

SIMM

système
d'information sur
le milieu marin



SERVICE D'ADMINISTRATION
DES RÉFÉRENTIELS MARINS



Conclusion du travail sur le référentiel taxonomique du SIMM

Historique du document

Date de création : 13 décembre 2018

Dernière modification : 13 mai 2019

Version : soumise à avis du groupe de pilotage du langage commun le
14 mars 2019

Version	Date	Auteur	Relecture	Commentaires
1	13/12/2018	C.Rabévol	O. Passafiume et S.Piel	
2	04/02/2019	C.Rabévol	O. Passafiume et S.Piel	
2.1	25/02/2019	C.Rabévol	O. Passafiume et S.Piel	
2.2	14/03/2019	C.Rabévol	Groupe langage commun	Corrections mineures

Contacts des membres du Service d'Administration des Référentiels (SAR) :

Nom	Téléphone	Mail	Organisme
Ornella Passafiume	0298224058	ornella.passafiume@ifremer.fr	Service SISMER – Ifremer centre Bretagne ZI de la pointe du Diable 29280 Plouzané
Steven Piel	0298338745	steven.piel@afbiodiversite.fr	Agence française pour la biodiversité Direction recherche, expertise et données Pôle maritime de Brest 16 quai de la Douane 29229 Brest
Clémence Rabévol	0298224695	clemence.rabevolo@ifremer.fr	Service SISMER – Ifremer centre Bretagne ZI de la pointe du Diable 29280 Plouzané

Table des matières

Remerciements	4
Objet de ce document.....	4
1. Préambule	5
1.1. Présentation du SIMM.....	5
1.2. Présentation du SAR	6
1.3. Définitions	6
1.4. Présentation des référentiels taxonomiques	8
1.4.1. Référentiels d'usage ou opérationnels au sens des systèmes d'information	8
1.4.2. Référentiels scientifiques	10
1.5. Présentation des banques de données du SIMM	11
2. Besoins	13
2.1. Rappel du cadre du SIMM	13
2.2. Besoins et services à rendre identifiés pour le SIMM	13
2.3. Besoins identifiés des banques	18
2.4. Synthèse des besoins et priorisation (SIMM et SI métiers)	20
3. Réponse des référentiels aux besoins du SIMM.....	22
3.1. Croisement des besoins et services à rendre avec l'organisation et les fonctionnalités de chaque référentiel	22
3.2. Conclusion et proposition du SAR pour le SIMM	33
Glossaire	34
Partenaires du SAR ayant apporté leur contribution à cet état des lieux	35

Remerciements

Merci à l'ensemble des partenaires du SAR (tableau en annexe) qui ont été consultés ou ont contribué à l'élaboration de ce document de proposition pour le référentiel taxonomique du SIMM.

Nous tenons notamment à remercier l'équipe de l'UMS Patrinat, du SANDRE et de Quadrige ; les experts qui ont participé à l'état des lieux ; les coordinateurs de banques, et enfin Laurent Coudercy, Antoine Huguet et Catherine Satra Le Bris qui suivent de près et conseillent les travaux du SAR.

Objet de ce document

L'objet de ce document est d'objectiver une proposition de choix de référentiel taxonomique pour le SIMM à partir des contributions des différents experts.

Nota bene

L'équipe du SAR a fait le choix d'utiliser l'appellation « taxonomie » (et non « taxinomie ») dans ces travaux dans un souci de compréhension par la majorité des acteurs du secteur et de cohérence internationale¹.

¹ Sources : http://www.btb.termiumplus.gc.ca/chroniq-srch?lang=eng&srchtxt=taxonomie&cur=1&nbr=1&lettr=indx_titls&page=9Jo4BkQesPkU.html ; <https://www7.inra.fr/opie-insectes/pdf/i97raimbault.pdf>

1. Préambule

1.1. Présentation du SIMM

Le Système d'Information sur le Milieu Marin (SIMM) est un des trois systèmes fédérateurs au niveau national (avec le Système d'Information sur l'Eau et le Système d'Information sur la Biodiversité). Il est porté conjointement par l'Agence française pour la biodiversité et par le Ministère de la transition écologique et solidaire. L'organisation et la structure du SIMM sont précisées dans le Schéma National des Données sur le Milieu Marin (SNDMM), prévu par l'article R131-34 du code de l'environnement, et en cours de finalisation.

Des outils de gestion d'informations particuliers relatifs au milieu marin existent déjà et produisent, collectent, échangent et détiennent de la donnée. L'objectif du SIMM est de les organiser afin de centraliser et pérenniser les données et de faciliter leurs échanges. Il a ainsi un rôle essentiel dans le rapportage vis-à-vis de l'Union Européenne pour plusieurs politiques publiques, comme la Directive Cadre Stratégie pour le Milieu Marin (DCSMM) et la Directive Cadre pour la Planification de l'Espace Maritime (DCPEM).

Dans le cadre du SIMM, la création du service public d'information « MilieuMarinFrance » offre un point d'accès centralisé à des informations fiables, tenues à jour et facilement compréhensibles, relatives au milieu marin, à destination des scientifiques, des décideurs publics et du grand public.

Les informations et données du SIMM, désignées comme « relatives au milieu marin », dans le périmètre des espaces maritimes et littoraux qui relèvent de la souveraineté ou de la juridiction française, doivent permettre de :

- caractériser les activités et les usages, en mer et sur le littoral,
- identifier les pressions engendrées par ces activités sur le milieu marin et littoral et évaluer leurs impacts,
- décrire et caractériser l'état des écosystèmes marins et littoraux,
- connaître les « réponses » des pouvoirs publics, c'est-à-dire les actions des politiques publiques visant à améliorer l'état des milieux marins et littoraux.

La gouvernance du SIMM est assurée par trois instances : une instance consultative, une instance de décision stratégique et des instances techniques. Les instances techniques sont composées d'un comité de coordination technique et de deux groupes spécialisés : un groupe de pilotage de l'urbanisation et un groupe de pilotage du langage commun. Le groupe de pilotage du langage commun établit les règles d'élaboration et d'emploi des jeux de données de référence et pilote les travaux du Service d'Administration des Référentiels (SAR).

1.2. Présentation du SAR

Le Service d'Administration des Référentiels (SAR) a pour rôles d'assurer l'élaboration des éléments du référentiel technique et de mettre en place les moyens techniques permettant de proposer un langage commun. Son objectif est de permettre notamment l'interopérabilité des systèmes d'informations au sein du SIMM (entre SI métiers, entre banques) tout en veillant à ce que le SIMM soit interconnecté à d'autres systèmes (autres systèmes fédérateurs, autres systèmes d'information publics, systèmes internationaux). En proposant une vision unique du référentiel de donnée au sein du SIMM, le SAR facilite ainsi la communication entre les différents systèmes d'information.

Pour cela, le SAR doit s'appuyer autant que possible sur des éléments de référence existants. Toutefois, si aucune réponse n'est adaptée au SIMM, le SAR créera de nouveaux modèles de données et de nouvelles données de référence.

L'équipe du SAR est constituée d'agents de l'Agence française pour la biodiversité (AFB) et de l'IFREMER (secrétariat), ainsi que des coordinateurs de banques du SIMM, et s'appuie sur des experts thématiques autant que de besoin.

Le secrétariat technique du SAR est chargé de :

- l'organisation de groupes de travail, rédaction de documents et consultation des usagers et producteurs de données,
- l'administration et la diffusion des données de référence,
- l'appui technique aux acteurs des SI métiers : expression de leurs besoins, prise en compte du référentiel dans les textes réglementaires, gestion et diffusion des données.

