



Le contexte du séminaire

Le domaine de l'environnement, et par déclinaison tous les systèmes d'information qui le servent, fait interagir des réseaux d'acteurs et donc des silos de données fortement distribués. Le partage et la consolidation des informations, nécessaires à leur exploitation globale, n'en sont que plus difficiles. Au-delà des difficultés évidentes d'échange et de synchronisation, les différentes représentations de connaissances métiers entre différents acteurs et leur mise en relation pour un usage efficace est un enjeu qui rejoint les problématiques de normalisation nationales et internationales, les linked data et les nouveaux modes d'interrogation. Ces difficultés, renforcées par certaines facettes des données manipulées – comme la gestion de leur spatialisation et de leur temporalité – rendent d'autant plus ardues les processus métiers constituant les systèmes d'information environnementaux, parmi lesquels la collecte, le croisement, l'exploitation et la réexposition d'information. Seule une veille et une recherche efficace autour des outils, des méthodes et de leurs solutions technologiques garantissent de maintenir des systèmes d'informations innovants capables de répondre aux besoins des utilisateurs et des décideurs.

Confrontés à ces enjeux, l'Onema et le BRGM se sont associés en 2014 pour créer un pôle de recherche, développement et innovation centré sur les problématiques d'interopérabilité inhérentes aux systèmes d'information distribués, pôle appuyé sur les besoins du domaine de l'eau et des milieux aquatiques. Le pôle INSIDE a ainsi marqué la volonté des deux établissements de collaborer pour préparer l'évolution technologique du système d'information sur l'eau (SIE).

2017 verra l'Agence Française pour la Biodiversité (AFB) regrouper, outre l'Onema (avec ses préoccupations eau et milieu aquatique), l'Agence des Aires Marines Protégées (orientée milieu marin et ultramarin) et des organismes en charge de la biodiversité en général (PNF, Aten, ...), et un rapprochement avec le Muséum National d'Histoire Naturelle (MNHN) dans le cadre d'une Unité Mixte de Service. Le pôle INSIDE doit donc réfléchir à son évolution dans ce nouveau contexte, en s'ouvrant aux compétences en matière d'interopérabilités existantes dans les domaines biodiversité et mer.

Les compétences des organisateurs en R & D sur les systèmes d'information

Le Muséum National d'Histoire Naturelle, dans le champ de la biodiversité et géodiversité, est un acteur majeur par la richesse de ses données qui alimentent les principales bases internationales : GBIF, Worms, Fishbase, GenBank, VertNet, BOLD. Ces flux, s'appuient sur les standards du domaine. Il a été amené, depuis plusieurs années à développer des systèmes avec des problématiques d'interopérabilité technique et scientifique importantes et cela en lien avec de nombreux acteurs et partenaires. Ce travail a notamment été mené pour répondre aux besoins lors de la mise en place de systèmes d'information partenariaux tels que le Système d'Information Nature et Paysage (SINP) dont l'Inventaire National du Patrimoine Naturel (INPN), E-Recolnat, GBIF ou plus récemment 65Mo... Au-delà d'une contribution



importante sur les standards nationaux ou internationaux, il a travaillé à la mise en place d'outils permettant la collecte d'information (y compris dans le cadre de programmes de sciences participatives), le contrôle, le traitement et le partage d'information. Un travail important sur des services web de diffusion (Pour TAXREF des formats xml, json et rdf, une plate-forme SPARQL...), des outils de réconciliation taxonomique ou le développement de clés multicritères de détermination (Xper, Biodiversiclé) a été mené.

Par ailleurs, le MNHN expérimente les technologies de traitement de gros volumes de données (exploiter les technologies Hadoop Solr...) ou de protocole d'échange de données entre plateformes comme dans le cadre du projet GINCO élaboré (avec l'IGN et le MEEM) ...

L'IFREMER, pour le milieu marin, a lui aussi développé des travaux dans ces domaines. Déployant à la mer des moyens d'observations variés, il étudie et met en place des systèmes, basés sur les familles de normes OGC-Sensor Web Enablement (SWE) et l'Internet des Objets (IOT), permettant d'optimiser les transferts de données du capteur à l'utilisateur, et de garantir une description précise des conditions d'observation et une traçabilité des résultats obtenus (suivi qualité de la chaîne de mesure, de contrôle et de traitement). D'autre part, fortement impliqués dans la mise en place de réseaux de centres de données environnementales marines au niveau européen (DG-Recherche-Infrastructure de Recherche SeaDataNet, DG-Mer – European Marine Observation and Data NetWork, DG-Industrie-Copernicus Marine Monitoring Environmental Services), l'Ifremer participe aux avancées en matière d'interopérabilité technique et sémantique par l'utilisation de vocabulaires contrôlés (Ontologies), l'élaboration de modèles de données et leur implémentation... L'Ifremer implémente également des outils de supervision de systèmes distribués. Enfin, il étudie les possibilités des services « en nuage » (Cloud Storage, Cloud Computing...) afin de proposer des services performants, hautement disponibles et proposant des fonctions avancées telles que la visualisation, la co-localisation des observations, les traitements en ligne de données volumineuses (Environnements Virtuels de Traitements, « big data »).

