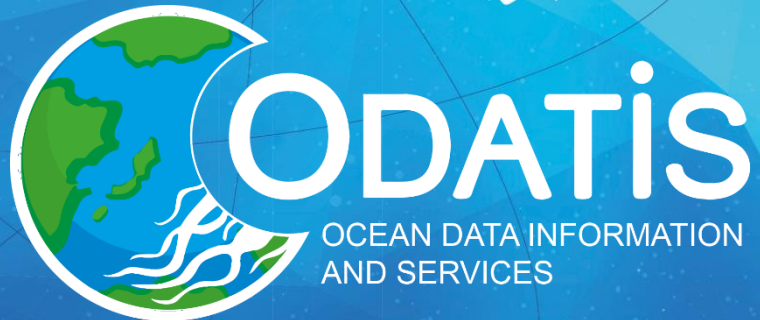


Atelier technique juin 2019

Mercredi 5 juin 2019

CDS-IS-SBR
Mark HOEBEKE





Présentation générale du CDS



Périmètre structurel du CDS

Station Biologique de Roscoff (CNRS/SU)

Services Communs FR2424

Services soutien

Service Informatique &
Bio-informatique

Plateforme
ABiMS

Service
Observation

Services support

Recherche

UMR 7144

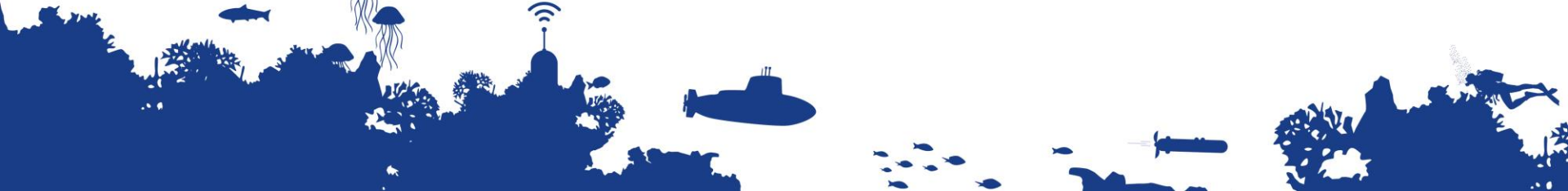
UMR8227

UMI3614

CDS – Périmètre des données

BDD PELAGOS du RESOMAR

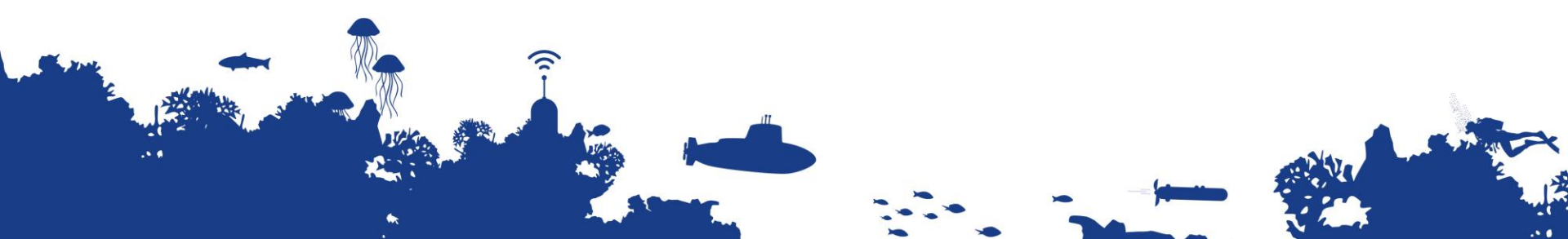
- Objectifs :
 - Bancarisation des données de biodiversité de l'écosystème pélagique côtier.
 - Mise en accès public des métadonnées
 - Mise en accès communautaire des données.
- Application Web en production depuis 2014 :
 - Export des suivis SOMLIT-ASTAN vers SeaDataCloud et EMODNet Biology
 - Exports de l'ensemble de suivis phytoplanctoniques vers la base du SNO PHYTOBS.



CDS – Périmètre des données

SNO PHYTOBS (BDD et services OGC)

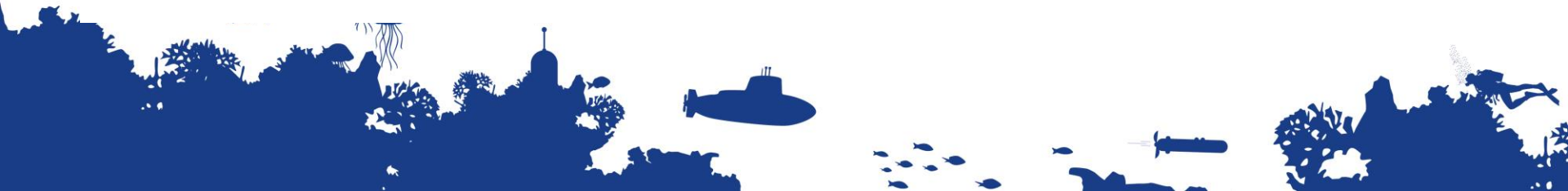
- Objectifs :
 - Développer la base de données dédiée au portail d'accès aux données du SNO PHYTOBS
 - Fournir la couche OGC d'accès à ces données
 - Fournir les outils d'alimentation à partir de la base PELAGOS du RESOMAR
- Réalisation en cours