1.3. Définitions

Les référentiels taxonomiques peuvent être de deux types :

- **les référentiels dits « scientifiques »** qui prennent la forme d'un ensemble de taxons structuré par la taxonomie. Cet ensemble est souvent lié à un territoire donné au niveau duquel, les taxons sont présents ou ont été présents. Ils reposent entièrement sur une détermination scientifique avec obligatoirement des publications associées. Ils respectent les règles appliquées dans le domaine de la classification phylogénétique et notamment les règles de nomenclature internationale des taxons. Leur objectif est de rendre compte des connaissances taxonomiques.

- **les référentiels dits « d'usage » ou « opérationnels » au sens des systèmes d'information** qui prennent la forme d'une liste de taxons ou de groupes de taxons qui ont été identifiés sur le terrain dans le cadre de protocoles d'observations scientifiques parfois normalisés afin de rendre les observations partageables (au sens des échanges entre acteurs), comparables et évaluables. Les référentiels d'usage peuvent contenir des regroupements de taxons dont le nom ne respecte pas la nomenclature scientifique et dont la composition dépend des capacités de détermination induites par le protocole d'observation. Leur objectif est d'offrir un langage commun permettant l'échange de données entre plusieurs bases de données.

Les référentiels peuvent être interdépendants entre eux. L'interdépendance d'un référentiel désigne l'ensemble des liens opérationnels entre un référentiel et les autres qui sont à considérer pour l'administration de ce référentiel.

Si nous prenons l'exemple du référentiel « d'usage » des appellations de taxons Sandre, certaines appellations de taxons sont employées pour l'application de protocoles d'observations, pour le calcul d'indicateurs d'évaluation, pour la définition de groupes de taxons, etc. Le Sandre doit s'assurer que ces appellations de taxons sont toujours référencées de la même manière dans son référentiel. En effet, le fait de changer la codification d'une appellation de taxon peut impacter le calcul d'un indicateur d'évaluation et le fait de basculer une appellation de taxon dans un autre groupe impacte les observations associées à ce groupe.

Pour un référentiel « scientifique » comme TAXREF, les informations provenant de ce référentiel sont associées à la base de connaissance sur les espèces (statuts de conservation, réglementations, statuts biogéographiques) et à la base sur les traits de vie (interactions entre espèces notamment). De la même manière que pour le SANDRE, les taxons TAXREF doivent toujours être référencés de la même manière afin de maintenir ces liens et permettre aux utilisateurs de continuer d'effectuer des requêtes entre ces bases de données via le web-service (API).

Par ailleurs le Cadre Commun d'Architecture des Référentiels de données, publié par le Secrétariat général pour la modernisation de l'action publique (SGMAP) le 18 décembre 2013 définit les référentiels de données comme : « les outils informatiques nécessaires à la gestion de ces données dans le temps et leurs mises à disposition des autres applications, systèmes d'information ou utilisateurs. » Il peut s'agir de données alphanumériques, comme les appellations de taxons, ou de données géographiques, comme l'emplacement de stations de mesures.

Compte tenu de ces définitions, le référentiel taxonomique utilisé dans le cadre du SIMM doit être un référentiel d'usage, permettant de s'adapter aux besoins de ses utilisateurs et assurant l'interopérabilité de leurs données.

Toutefois, les référentiels scientifiques et d'usage interagissent entre eux. Le référentiel d'usage des taxons s'appuie sur un référentiel scientifique, notamment en tant que référence pour contrôler les demandes utilisateurs de mise à jour du référentiel. Tandis que le référentiel scientifique des taxons s'appuie sur le référentiel d'usage pour s'assurer de sa complétude.

L'objectif de ce document est donc de formaliser un choix de référentiel d'usage. Néanmoins, compte tenu des interactions entre référentiels scientifiques et d'usage, les référentiels scientifiques ont également été étudiés afin d'identifier lequel de ces référentiels constituerait l'expertise la plus adéquate comme appuie au référentiel d'usage du SIMM.

1.4. Présentation des référentiels taxonomiques

Les référentiels choisis pour être étudiés lors de l'état des lieux effectué avec les experts ont été sélectionnés car ils sont utilisés par les banques de données qui participeront au SIMM. Comme vu précédemment, les référentiels d'usage et scientifiques interagissent entre eux, c'est pourquoi des référentiels de ces deux catégories ont été examinés.

1.4.1. Référentiels d'usage ou opérationnels au sens des systèmes d'information

Le référentiel d'appellation taxonomique du SANDRE

Le Service d'Administration Nationale des Données et Référentiels sur l'Eau (SANDRE) est assuré par l'Office international de l'Eau (OIEau), sous le pilotage de l'AFB. Il a pour objectif de permettre le partage des données sur l'eau et l'environnement dans le cadre du Système d'Information sur l'Eau. Depuis 1993, le SANDRE a mis en place plusieurs référentiels, notamment concernant les appellations taxonomiques. Le référentiel d'appellation taxonomique est un référentiel d'usage composé de deux jeux de données : « appellations de taxons » et « groupes d'appellations de taxons ».

Ce référentiel accessible sur le site du SANDRE est défini par plusieurs documents de spécification:

- une présentation générale des données vulgarisant les pratiques et principes de fonctionnement d'un domaine particulier,
- un dictionnaire qui décrit d'un point de vue informatique les données ; il structure la base de données en la définissant et permet ainsi son interopérabilité,
- un scénario d'échanges définissant le format et les règles d'échange de certaines données,
- un document d'administration des référentiels qui détermine le périmètre et les règles organisationnelles et techniques des données de référence,
- des données de référence qui constituent les données élémentaires.

Pour mettre à jour son référentiel, le SANDRE s'appuie sur TAXREF, et à défaut sur d'autres référentiels scientifiques comme WoRMS.

Aquatic Sciences and Fisheries Information System (ASFIS)

L'Aquatic Sciences and Fisheries Information System (ASFIS) est un référentiel taxonomique fournissant un système de codification normalisé couvrant la plupart des catégories d'espèces liées aux activités de pêche. Il est publié par le Service des statistiques et de l'information sur les pêches et l'aquaculture de l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO). C'est un référentiel d'usage au service des statistiques de pêche et qui n'a pas vocation à faire autorité d'un point de vue taxonomique. Ses sources principales d'informations sont les publications du FAO FishFinder et de la base de données ASPHA. Le référentiel ASFIS assigne trois types de codes à chaque catégories d'espèces :

- un code CSITAPA (Classification Statistique Internationale Type des Animaux et des Plantes Aquatiques de la FAO) qui répartit les espèces commerciales en 50 groupes
- un code taxonomique utilisé par la FAO pour une classification plus détaillée des catégories d'espèces
- un code alpha-3 utilisé pour les échanges de données avec les correspondants nationaux et entre organismes de pêches, et qui nous intéresse donc dans le cadre du SIMM.

1.4.2. Référentiels scientifiques

World Register of Marine Species (WoRMS)

Créé en 2007, WoRMS découle du registre européen des espèces marines (ERMS) et de la combinaison de plusieurs registres spécifiques gérés par l'Institut des sciences marines de Flandre en Belgique (VLIZ). Il est le référentiel « scientifique » international sur le milieu marin, utilisé dans le rapportage des données de biodiversité pour la DCSMM (comme recommandé par le Groupe de Travail DATA de la Commission européenne).

