, que sont les chercheurs des autres laboratoires, mais aussi les pôles de données, et particulièrement les CDS de chacun de ses pôles.

Depuis que l'OSU-Réunion a été créé, en 2010, l'équipe informatique s'est organisée afin de créer un environnement permettant une gestion des données la plus complète possible, permettant de répondre aux besoins des chercheurs ainsi qu'aux réseaux. Aujourd'hui, notre stratégie repose sur 5 outils complémentaires : Zarlou, pour le stockage de la donnée, Eobs, pour le stockage et la gestion des métadonnées, Airflow pour la gestion des flux de données, Licorne, pour le traitement des données, et enfin, GeOsur, pour l'accès à la donnée. Tous ces outils ne peuvent fonctionner qu'à travers une infrastructure système réseau, que nous gérons avec l'aide de la DSI de l'université. Nous avons développé plusieurs services afin de nous aider à gérer cette infrastructure. Certains de ces services ont été ouverts, que ce soit aux équipes techniques ou aux chercheurs, afin de les aider dans leur travail. Enfin, des services spécifiques ont été créés pour répondre aux besoins des différentes équipes techniques et aux chercheurs de l'OSU-Réunion.

Notre stratégie de gestion de la donnée est maintenant arrivée à un niveau de maturité important. Cependant, nos outils ne sont pas encore pleinement opérationnels, et avec le plan de charge actuel, il faudra encore deux ans de travail afin de les rendre pleinement opérationnels. Pour ce faire, nous sommes impliqués dans plusieurs projets, qui nous apporteront des ressources humaines. Une fois ce travail réalisé, nous rentrerons ensuite dans une deuxième phase, la phase d'exploitation nominale de nos outils et de nos services. Il s'agira de fournir aux chercheurs de l'OSU-Réunion la possibilité d'utiliser pleinement tous nos outils de gestion de la donnée, et de pouvoir suivre en temps réel les évolutions des demandes. Cette deuxième phase nécessitera de nouvelles compétences pour l'équipe, avec la mise en place d'une routine d'exploitation de nos services. Pour réussir cela, nous aurons besoin de deux nouveaux personnels, des assistants-ingénieurs, qui permettront d'apporter cette réponse.

Nous allons également continuer à travailler sur la communication. Un travail important sera réalisé afin de mettre en place des indicateurs sur l'utilisation de nos services, quantitatif et qualitatif. Il nous permettra de mieux cerner les outils les plus utilisés, et donc de mieux prévoir notre plan de travail sur l'amélioration et l'évolution de nos services. Ces indicateurs seront également intégrés à une communication visuelle (page web, prospectus, présentations) afin de promouvoir nos services à

tout l'OSU. Enfin, nous continuerons nos formations sur nos outils, à destination des personnels. Toute cette communication permettra de faire le lien entre notre équipe et les différents utilisateurs de nos services, ce qui nous permettra de déployer notre stratégie de travail en fonction des besoins, le tout validé par le conseil de l'OSU-Réunion.