CDS – Périmètre des données

Base locale BENTHOBS

- Objectifs :
 - Bancarisation des données de biodiversité des populations benthiques
 - Accès réservé aux collaborateurs du projet
- Application Web en fonctionnement depuis 2011
- Extractions à la demande pour alimenter la base QUADRIGE



CDS – Périmètre des données

Base de données Inventaires

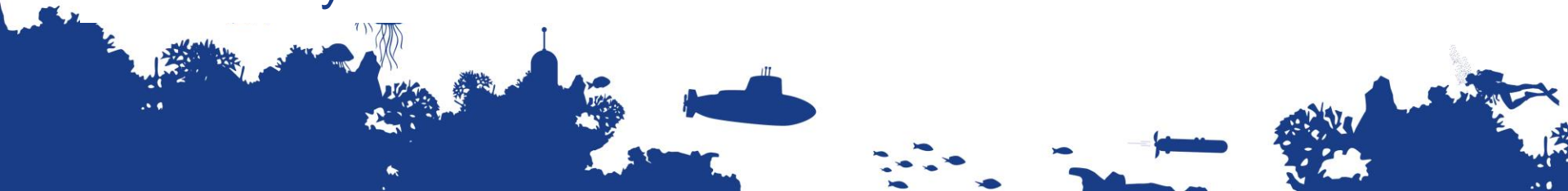
- Objectifs :
 - Mettre en accès des données de biodiversité (faune + flore) d'une zone allant de Portsall aux Sept Îles
 - Données primaires : observations de terrain + observations « historiques » (issues de faunes & flores de référence)
- Application Web en fonctionnement depuis 2011
 - Export vers EMODNet Biology

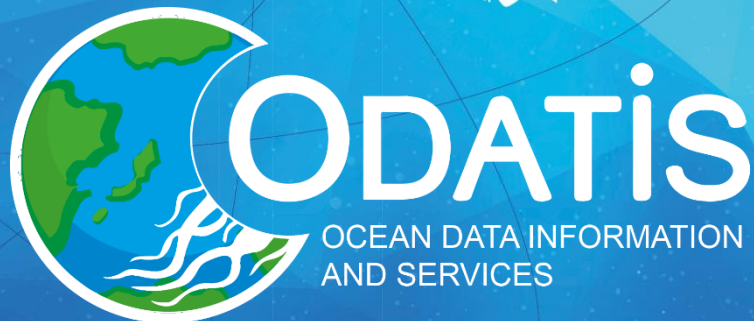


CDS – Périmètre des données

Entrepôt de données « HF »

- Objectifs :
 - Bancarisation de mesures physico-chimiques générées par des dispositifs d'acquisition « embarqués » (bouées, sondes, ferrybox)
 - Accès aux données brutes pour les collaborateurs
 - Accès à des représentations graphiques pour le grand public
- Application Web en fonctionnement depuis 2012
- Exports automatisés vers la base Coriolis :
 - Bouée ASTAN (dans le cadre du SNO COAST-HF)
 - Ferrybox