Il est basé sur une base de données appelée « Aphia » composée de noms d'espèces validés, de synonymes, de noms vernaculaires, et d'informations supplémentaires (références bibliographiques, répartition biogéographique, photographies et/ou dessins, liens vers pages web). Chaque nom scientifique est également lié à son taxon parent. Ces informations sont accessibles via l'interface web « MarineSpecies.org ». Aphia et MarineSpecies.org sont principalement financés par l'Union Européenne et sont hébergés par le VLIZ. WoRMS combine les informations d'Aphia avec celles d'autres listes comme AlgaeBase et FishBase.

TAXREF

TAXREF est le référentiel taxonomique national pour la faune, la flore et la fonge de France métropolitaine et d'outre-mer, élaboré et diffusé par le Muséum national d'Histoire naturelle (MNHN) dans le cadre de la mise en œuvre du Système d'Information sur la Nature et les Paysages (SINP). Ce référentiel « scientifique » est porté actuellement par l'UMS PatriNat (AFB-CNRS-MNHN). Il a pour but premier de lister et d'organiser les noms scientifiques de l'ensemble des êtres vivants recensés sur l'ensemble des territoires français. Il répond aussi à des besoins métiers dans le cadre de programmes européens concernant des taxons ne se trouvant pas sur le territoire national. Le référentiel TAXREF est consultable et téléchargeable sur le site internet de l'INPN. Une nouvelle version est mise en ligne annuellement (version 12 en 2018). L'ensemble des sources et la méthodologie mise en œuvre pour élaborer et alimenter TAXREF sont décrits dans un document associé aux versions de TAXREF. Les sources utilisées pour l'alimenter sont les bases de données validées et officielles (comme WoRMS), des publications scientifiques et les données de groupes d'experts (comme l'initiative française pour les récifs coralliens).

1.5. Présentation des banques de données du SIMM

Plusieurs banques de données qui devront participer au SIMM ont été étudiées durant le travail d'état des lieux effectué en amont de la rédaction de ce document. Il ne s'agit pas d'une liste exhaustive mais des principales banques, existantes ou encore à l'état de projet, utilisant un référentiel taxonomique et déjà identifiées par le groupe de travail sur l'urbanisation au début du travail du SAR sur ce mandat. L'ensemble de ces banques sont présentées ci-dessous. Les besoins de ces banques par rapport au référentiel taxonomique du SIMM sont exposés dans la suite de ce document (partie 2.3).

Quadrige

La banque Quadrige a été développée par l'IFREMER pour gérer les données de la surveillance du littoral, étendu depuis à la surveillance marine. Elle est utilisée pour la saisie de données sur le terrain et en laboratoire, leur diffusion au grand public et le calcul d'indicateurs (DCE, DCSMM, etc.). Elle constitue déjà un élément du Système d'Information sur l'Eau et contribue donc aux travaux du Secrétariat d'Administration des Données Relatives à l'Eau (SANDRE), auquel elle est conforme.

HARMONIE

La banque HARMONIE du Système d'Informations Halieutiques (SIH) intègre un référentiel taxonomique interne. Elle est rattachée à un système d'information métier du SIMM, est portée par l'Ifremer et constitue le réseau opérationnel d'observation des ressources halieutiques et des usages associés. La couverture thématique de ce référentiel concerne donc les espèces d'intérêt halieutique, scientifiques et commerciales.

AMP

La banque relative aux aires marines protégées (AMP) est portée par l'Agence française pour la biodiversité et elle est rattachée à un système d'information métier du SIMM. Elle permet de localiser et d'obtenir des informations sur les aires marines protégées (dont les AMP françaises sur la totalité de l'espace maritime français). Elle est utilisée dans la gestion des aires marines protégées et le suivi d'espèces et d'habitats « à statuts », ainsi que la production d'indicateurs et de chiffres clés pour les conventions de mers régionales ou les directives Natura 2000 (Habitats-faune-flore et Oiseaux).

RESOMAR PELAGOS BENTHOS

Le Réseau des Stations et Observatoires Marins (RESOMAR) est une structure financée par le CNRS et l'IFREMER qui rassemble des stations, observatoires et laboratoires marins français. Il gère deux bases de données : « Benthos » et « Pelagos », dont l'objectif est de collecter et sécuriser les données biologiques des stations marines puis de les exploiter à l'échelle nationale.

ESTAMP

La base de données ESTAMP créée par l'Agence française pour la biodiversité a pour objectif de sécuriser et valoriser les données issues des suivis écologiques et d'usages sur les estrans au sein des aires marines protégées. Cette base de données est accompagnée d'une application web adaptée aux gestionnaires permettant l'accompagnement de la saisie, la bancarisation et la diffusion des données depuis un PC ou une tablette.

BIGOOD

La base de données BIGOOD (anciennement BIOCEAN), opérée par IFREMER, couvre la thématique des environnements profonds. Elle contient des données issues de campagnes océanographiques. Les taxons prélevés sont identifiés, généralement par étapes de tris successifs et expertise finale par un taxonomiste. La diffusion des données se fait au travers de systèmes tels qu'OBIS (Ocean Biogeographic Information System).

Banque mégafaune marine (en projet)

La banque mégafaune marine est portée par l'Agence française pour la biodiversité. Elle est rattachée à un système d'information métier du SIMM. Encore à l'état de projet, elle a pour objectif de rassembler, donner accès et diffuser les données des bases PELAGIS, OBSenMER, Observatoire Oiseaux et d'autres sources de données existantes ou à venir. La couverture thématique de cette banque de données concerne donc tous les suivis de la mégafaune marine (oiseaux, mammifères marins, tortues et élastombranches) par observations visuelles et acoustiques, échouages et biopsies.

Espèces exotiques envahissantes (en projet)

Encore à l'état de projet, un Système d'Information sur les espèces exotiques envahissantes verra le jour d'ici quelques années. Il est important que le SAR le prenne en compte afin de permettre son interopérabilité avec le SIMM dès sa mise en service.

En effet, les espèces exotiques envahissantes sont une problématique de plus en plus présente en France avec une réelle nécessité de mutualiser les informations pour permettre une meilleure gestion de ces espèces. A titre d'exemple, l'AFB et l'UICN ont constitués une base de données sur les espèces exotiques envahissantes en outre-mer. Il s'agit également de données qui pourront être impliquées dans le rapportage pour la Commission européenne puisque le descripteur 2 de la DCSMM porte sur les espèces non indigènes.

2. Besoins

2.1. Rappel du cadre du SIMM

Comme le rappelle le dossier d'urbanisation du SIMM (en cours d'écriture), l'organisation actuelle entre les différents partenaires, caractérisée par une mise en réseau limitée et un manque de mutualisation des outils de bancarisation des données, ne permet pas un accès aisé à l'information. Le rôle du SIMM est donc de mettre en place des dispositifs d'échanges normalisés, de communication et d'intégration des informations sur le milieu marin.

Afin d'aboutir à un système cohérent et interopérable, des référentiels de données et des données de référence sont nécessaires. En effet, pour les données sur les taxons par exemple, les banques qui seront incluses dans le SIMM utilisent divers référentiels taxonomiques (TAXREF, WoRMS, ASFIS) ainsi que parfois leurs propres identifiants. Cette hétérogénéité sémantique et de codification entre les banques ne permet pas l'interopérabilité entre elles car il n'y a pas de langage commun défini. Le référentiel taxonomique du SIMM est donc essentiel afin de permettre l'échange des données contenues dans ces banques et servir de pivot entre les SI métiers du SIMM et ceux des autres SI fédérateurs (Eau et Biodiversité).