Le pôle INSIDE a su en quelques années d'existence mener à bien des travaux variés, en commençant par une forte contribution à l'activité de standardisation internationale dans les domaines de l'eau, et ce en s'inscrivant dans des initiatives telles qu'Inspire et l'OGC ; on citera sur ce point la maturation, au sein de l'OGC:HDWG, du standard GroundWaterML 2.0 et sa mise en correspondance avec les modèles Inspire et Sandre, l'expérimentation de solutions serveur afin de diffuser les données conformément à ces standards, et le développement ouvert d'une brique cliente QGIS simplifiant la consommation des précédents flux. Les travaux du pôle ont également porté sur le développement expérimental de services web de traitement de données (WPS), avec l'atteinte pour certains d'un niveau de maturité permettant une utilisation opérationnelle ; la possibilité de les utiliser dans des problématiques plus larges a pu être démontrée, comme par exemple dans le cadre de la priorisation d'aménagement des obstacles à l'écoulement ; ces travaux visent désormais à allier l'exploitation de ressources de calculs en arrière-plan de tels services, expérimentée actuellement sur le calcul de bassins versants. Par ailleurs, dans le cadre d'un projet d'investissement d'avenir (PIA), le pôle a réalisé un prototype, hubeau.fr, plateforme d'agrégation des données sur l'eau exploitant les technologies Hadoop et consorts pour offrir



une restitution des gros volumes de données du SIE au travers d'une API REST. Enfin des travaux sont en cours sur les linked data, concrétisés récemment par un travail commun avec les équipes d'Atos visant à transposer certains standards issus de « l'ingénierie classique de modélisation » dans le monde du WebSemantique. Tous ces travaux ont été finalement l'occasion pour le pôle de nouer des relations avec d'autres acteurs, tant publics (IGN, ...) que privés (Atos, Géo-Hyd, Oslandia, Géomatys, Akka...).

Le sujet du séminaire

La création de l'Agence Française pour la Biodiversité est pour le pôle INSIDE l'occasion de s'ouvrir sur des préoccupations numériques communes aux acteurs de l'eau, du marin et de la biodiversité, afin de répondre autant aux besoins du SIE que du SINP (système d'information nature et paysage) et du futur SIMM (système d'information milieu marin).

Pour préparer au mieux cette ouverture inter-domaine, le pôle, le MNHN et l'IFREMER organisent donc, les **28 et 29 mars prochains un séminaire de deux jours, dans les locaux du MNHN.**

Son titre :

Le numérique au service de l'eau, de la biodiversité et du milieu marin

Venez définir les prochaines orientations du pôle INSIDE, dans sa dynamique d'ouverture aux enjeux numériques communs à l'eau, au milieu marin et à la biodiversité.

L'organisation de la réflexion

Une approche originale et collaborative :

Sur la base non seulement de retours d'expérience du pôle INSIDE, de l'IFREMER et du MNHN, mais aussi et surtout d'idées collectées parmi les futurs participants de l'événement (via un espace dédié créé pour l'occasion plusieurs mois auparavant sur la plateforme Imagine d'Orange), une trame de 4 à 6 thèmes majeurs sera dégagée. A chacun de ces thèmes seront associés des défis que devront très probablement relever les SI liés à l'eau, la biodiversité et le milieu marin dans les prochaines années. Ces défis, quoique nécessairement dans les problématiques d'interopérabilité technique de systèmes d'information distribués, pourront « résonner » différemment d'un domaine à l'autre, et adresser tout autant le numérique que d'autres aspects souvent indissociables (organisationnel, sociétal...).

L'événement en lui-même sera intégralement structuré autour de ce travail préliminaire, puisque, mis à part les premières heures en plénière consacrées à une mise en perspective des enjeux d'interopérabilité aux niveaux national, européen et international, la majeure partie des deux jours se déroulera en groupes de travail, selon les thèmes définis au préalable. Afin de laisser le maximum de temps aux participants pour préciser les défis pré-fléchés et co-construire un début de réponse, l'alimentation des réflexions par l'exposé de travaux existants ou en cours se fera principalement lors d'une séance de « présentations courtes » suivie de discussions autour de posters pendant les pauses. Le fruit de tous les travaux sera restitué lors



des dernières heures de l'événement, afin de garantir au mieux leur poursuite au-delà du séminaire.

A titre d'indication, voici quelques exemples de thèmes, inspirés des derniers travaux des organisateurs et d'échanges avec leurs partenaires :

- Entre démarches classiques et approches dites de rupture, quelles organisations de collecte d'information répondront demain au mieux pour chacun des 3 grands domaines métier abordés ? Quels sont les opportunités et les contraintes offertes par les réseaux de capteurs, l'internet des objets, le crowdsourcing... ? Comment récupérer/partager les données de capteur, y compris humains, utiles à l'environnement ?
- Comment organiser efficacement nos connaissances ? Le Web Sémantique est-il orthogonal aux nombreux travaux déjà réalisés par les multiples groupes de standardisation métier, ou peut-il apporter une approche complémentaire seule capable de résoudre certains écueils difficilement contournables dans l'ingénierie classique de la connaissance ? Quel est l'apport des standards d'interopérabilité pour les acteurs des domaines, comment aller vers plus d'interopérabilité opérationnelle ?
- De nombreuses données relatives à l'eau, la biodiversité et le milieu marin sont ouvertement diffusées de longue date ; néanmoins, force est de constater que leur appropriation et leur réutilisation par des sphères différentes de celles les ayant collectées est encore très marginale. Comment dynamiser cette création de valeur ? Faut-il renouveler et étendre les projets de création de plateforme d'accès, ou les hackathons, ou faut-il imaginer de nouveaux modes collaboratifs ? Ou faut-il se poser d'autres questions, tenter d'autres voies ?
- A l'instar des données, quel modèle de diffusion pour nos algorithmes, traitements et autres transformations appliquées à nos informations ? Doit-on livrer données et codes indépendamment, et/ou les assembler nous-mêmes pour n'en extraire que les données transformées, bien souvent plus rapidement intégrables dans les utilisations aval ?