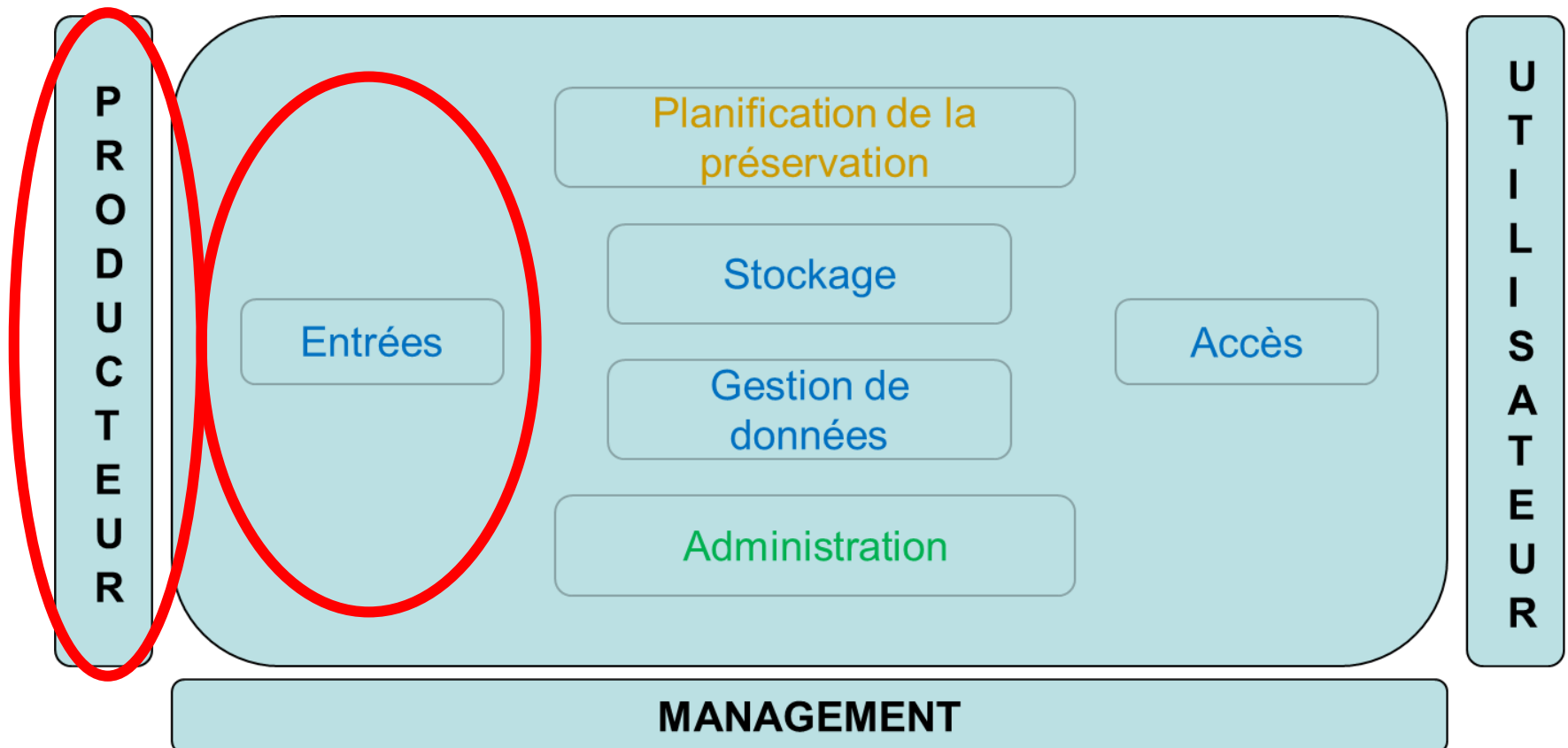




Modèle OAIS

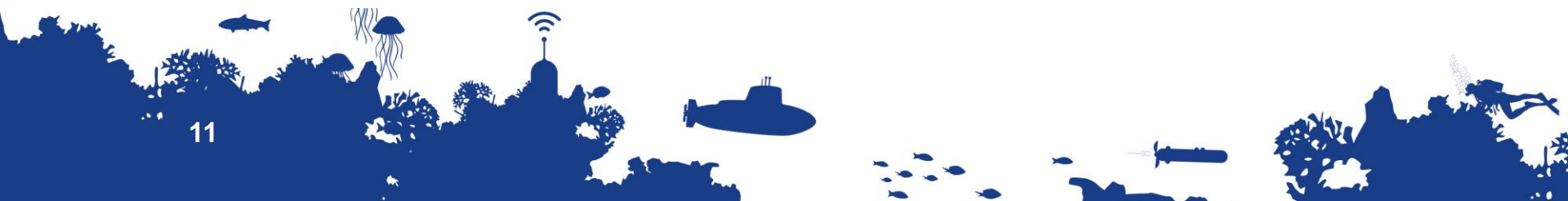


Modèle OAIS – Production & Entrées



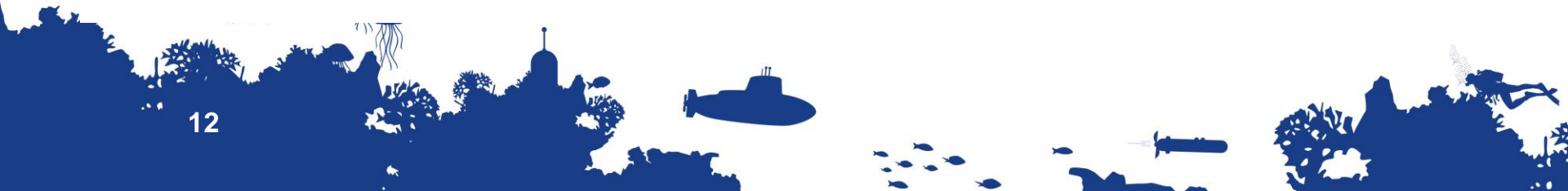
Fonction « production »

- Chaines de traitement :
 - Données physico-chimiques : acquisitions automatiques
 - Données de biodiversité : insertions manuelles
 - Dans le cadre du CDS : pas de traitements des données acquises
- Contrôle de cohérence des résultats
 - Données physico-chimiques : valeurs numériques spécifiques indiquant des défaillances matérielles, contrôle a posteriori par les responsables scientifiques, avec correction possible.
 - Données de biodiversité : le contrôle de cohérence est réalisé en amont de l'insertion par les responsables des suivis. L'outil d'insertion effectue des contrôles de cohérence supplémentaires (PELAGOS, Benthobs, Inventaires)
- Description des résultats
 - Les données insérées sont les résultats.