2.2. Besoins et services à rendre identifiés pour le SIMM

Le référentiel taxonomique utilisé dans le cadre du SIMM doit répondre à ses besoins et aux règles nationales inscrites dans le Cadre Commun d'Architecture des Référentiels de données. L'ensemble de ces besoins généraux et les services qui y sont rattachés sont présentés ci-dessous.

Référentiel opérationnel

Le SIMM nécessite un référentiel taxonomique opérationnel au sens des systèmes d'information (voir partie 1.3), permettant de rattacher ou de positionner des données métiers. Afin de répondre à ce besoin, le référentiel utilisé doit rendre plusieurs services :

- Le référentiel doit être en capacité de gérer des groupes de taxons.
- Le référentiel doit pouvoir contenir tous les taxons nécessaires aux utilisateurs du SIMM (y compris ceux non connus scientifiquement ou éventuellement pas encore référencés dans des territoires français) en leur attribuant un identifiant stable dans le temps.
- L'attribution de ces identifiants doit se faire dans un laps de temps court compatible avec les contraintes de bancarisation des données et de leur traitement.
- Le référentiel doit disposer de services d'abonnement aux mises à jour afin de faciliter le travail de mise en conformité au SIMM des coordinateurs de banques.

Codification

Le SIMM nécessite un référentiel taxonomique disposant d'un identifiant stable, unique, non réutilisé, et géré, permettant d'identifier chaque objet sans ambiguïté ni doublon. Afin de répondre à ce besoin, le référentiel utilisé doit suivre des règles précises pour la gestion de ses identifiants. Ces règles sont énoncées dans le Cadre Commun d'Architecture des Référentiels de données :

- l'identifiant doit être facilement partageable (dans un format interopérable) ;
- il doit être non ambigu ;
- il doit être non signifiant ;
- il doit être non modifiable : une fois défini et attribué, il ne doit plus changer ;
- il ne doit pas être réaffecté à un autre objet métier ;
- Il ne doit pas être supprimable, même si l'objet n'a plus lieu d'être ;
- Il doit être persistant : c'est-à-dire qu'il doit être réellement stocké, conservé et archivé dans le temps.

Mise à jour

Conformément à la règle RS4 du Cadre Commun d'Architecture des Référentiels de données (énoncée ci-dessous), le référentiel taxonomique utilisé par le SIMM doit adapter la réactivité de ses mises à jour aux besoins opérationnels des utilisateurs.

Règle RS4 du Cadre Commun d'Architecture des Référentiels de données :

Les dispositifs de mise à jour des données d'un référentiel doivent être ancrés dans des processus métiers opérationnels liés aux traitements du cycle de vie des objets métiers considérés. En fonction de la nature des données de référence, le juste compromis entre d'une part la réactivité des mises à jour (et des corrections) et d'autre part le contrôle à la source, doit être définie en toute transparence. Dans tous les cas, l'opérationnalité du référentiel doit être privilégiée, à savoir : la réponse aux besoins opérationnels des utilisateurs des données de référence. « Il est indispensable de privilégier une très forte réactivité dans la mise à jour et la correction des données de référence. La capacité des dispositifs de mise à jour doit le permettre.

Afin de répondre à ce besoin, le référentiel utilisé doit être en mesure d'effectuer des mises à jour en continu, au fur et à mesure des demandes de modification ou d'ajout de taxons par les utilisateurs.

Accessibilité

Le SIMM nécessite un référentiel taxonomique dont les données et les outils doivent être accessibles à tous librement, selon des modalités techniques adaptées aux besoins. En effet, selon le Cadre Commun d'Architecture des Référentiels de données, la logique générale de mise en place d'un référentiel et notamment de ses services doit répondre aux principes *API first* et *open first*. Afin de répondre à ce besoin, le référentiel utilisé doit :

- disposer d'une sémantique connue, et limitée au strict nécessaire à la fonction de référence ;
- concevoir l'ensemble du dispositif sous une forme ouverte et accessible largement et simplement à travers des API utilisant les technologies et les principes d'interopérabilité du web ;
- permettre de disposer des données sous différents formats de données et différents protocoles d'échanges normés.

Qualité des données

Le SIMM nécessite un référentiel taxonomique proposant des données de référence de qualité connue et contrôlée. Afin de répondre à ce besoin, le référentiel utilisé doit faire l'objet de contrôles qualité et d'un processus d'amélioration en continu comme décrit dans le Cadre Commun d'Architecture des Référentiels de données :

- Le référentiel doit décrire, mesurer et communiquer sur le niveau de qualité atteint pour chaque donnée de référence et progresser en communiquant sur le niveau visé.
- Selon la règle RM3 (citée ci-dessous), le référentiel doit avoir défini des indicateurs de mesure de la qualité des données, qui sont mesurés et publiés. Il doit également disposer d'outils permettant d'analyser ces indicateurs et d'agir en conséquence sur les données.
- Selon la règle RM4 (citée ci-dessous), le référentiel doit proposer un service d'alerte permettant aux utilisateurs de faire remonter toutes anomalies sur les données.

Règle RM3 du Cadre Commun d'Architecture des Référentiels de données :

Les indicateurs de qualité sont identifiés dans les dispositifs d'acquisition et de distribution des données de référence. Ils sont mesurés et publiés.

Règle RM4 du Cadre Commun d'Architecture des Référentiels de données :

Chaque référentiel doit intégrer un dispositif d'alerte ou de signalement permettant à un utilisateur du référentiel de faire remonter au responsable du référentiel toutes anomalies sur les données détectées en aval (incomplétude, incohérence, doublon, amalgame, problèmes d'intégrité, etc.). Le processus de traitement des signalements doit être également décrit, à jour et publié. Il est recommandé également de rechercher son automatisation et donc son outillage.

Gouvernance

Le SIMM nécessite un référentiel taxonomique disposant de règles d'administration connues et contrôlées. Afin de répondre à ce besoin, la gouvernance du référentiel utilisé doit être conforme au Cadre Commun d'Architecture des Référentiels de données :

- Le référentiel doit disposer d'un cadre de gouvernance accessible et à jour, conformément à la règle RS3 (énoncée ci-dessous).
- Le référentiel doit faire l'objet d'un pilotage opérationnel assurant la mise en place, l'évolution, voire la transformation dans le temps du référentiel, en fonction des besoins des acteurs. Un dialogue régulier doit être assuré entre les utilisateurs et ceux qui administrent le référentiel pour le faire évoluer dans la concertation et selon des priorités définies collégialement.

Règle RS3 du Cadre Commun d'Architecture des Référentiels de données :

Tout référentiel de données dispose d'un cadre de gouvernance à jour, accessible librement à tous.

Interopérabilité

La Commission Européenne² définit l'interopérabilité comme : « l'aptitude d'organisations disparates et diverses à interagir en vue de la réalisation d'objectifs communs mutuellement avantageux, arrêtés d'un commun accord, impliquant le partage d'informations et de connaissances entre ces organisations à travers les processus métiers qu'elles prennent en charge, grâce à l'échange de données entre leurs systèmes de TIC respectifs. »

Le SIMM nécessite un référentiel taxonomique permettant d'assurer son interopérabilité à tous les niveaux. En effet, les référentiels du SIMM serviront de pivot entre les SI métiers du SIMM et avec les autres SI fédérateurs. Selon le Schéma National des Données sur le Milieu Marin, le SIMM doit être interopérable avec les autres Systèmes d'Information de l'Etat, notamment le SI Eau et le SI Biodiversité ; mais également avec des SI régionaux ou internationaux (comme dans le cadre du rapportage à la Commission européenne ou encore à un secrétariat d'une convention de mer régionale telle que OSPAR).

Afin de répondre à ce besoin, le référentiel utilisé doit :

- Diffuser la correspondance de ses identifiants avec ceux d'autres référentiels utilisés dans certaines banques (TAXREF, WoRMS, ASFIS, etc.) pour les échanges, et permettre de passer de l'un à l'autre facilement. De plus, ces correspondances permettent potentiellement de proche en proche de se raccrocher à d'autres référentiels connexes.
- Proposer des services permettant d'attribuer des identifiants à une liste de noms d'espèces, convertir des codes d'un référentiel en codes d'un autre référentiel, etc.

² Article 2 de la décision n° 922/2009/EC du 16 septembre 2009

2.3. Besoins identifiés des banques

Le Système d'Information sur le Milieu Marin (SIMM) centralise les données de plusieurs banques de données (voir partie 1.5.). Il est donc important de prendre en compte les caractéristiques de ces banques dans le choix du référentiel taxonomique du SIMM. L'ensemble des besoins des banques et les services qui y sont rattachés sont listés ci-dessous.

Transcodage

Afin d'assurer l'interopérabilité des Systèmes d'Information entre eux et avec le SIMM, les banques de données doivent être en mesure de transcoder facilement leurs identifiants avec ceux utilisés dans le cadre du SIMM. Le référentiel taxonomique du SIMM doit donc assurer un transcodage avec les référentiels utilisés par les banques. Afin de répondre à ce besoin, le référentiel utilisé devra comporter à minima une correspondance avec les identifiants de :

- WoRMS : utilisé par Quadrigé, Harmonie, SI AMP, SI RESOMAR PELAGOS BENTHOS, banque mégafaune marine, ESTAMP et BIGOOD.
- TAXREF : utilisé par Quadrigé, SI AMP, banque mégafaune marine et ESTAMP.
- ASFIS : utilisé par HARMONIE et ESTAMP

Certaines banques utilisent des référentiels métiers comme source de données. Ce sont des référentiels plus spécifiques dont le transcodage n'est actuellement pas géré par les référentiels étudiés. Afin d'assurer l'interopérabilité des banques utilisant ce type d'identifiants, le référentiel taxonomique du SIMM devra être en capacité de gérer le transcodage de ses identifiants avec des référentiels supplémentaires.

Réponse aux demandes des utilisateurs

En vue de répondre aux besoins des banques en termes de réactivité, le référentiel taxonomique du SIMM doit assurer une codification des taxons dans un délai raisonnable. Cela assure à l'utilisateur la possibilité d'obtenir rapidement le code SAR correspondant à un taxon, avant tout échange de données avec le code provisoire saisi en local. En effet, ces identifiants sont indispensables aux échanges de données et à la diffusion qui doit se faire la plus proche possible du temps réel afin de permettre une utilisation rapide (calculs d'indicateurs par exemple). Par exemple, au retour des campagnes halieutiques de nombreuses données sont intégrées dans les banques (BIGOOD, Harmonie, Quadrigé, etc.) ce qui nécessite la création rapide des taxons non nécessairement présents dans le référentiel.

A titre d'exemple, Quadriga assure un traitement des demandes de ses utilisateurs en moins de sept jours.

Afin de répondre à ce besoin, le référentiel utilisé doit se limiter à un temps de réponse pour la codification d'un taxon de moins d'un mois.

Les thématiques et les zones d'études des banques de données du SIMM sont très diverses. Il est donc nécessaire que la couverture thématique et spatiale du référentiel taxonomique du SIMM soit assez souple pour s'adapter aux demandes des banques. Il doit être en mesure de permettre la codification de tous les taxons qui y sont bancarisés :

- Il doit pouvoir intégrer les taxons non présents de manière documentée en France. C'est par exemple le cas pour les données de suivi traitant d'espèces de passage ou nouvellement naturalisées en France. Certaines banques peuvent également comporter des données portant sur des observations en dehors du territoire national, par exemple en haute mer.
- Il doit être en mesure de gérer aussi bien les taxons marins que les taxons d'eau douce. En effet, des informations sur des taxons d'eau douce pourront être échangées dans le cadre du SIMM au niveau de la zone commune entre DCE et DCSMM, en particulier sur les masses d'eaux côtières (exemple des amphihalins qui sont présent dans les deux milieux selon leur cycle de vie).

Pour des raisons très diverses liées notamment à des facilités de suivi des espèces, les banques créent des regroupements de taxons. Le référentiel taxonomique du SIMM doit donc être en mesure de gérer ces regroupements de manière à permettre l'interopérabilité des données et le rapportage. Cette exigence a été émise par plusieurs banques de données : Harmonie, la banque de données AMP, Quadriga et la base de données ESTAMP.

Afin de répondre à ce besoin, le référentiel utilisé doit être en mesure de répondre favorablement aux demandes de création de groupes de taxons quel que soit leur type (et non nécessairement selon un rang taxonomique). Par exemple, Quadriga gère des regroupements correspondant à des unités taxonomiques identifiables sur le terrain (exemple : algues filamenteuses) ou pratiques pour l'analyse des échantillons (exemples : regroupement des « espèces carnivores » pour les mesures de biomasse des invertébrés benthiques).

Autres services aux banques de données

Compte tenu du nombre de données de référence dans le référentiel taxinomique (plusieurs dizaines de milliers de données), la mise à jour manuelle de l'ensemble des données est impossible. Afin de permettre aux banques de données contribuant au SIMM d'effectuer les mises à jour de manière automatique le référentiel taxinomique du SIMM doit proposer un service d'abonnement aux mises à jour, sous forme utilisable directement par les systèmes informatiques, au rythme et à la fréquence jugée utile par les systèmes d'information.

Le domaine thématique du SIMM couvre tous les groupes taxonomiques, des animaux aux végétaux en passant par les organismes unicellulaires. Dans l'ensemble de ces taxons, certains portent le même nom mais à des niveaux taxonomiques ou avec une classification différente (ex : le genre *Planaria* T. Brown, 1827 correspond à un mollusque alors que le genre *Planaria* Müller, 1776 correspond à un vers némerte). Pour garantir l'identification des taxons de manière non ambiguë, le référentiel taxinomique du SIMM doit donc inclure dans son modèle de données un arbre taxonomique permettant de distinguer ces taxons ainsi que l'autorité de classification qui distingue la plupart des homonymes.

2.4. Synthèse des besoins et priorisation (SIMM et SI métiers)

A partir de l'état des lieux précédent des besoins du SIMM et des banques de données qui y seront rattachées, une synthèse des principaux besoins a été établie afin de les organiser et les prioriser :

- **Gouvernance du référentiel**

Le référentiel taxinomique du SIMM doit disposer de règles d'administration connues et contrôlées. Il doit faire l'objet d'un pilotage opérationnel assurant la mise en place, l'évolution, voire la transformation dans le temps du référentiel, en concertation avec les utilisateurs. Les besoins des utilisateurs doivent pouvoir être exprimés facilement et traités de manière régulière et/ou rapide. Ils doivent pouvoir être tracés et les utilisateurs tenus informés des avancées.