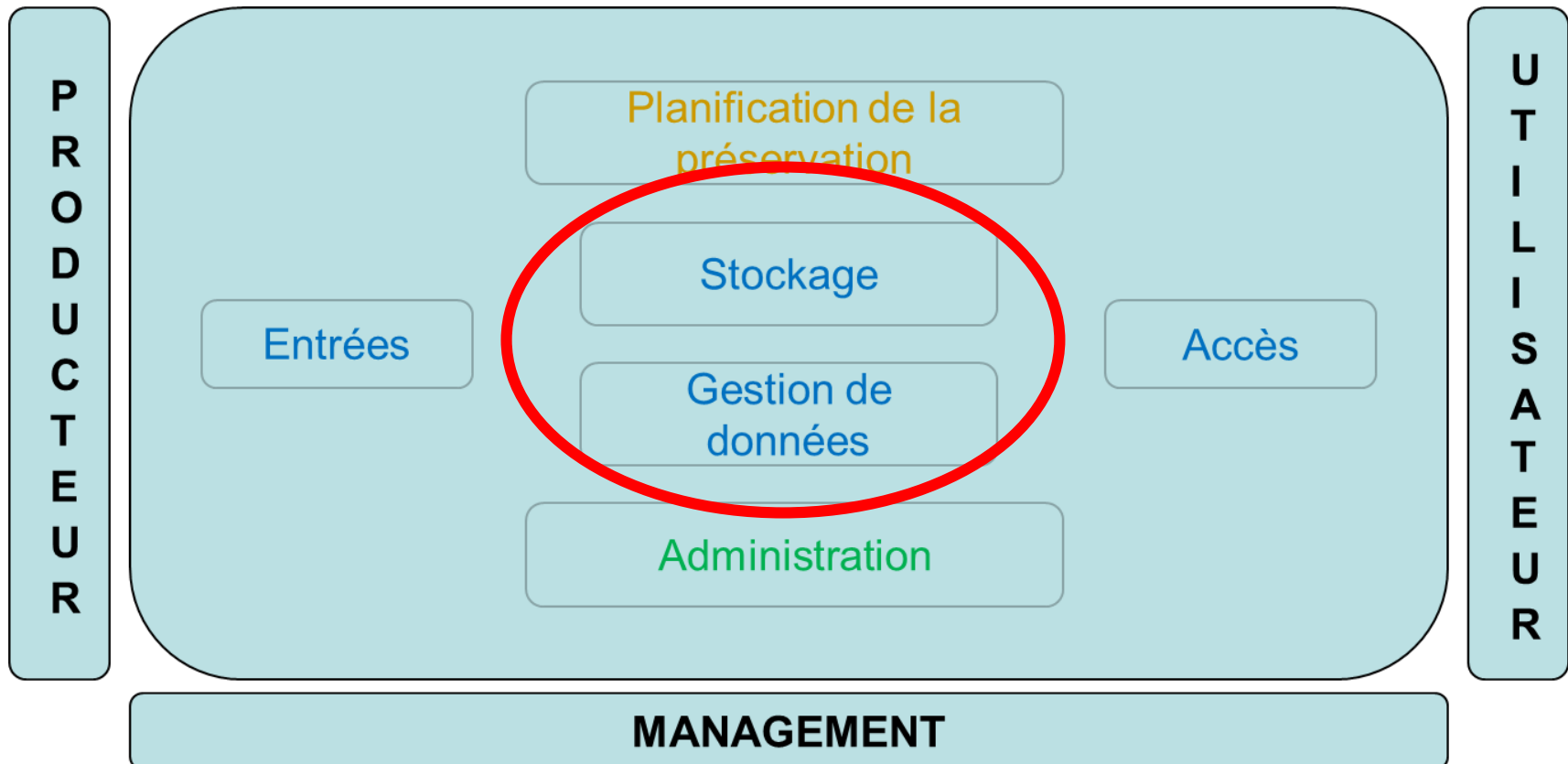


Fonction « entrées »

- Harmonisation des données
 - Données physico-chimiques : métadonnées mutualisées par type de capteur
 - Données de biodiversité : les protocoles et référentiels taxonomiques sont communs à l'ensemble des suivis.
 - Pour la base PELAGOS : Des ateliers d'experts sont organisés afin d'améliorer l'harmonisation de la détermination des spécimens.
 - Pour les données du SNO Phytobs : Une liste exhaustive des taxons ou regroupements taxonomiques est en cours de validation par les partenaires.
- Contrôle qualité
 - Pas de procédures standardisées pour le contrôle qualité.