- **Accessibilité du référentiel**

Le référentiel taxinomique du SIMM doit être accessible à tous librement. Sa sémantique doit être connue, ses données disponibles sous différents formats et ses outils conçus sous une forme ouverte (API).

- **Codification de tous les taxons**

Le référentiel taxonomique du SIMM doit pouvoir codifier tous les taxons nécessaires aux utilisateurs (y compris s'ils ne sont pas validés par une publication scientifique dans l'immédiat ou dont la présence n'est pas connue à l'échelle de l'espace maritime français).

- **Gestion de groupes de taxons**

Le référentiel taxonomique du SIMM doit être en mesure de gérer des groupes de taxons et de créer les groupes nécessaires aux besoins des utilisateurs tant pour la bancarisation des données que pour leur traitement.

- **Transcodage avec d'autres référentiels**

Le référentiel taxonomique du SIMM doit être en mesure de fournir à ses utilisateurs les équivalences de ses identifiants avec ceux d'autres référentiels utilisés par les banques de données (WoRMS, TAXREF, ASFIS). Il doit proposer des services permettant de convertir des codes ou d'attribuer des codes à une liste de noms d'espèces.

- **Identification unique des taxons**

Le référentiel du SIMM doit permettre d'identifier sans ambiguïté chaque donnée de référence et doit donc contenir dans son modèle les informations nécessaires à la distinction des différents taxons (arborescence = parentalité des taxons, autorité de classification, niveau taxinomique, etc.).

- **Réactivité des mises à jour**

Le référentiel taxonomique du SIMM doit offrir un service permettant aux utilisateurs de demander l'ajout de nouveaux taxons. Le temps de réponse pour la codification d'un taxon doit être de moins d'un mois. Les mises à jour doivent être effectuées en continue au fur et à mesure des demandes des utilisateurs.

- **Gestion des identifiants**

Le référentiel taxonomique du SIMM doit respecter les règles énoncées dans le Cadre Commun d'Architecture des Référentiels de données pour la gestion de ses identifiants.

3. Réponse des référentiels aux besoins du SIMM

3.1. Croisement des besoins et services à rendre avec l'organisation et les fonctionnalités de chaque référentiel

Les besoins du SIMM pour son référentiel taxonomique, identifiés dans la partie précédente ont été croisés avec les informations sur les référentiels, obtenues lors d'un travail d'état des lieux effectué avec divers experts de la thématique. Un tableau synthétise les résultats de cette jonction en fin de partie.

Bien que les référentiels étudiés aient des objectifs différents, ceux-ci ont été croisés avec les mêmes besoins du SIMM car il n'y a pas eu de différenciation entre besoins liés aux référentiels scientifiques et besoins liés aux référentiels d'usage dans leurs caractérisations. La différenciation entre ces deux types de référentiels interviendra dans la conclusion.

- **Gouvernance du référentiel**

WoRMS

WoRMS ne répond pas à ce besoin du SIMM car sa gouvernance ne permet pas une prise en compte rapide des demandes d'évolutions. Il n'y a pas de réelle instance / outil permettant de recueillir les besoins et les avis des utilisateurs quant au développement du référentiel. On notera tout de même que la gouvernance de WoRMS est bien structurée et présentée de manière assez détaillée et facilement accessible sur leur site. De plus, de nombreux rapports (rapport d'activité annuelle de l'équipe de gestion des données, compte-rendu de réunions et notamment du comité de pilotage, conseils à l'organisation des groupes de travail, etc.) sont disponibles librement en ligne.

WoRMS est géré à trois niveaux, chacun ayant ses propres responsabilités :

- Le comité de pilotage qui est constitué de membres du comité éditorial élus pour trois ans se réunissant trois fois par an. Le pilotage opérationnel du référentiel est effectué par ce comité qui décide des priorités pour les futures activités.
- Le comité éditorial qui est composé de tous les fournisseurs de données. Ils sont responsables d'un ou plusieurs taxons, thèmes ou régions qu'ils prennent sous leur expertise.

- L'équipe de gestion des données qui est responsable du fonctionnement de la base de données en ligne, de l'intégrité des identifiants uniques et de l'assistance aux éditeurs et aux utilisateurs.

La gouvernance des référentiels du SIMM doit garantir que les besoins impératifs liés au SIMM soient pris en compte dans les délais compatibles à son bon fonctionnement. Or, le WoRMS est un référentiel international où les besoins du SIMM pourraient être remontés aux instances de pilotage WoRMS, mais une fois intégrés aux autres besoins exprimés par les autres contributeurs leur prise en compte ne peut être garantie.

TAXREF

TAXREF ne répond pas à ce besoin du SIMM car peu d'informations existent sur la gouvernance de ce référentiel, et celle-ci ne semble pas convenir aux besoins du SIMM. Les règles de gestion de TAXREF reposent sur des liens avec les groupes d'experts. En effet, l'expertise de TAXREF s'appuie sur de nombreux spécialistes, qui collaborent soit directement en effectuant eux-mêmes les modifications, soit en assurant une veille scientifique et en signalant les sources bibliographiques à prendre en compte. La gouvernance de TAXREF semble varier suivant ces groupes et s'organise sous la forme de conventions et de Comité de Pilotages. Toutefois, la concertation avec les utilisateurs ne semble pas être assez développée. Le dernier séminaire avec les utilisateurs remonte à 2016, il n'y a donc pas de réelles instances planifiées de manière régulière durant lesquelles les utilisateurs peuvent s'exprimer.

SANDRE

Le SANDRE répond à ce besoin du SIMM car sa gouvernance est clairement décrite au sein du Schéma National des Données sur l'Eau, et un document d'administration du référentiel taxonomique est téléchargeable librement sur son portail. La production de chaque référentiel SANDRE est associée à un groupe de travail intégrant notamment des utilisateurs de celui-ci. En pratique, il n'y a un groupe de travail que quand il y a un besoin spécifique d'administration du référentiel. N'ayant plus de besoin spécifique d'administration du référentiel d'appellation taxonomique, ce groupe ne se réunit plus. Toutefois, les utilisateurs peuvent toujours laisser des avis en ligne sur le référentiel et les contenus liés à celui-ci sur le site du SANDRE.

Par ailleurs, les organismes contribuant aux SIE sont représentés au sein des groupes des Administrateurs De Données (ADDs) qui remontent tout besoin utilisateur lié aux référentiels. Les réunions ADDs ont lieu environ tous les trimestres. Le Groupe de Pilotage du SANDRE (GPS) se réunit trois fois par an et inclut déjà plusieurs membres des instances de gouvernance du SIMM (notamment l'AFB et l'Ifremer) et du SIB.

ASFIS

Le référentiel ASFIS ne répond pas à ce besoin du SIMM car aucune information sur sa gouvernance n'est accessible. Il ne paraît pas y avoir d'instances ou d'outils permettant aux utilisateurs de faire remonter leurs avis sur le référentiel. De plus, comme pour WoRMS, le caractère international de ce référentiel rend peu probable la prise en compte des besoins du SIMM par ses instances de pilotage.

- **Accessibilité du référentiel**

WoRMS

WoRMS répond partiellement à ce besoin du SIMM car l'accès au référentiel se fait de manière libre mais la sémantique des données n'est pas accessible. L'accès à la majeure partie du référentiel ne nécessite pas d'identifiants : accès aux mises à jour en continue en ligne, outil Match_Taxa et web-services sous la forme d'API. Seul le téléchargement mensuel des mises à jour n'est possible que sous accès connecté.

TAXREF

TAXREF répond partiellement à ce besoin du SIMM car les versions annuelles du référentiel et les web-services sous la forme d'API sont accessibles à tous. La sémantique du référentiel est accessible sur le document de présentation des données lié à chaque version annuelle du référentiel. Toutefois, toute la version de travail, ainsi que l'outil de réconciliation TAXREF_MATCH, ne sont accessibles que via TAXREF-web nécessitant une authentification. En effet, l'application TAXREF-web a initialement été créée afin de faciliter le travail des experts. Toutefois l'objectif est de l'ouvrir à tous à l'image de WoRMS. L'accès aux données et aux outils du référentiel pour le grand public n'est donc pas total pour le moment.

SANDRE

Le référentiel du SANDRE répond à ce besoin du SIMM car les données du référentiel du SANDRE sont diffusées comme des données ouvertes (Open data) sans restriction d'usage, sous licence ouverte, accessibles à tous sur le site du SANDRE. Elles sont téléchargeables en différents formats (xml, csv, json). Les différents outils proposés sont également accessibles à tous, notamment des API. La sémantique des données est formalisée dans un dictionnaire de données propre au référentiel. Un scénario d'échange permet également de définir le format et les règles d'échange de certaines données.

ASFIS

Le référentiel ASFIS répond partiellement à ce besoin du SIMM car le téléchargement des données est accessible librement en ligne (formats txt et xls) mais aucun outil n'est proposé. La sémantique des données est formalisée dans un document présentant la structure de la base de données.

- **Codification de tous les taxons**

WoRMS

WoRMS ne répond pas à ce besoin du SIMM car il s'agit d'un référentiel scientifique, qui n'a pas pour vocation d'être opérationnel. Il ne permet que difficilement aux utilisateurs de faire des demandes d'ajout de nouveaux taxons (à cette adresse : info@marinespecies.org). De plus, l'ajout des taxons se fait en fonction des contributions volontaires des experts et peut donc être très inégale en fonction des branches de l'arbre taxonomique. Tous les taxons sur lesquels les banques ont de la donnée ne seront donc pas forcément codifiés sur WoRMS (notamment pour les taxons récemment découverts, sur BIGOOD par exemple) ce qui posera un problème d'interopérabilité car ces données ne seront pas échangeables.

TAXREF

TAXREF ne répond pas à ce besoin du SIMM car il s'agit d'un référentiel scientifique, qui n'a pas pour vocation d'être opérationnel. Il ne gère que les taxons présents de manière certaine en France, ayant fait l'objet d'une publication scientifique de rang A. Cela pose donc un problème pour les banques possédant des données sur les espèces exotiques envahissantes ou de passage, ou pour le reporting (pour le critère concernant les espèces exotiques envahissantes de la DCSMM par exemple).

SANDRE

Le SANDRE répond à ce besoin du SIMM. L'objectif du SANDRE, en tant que référentiel d'usage, n'est pas d'être exhaustif mais de codifier les taxons sur lesquels portent de la donnée. Le référentiel taxonomique du SANDRE étant lié au SIE, la couverture thématique est pour l'instant limitée au niveau des taxons marins aux données acquises dans le cadre de la Directive Cadre sur l'Eau. L'équipe du SANDRE ajoute de nouveaux taxons au fur et à mesure des demandes des utilisateurs, quel que soit le taxon demandé. Bien que se basant sur TAXREF pour référence, il est possible d'ajouter des taxons ou des groupes de taxons non valides scientifiquement ou non connus pour vivre en France, si la gestion des données l'impose. L'ouverture du référentiel à d'autres parties de la classification et la codification de tous les taxons utiles pour le SIMM est totalement possible. A titre d'exemple, 100% des taxons portant des données dans Quadriga sont codifiés SANDRE.

ASFIS

Le référentiel ASFIS ne répond pas à ce besoin du SIMM car il ne gère que les espèces exploitées ou capturées accidentellement par la flotte de pêche mondiale.

- **Gestion de groupes de taxons**

WoRMS

WoRMS ne répond pas à ce besoin du SIMM car il ne couvre pas le concept de regroupement de taxons.

TAXREF

TAXREF répond partiellement à ce besoin du SIMM puisqu'il permet la gestion de certains groupes de taxons. Toutefois, leur création répond à des besoins très spécifiques, ce qui ne permettra pas au référentiel de s'adapter à toutes les demandes des utilisateurs du SIMM, notamment pour des groupes de taxons ne reposant pas sur des bases scientifiques mais permettant de faciliter les suivis ou le traitement des données par exemple.

TAXREF gère des « groupes opérationnels » qui correspondent à des groupes biologiques renfermant un ensemble de taxons liés entre eux par une ou plusieurs relations écologiques, systématiques et/ou sociétales (ex: espèces marines introduites en métropole).

Ces groupes sont consultables sur TAXREF-web et téléchargeables dans le web-service grâce à une API.

Dans le document méthodologique de la version 12 de TAXREF (2018)³, il est cité : « *En conséquence, TAXREF refuse (...) les groupes dits fonctionnels, c'est-à-dire des groupes pratiques pour identifier un ensemble d'espèces mais pas nécessairement clairement définis taxonomiquement et non publiés validement (...).* » Toutefois, l'équipe de l'UMS Patrinat travaille sur l'élaboration de « groupes informels », afin de rattacher des données imprécises qu'il n'est pas possible de rattacher à une espèce ou un genre donné pour cause de difficultés de détermination (par exemple "poissons rouges"). Une réflexion est donc en cours afin de formaliser la façon de nommer ces groupes et de contextualiser leur contenu afin qu'ils soient exploitables dans le temps.

SANDRE

Le SANDRE répond à ce besoin du SIMM car il permet les regroupements de taxons en fonction des besoins des utilisateurs, quels qu'ils soient. Ces regroupements sont gérés via un jeu de données de référence appelé « groupes d'appellations de taxon », qui est lié au jeu de données « appellations de taxon ». Ils peuvent contenir plusieurs sous-groupes, selon un ordre hiérarchique pouvant aller jusqu'à plusieurs niveaux inférieurs, et peuvent être rattachés à plusieurs groupes d'appellation de taxon parents.

Les groupes d'appellations de taxon du SANDRE sont élaborés suite aux demandes de codification d'utilisateurs, selon des critères complexes et multiples. Ces critères sont en rapport avec la finalité pour laquelle un regroupement de taxons est défini.

ASFIS

Le référentiel ASFIS répond partiellement à ce besoin du SIMM car il est en mesure de gérer des groupes de taxons mais seulement pour les assemblages d'espèces commerciales et pour les espèces pour lesquelles une détermination précise du taxon n'est pas possible par des non-experts.

³ Gargominy, O., Tercerie, S., Régnier, C., Ramage, T., Dupont, P., Daszkiewicz, P. & Poncet, L. 2018. TAXREF v12, référentiel taxonomique pour la France : méthodologie, mise en oeuvre et diffusion. Muséum national d'Histoire naturelle, Paris. Rapport Patrinat 2018-117. 156 pp.

- **Transcodage avec d'autres référentiels**

WoRMS

Worms ne répond pas à ce besoin du SIMM car il ne fournit pas de correspondances avec d'autres référentiels taxonomiques.