Modèle OAIS – Stockage & Gestion des données



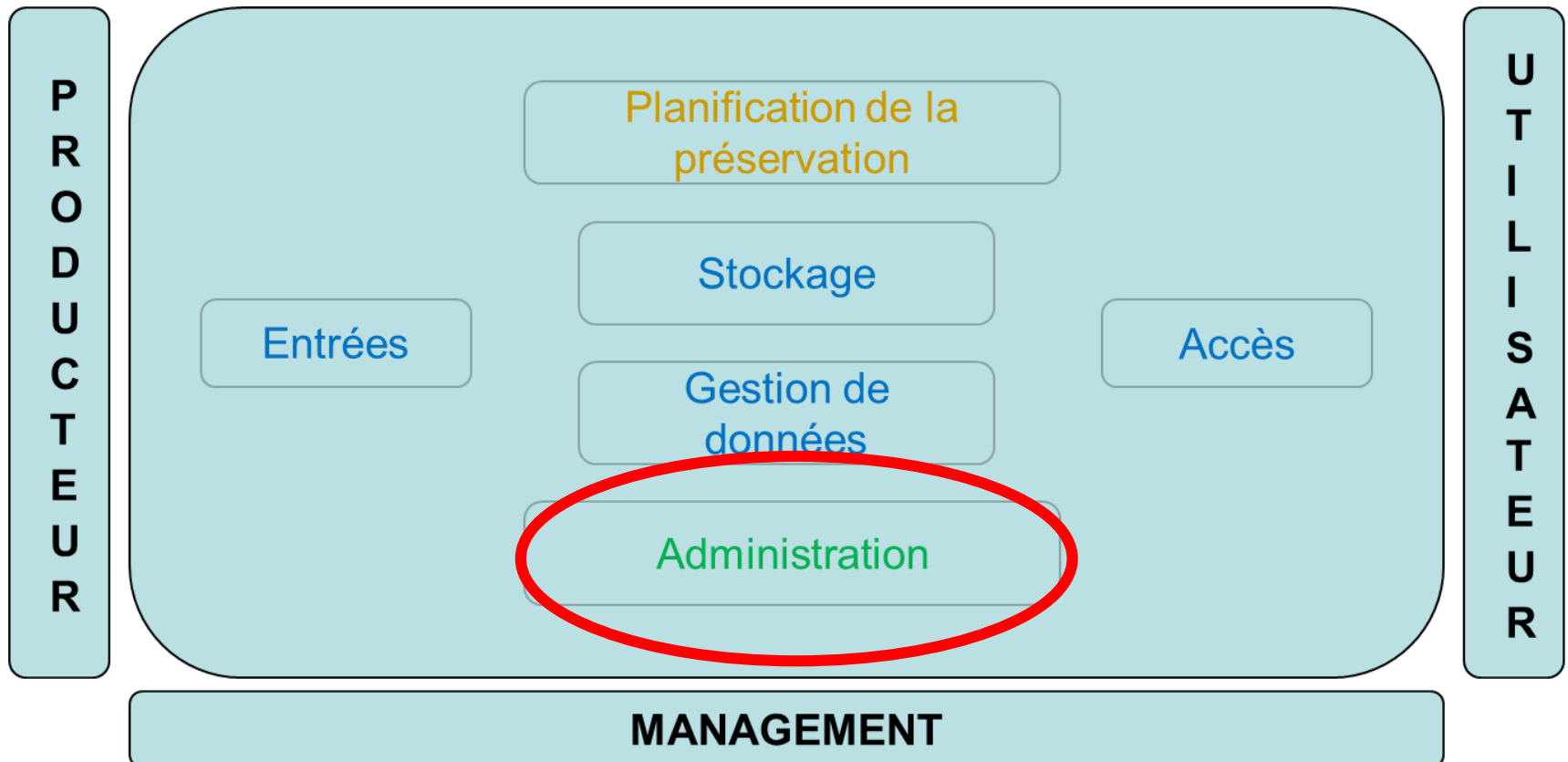
Fonction « stockage »

- Deux méthodes de stockage sont utilisées pour l'ensemble des données :
 - Bases de données relationnelles
 - Espaces de stockage dédiés pour les données annexes (images, espaces de travail, tableaux de données pré-extraits pour mise à disposition)
- Toutes les données (volumétrie totale < 1 To) sont stockées sur l'infrastructure de stockage sécurisé de la plateforme ABiMS (2 Po)

Fonction « gestion des données »

- Description des données gérées :
 - Un document intégré au SMQ recense les instances des bases de données hébergées.
 - Des descriptions sommaires figurent dans les pages documentation des applications Web permettant l'accès aux données.
- Description des résultats des chaînes de traitement
 - Pas de chaînes de traitement
- Gestion des droits d'accès et des conditions d'accès
 - Un référent scientifique par projet décide de l'octroi des droits sur les espaces de stockage, et utilise le *help desk* pour leur mise en place effective.
 - La gestion d'accès aux parties restreintes des applications est gérée dans des tableaux partagés avec les responsables scientifiques.

Modèle OAIS – Administration

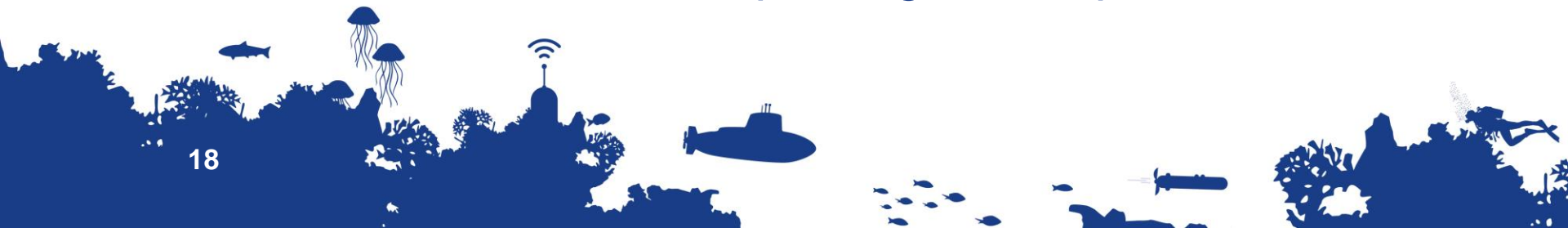


Fonction « Administration »

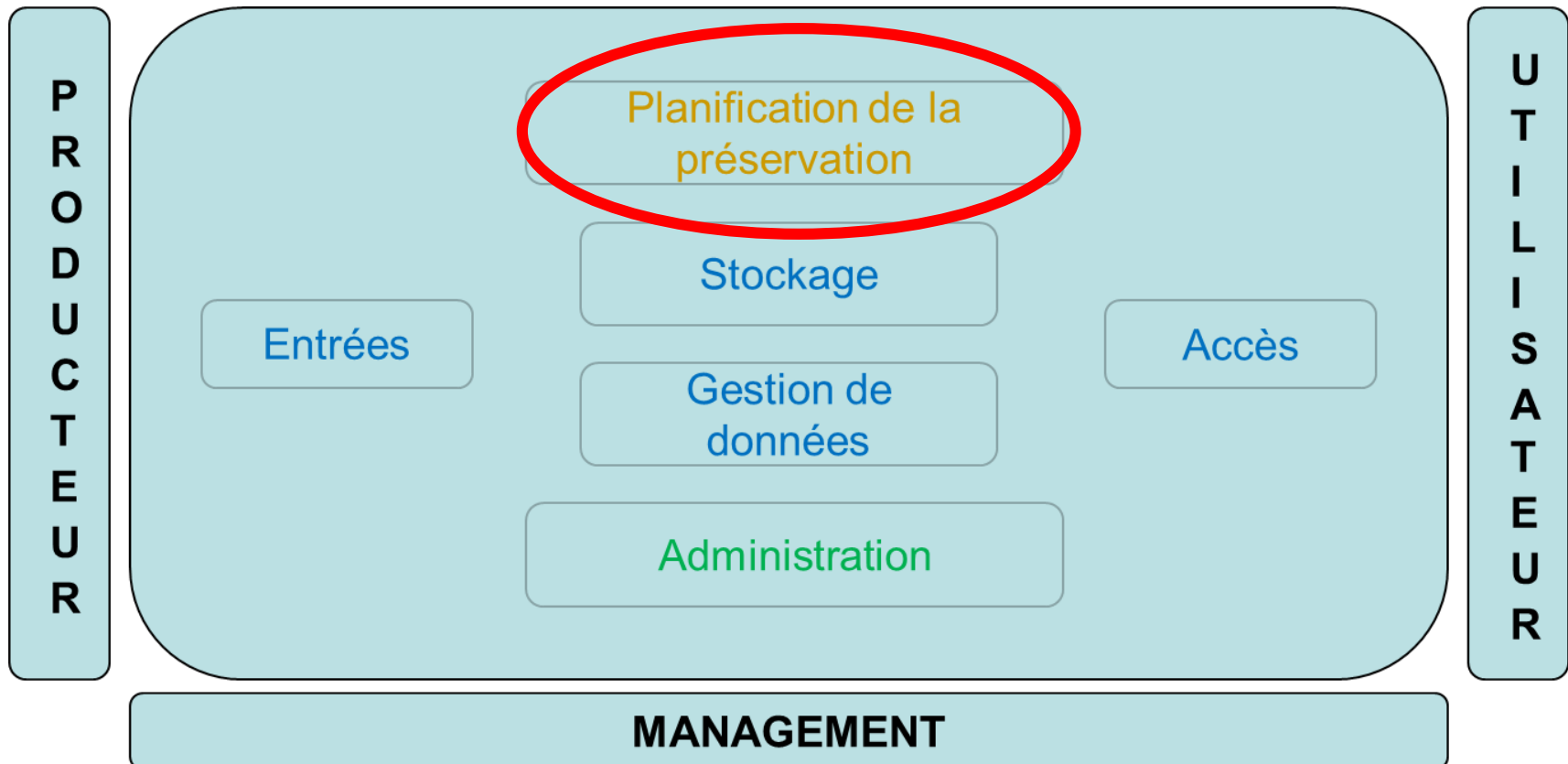
- Moyens techniques informatiques :
 - Éléments matériels de l'infrastructure hébergés dans des locaux dédiés et sécurisés
 - Administration de l'infrastructure de stockage et d'hébergement d'applications assurée par l'équipe technique locale (des %ETP ITA du Service Informatique & Bioinformatique)
- Maintien en conditions opérationnelles
 - Pour les composants de l'infrastructure relevant de la plateforme ABiMS : pris en charge par l'équipe d'administration.
 - Pour les composants tiers : pris en charge sur les %ETP dédiés à l'activité du CDS
 - Pour les développements internes : inclus dans le processus certifié « ingénierie logicielle » de la plateforme.

Fonction « Administration »

- **Maintien en configuration**
 - Pour les éléments d'infrastructure relevant de la plateforme ABiMS : pris en charge par l'équipe d'administration
 - Pour les composants tiers dédiés aux activités du CDS (IPT, Download Manager) : pris en charge sur les % ETP affectés au CDS.
- **Passage en exploitation des logiciels**
 - Pour les applications développées en interne, le processus de développement inclut une phase de test/validation avant le passage en exploitation

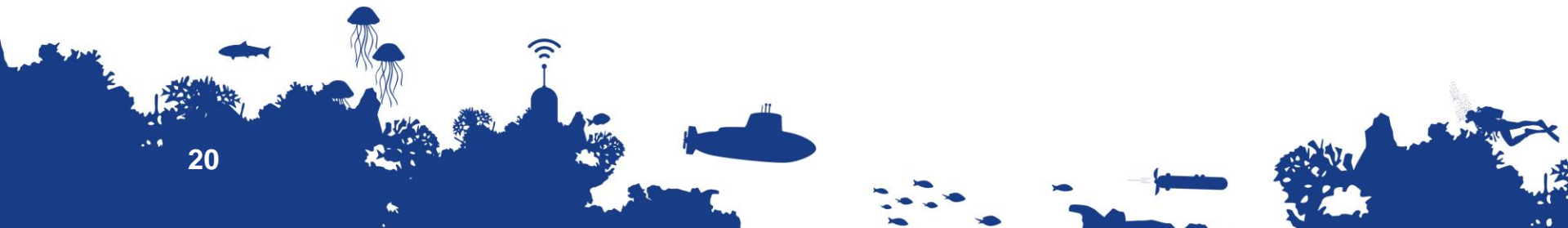


Modèle OAIS – Planification de la pérennisation



Fonction « planification de la pérennisation »

- Données et produits non reproductibles :
 - Les données de l'ensemble des suivis ne sont pas reproductibles.
 - Dans certains cas, des copies partielles sont stockées localement sur des postes de travail ou sur les dispositifs d'acquisition avec un temps de rétention limité.

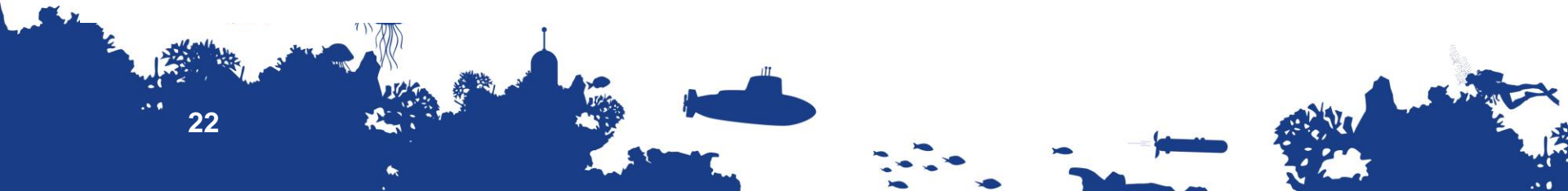


Fonction « planification de la pérennisation »

- Métadonnées :
 - Des métadonnées sommaires (horodatage, description des protocoles) sont disponibles pour l'ensemble des suivis.
 - Des métadonnées plus riches sont disponibles pour les jeux de données « publiés » dans des infrastructures/projets externes (SeaDataCloud, EMODNet Biology...) : BODC, Darwin Core

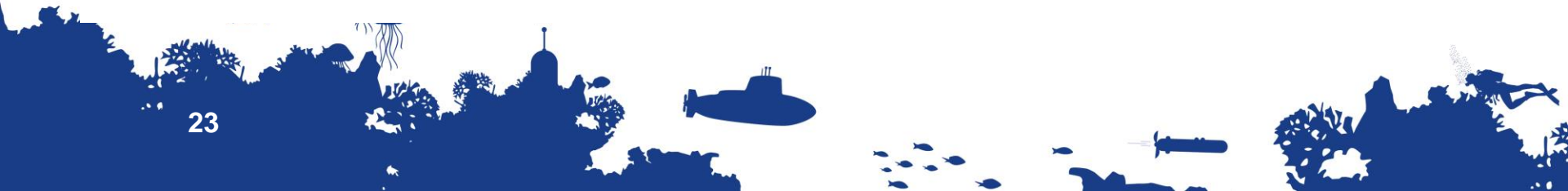
Fonction « planification de la pérennisation »

- Description des formats, vocabulaires
 - Les métadonnées « riches » utilisent les standards préconisés par les projets (vocabulaires BODC, EML)
 - Les données exportées sont au format tabulaire (COAST-HF), Bio-ODV (SeaDataCloud) ou Darwin Core Archive (EMODNet Biology)
 - Les données téléchargeables au travers des applications Web dédiées sont au format tabulaire

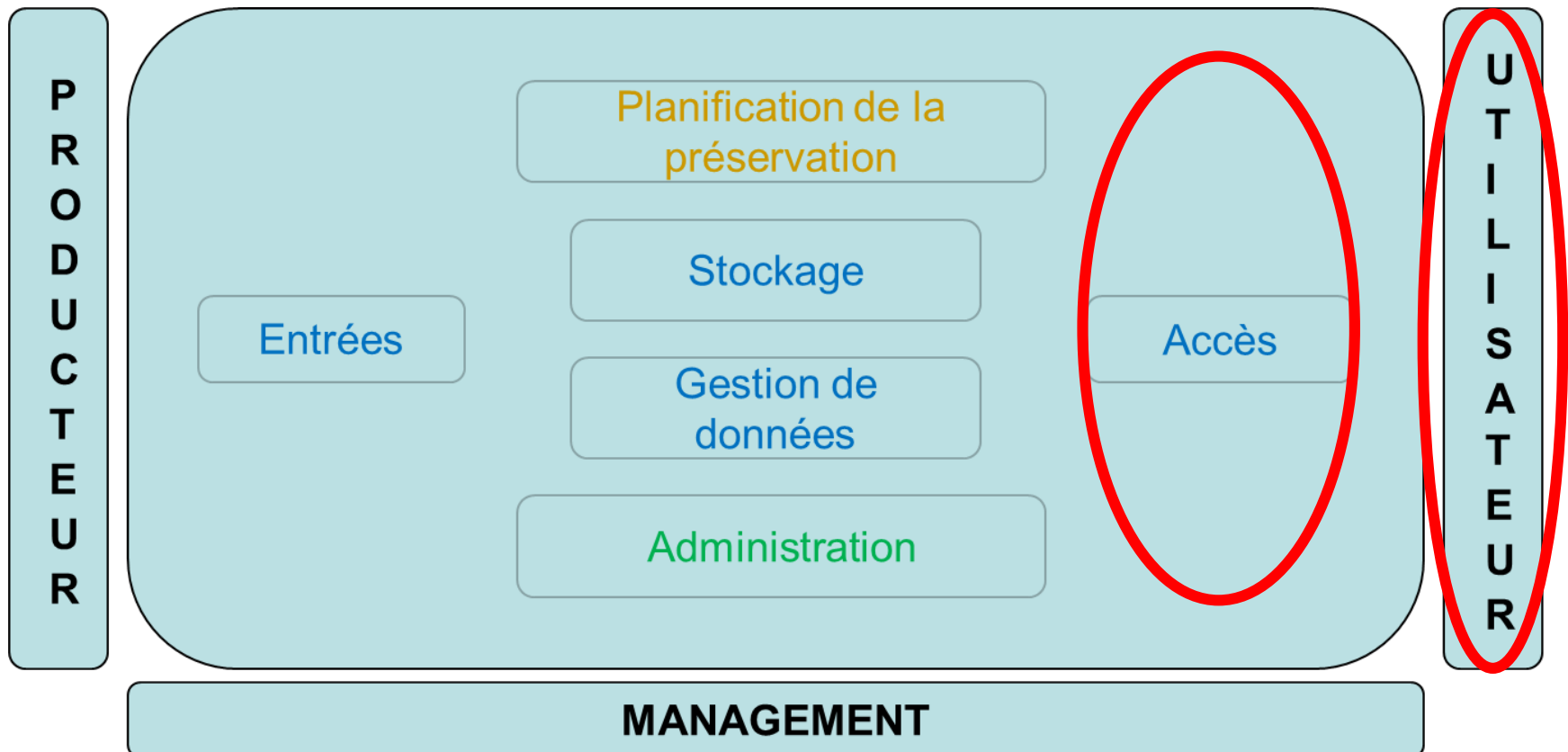


Fonction « planification de la pérennisation »

- Logiciels :
 - Base de données relationnelles : plusieurs versions de PostgreSQL sont maintenues. Les bases sont migrées vers les versions les plus récentes en fonction des besoins.
 - Applications Web dédiées développées en interne pour l'insertion et/ou la consultation des données : gestion du code et de la documentation par GitLab.
 - Applications hébergées pour les besoins de projets externes : si nécessaire, déployées dans des machines virtuelles (avec des environnements figés)

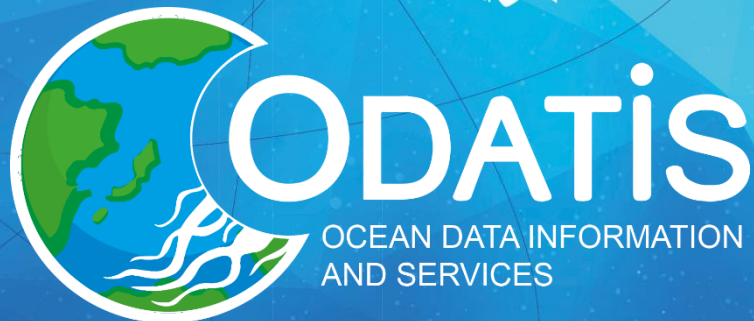


Modèle OAIS – Accès & Utilisateurs



Fonction « Accès »

- Protocoles :
 - HTTP, FTP
- Contrôle d'accès :
 - Accès aux espaces de stockage : mise en place d'ACLs par projet.
 - Accès aux bases de données : un compte dédié par base de données
 - Accès au travers des applicatifs : gérés au niveau applicatif, dans certains cas en utilisant l'annuaire LDAP local.



Besoins



Besoins techniques

- Renforcement de la sécurisation des espaces de stockage par réplication sur des sites distants.