TAXREF

TAXREF répond partiellement à ce besoin du SIMM. Il diffuse des correspondances avec d'autres référentiels taxonomiques dans la table TAXREF_LIENS téléchargeable avec le référentiel TAXREF dans la version publiée officiellement. Les équivalences avec les taxons présents seulement dans la version de travail ne sont donc pas accessibles librement. Un outil de réconciliation TAXREF_MATCH permet d'attribuer des codes TAXREF (CD_NOM) à une liste de noms d'espèces afin de retrouver les noms validés par le référentiel et le lien éventuel avec WoRMS, mais celui-ci nécessite une authentification. Malgré la possibilité de fournir des identifiants aux coordinateurs des banques du SIMM, cet outil n'est pas adaptable aux besoins du SIMM car les services des référentiels du SIMM doivent être en accès libre.

SANDRE

Le référentiel du SANDRE répond à ce besoin du SIMM car il intègre les codes de plusieurs référentiels taxonomiques (dont TAXREF, WoRMS et même les codes internes à certaines banques comme Quadrigé) à sa base de données en tant que codes alternatifs. Ces correspondances sont diffusées auprès des utilisateurs via un tableau des correspondances des codes taxonomiques (en format csv et libre accès), ainsi que par un service de convertisseur de codes. De plus, l'équipe du SANDRE est en mesure d'intégrer les codes d'autres référentiels.

ASFIS

Le référentiel ASFIS ne répond pas à ce besoin du SIMM car il ne fournit pas de correspondances avec des référentiels autres que ceux de la FAO.

- **Identification unique des taxons**

WoRMS

WoRMS répond à ce besoin du SIMM car les niveaux taxonomiques sont inclus dans son modèle de données.

TAXREF

TAXREF répond à ce besoin du SIMM car les niveaux taxonomiques sont inclus dans son modèle de données.

SANDRE

Le référentiel du SANDRE répond à ce besoin du SIMM car les niveaux taxonomiques sont inclus dans son modèle de données.

ASFIS

Le référentiel ASFIS répond à ce besoin du SIMM car les niveaux taxonomiques sont inclus dans son modèle de données.

- **Réactivité des mises à jour**

WoRMS

WoRMS ne répond pas à ce besoin du SIMM car il ne permet que difficilement de faire des demandes d'ajout de nouveaux taxons. L'ajout et la modification de taxons dépend des contributions volontaires des experts. Même si globalement les experts sont très réactifs, le temps de réponse aux demandes envoyées varie selon les groupes taxonomiques. Les mises à jour sont ensuite effectuées en continu dans les pages WoRMS en ligne et de manière mensuelle pour la version en téléchargement sous accès connecté.

TAXREF

TAXREF répond partiellement à ce besoin du SIMM car les taxons sont ajoutés au gré des demandes des utilisateurs mais seulement dans la version de travail, nécessitant une authentification pour y accéder. En effet, TAXREF est très réactif aux demandes de codification de nouveaux taxons, qui peuvent être effectués dans la journée si elles sont accompagnées d'une justification pour sa création. Toutefois, ceux-ci ne sont accessibles que via TAXREF-web entre deux versions annuelles. Malgré la possibilité de fournir des identifiants aux coordinateurs des banques du SIMM, cet outil n'est donc pas adapté aux besoins du SIMM car les données des référentiels du SIMM doivent être en accès libre. La réelle diffusion des données n'est qu'annuelle ce qui limite la fréquence de mise à jour via des outils automatisés. En attendant les nouvelles mises à jour, les utilisateurs du SIMM devront donc créer des identifiants provisoires pour stocker leurs données, ce qui est compliqué à gérer et entraîne des problèmes d'interopérabilité.

TAXREF est un référentiel scientifique, qui n'a pas pour vocation d'être opérationnel. Tout ajout ou modification d'un taxon doit reposer sur des publications scientifiques. TAXREF ne répond donc pas aux besoins du SIMM en terme de réactivité des mises à jour dans un temps court puisqu'il nécessite une recherche bibliographique. En effet, les demandes d'ajout d'un nouveau taxon se font par mail ou via la rubrique contact du site de l'INPN. Chaque demande doit être appuyée par une référence : description d'espèce valide respectant les codes de nomenclature dans le cas d'un ajout d'espèce, publication scientifique pour les changements de synonymie ou de classification. D'une part la recherche des articles scientifiques nécessaires aux ajouts dans TAXREF nécessite des moyens humains que les producteurs de données n'ont pas nécessairement, et d'autre part, il se peut, notamment pour les nouvelles espèces, qu'aucune référence n'ait encore été publiée. Cela pose alors un problème pour les banques stockant des données sur des espèces non encore décrites, comme BIGOOD par exemple.

Remarque : dans le cadre des échanges de données entre Quadriga et le SINP, des expertises taxonomiques ont été réalisées par le MNHN *a posteriori* de la bancarisation. Les corrections identifiées par le MNHN ont pu être prises en compte dans la banque Quadriga avec l'accord des producteurs de données. Ce fonctionnement de l'expertise taxonomique s'est révélé très efficace.

SANDRE

Le SANDRE répond à ce besoin du SIMM car il respecte un délai de réponse de dix jours ouvrés aux demandes de création, mise à jour et gel sur le référentiel. Celles-ci sont traitées via un outil d'administration des demandes d'évolution des référentiels SANDRE (le MDM du SANDRE).

ASFIS

Le référentiel ASFIS ne répond pas à ce besoin du SIMM car sa fréquence de mise à jour est annuelle (en février-mars de chaque année). Actuellement il s'agit de la version de février 2018. Lorsqu'une expertise documentée est fournie, elle est analysée et validée le cas échéant par la FAO pour être prise en compte lors de la mise à jour annuelle suivante.

- **Gestion des identifiants**

WoRMS

Le travail collaboratif d'état des lieux effectué en amont de la création de ce document n'a pas permis d'obtenir des informations sur la gestion des identifiants par WoRMS.

TAXREF

TAXREF ne répond pas parfaitement à ce besoin du SIMM qui nécessite un référentiel taxonomique avec des identifiants non supprimables et persistants. En effet, le guide méthodologique précise que les identifiants utilisés dans TAXREF (CD_NOM) peuvent disparaître. Ceux-ci sont alors stockés dans un tableau listant les identifiants qui ne sont plus diffusés dans la dernière version et les précédentes (CDNOM_DISAPRUS.xls). L'utilisation de deux fichiers différents rend l'utilisation opérationnelle compliquée, en particulier lors de contrôle de cohérence de données métier avec le référentiel.

SANDRE

Le SANDRE répond à ce besoin du SIMM car il suit les règles du Cadre Commun d'Architecture des Référentiel de données dans la gestion de ses identifiants. La manière dont sont gérés les identifiants est décrite dans le « Guide – Les référentiels du Système d'information sur l'eau - Règles, contraintes et services ».

ASFIS

Le référentiel ASFIS ne répond pas à ce besoin du SIMM car la manière dont sont gérés ses identifiants alpha-3 ne suit pas totalement les règles du Cadre Commun d'Architecture des Référentiel de données. En effet, certains de ces codes sont signifiants car les premiers codes produits faisaient référence aux noms scientifiques ou anglais des catégories d'espèces.

Synthèse du croisement entre les besoins du SIMM et les caractéristiques des référentiels

Besoins du SIMM	Référentiels d'usage		Référentiels scientifiques	
	SANDRE	ASFIS	TAXREF	WoRMS
• Gouvernance du référentiel	Répond au besoin	Ne répond pas au besoin	Ne répond pas au besoin	Ne répond pas au besoin
• Accessibilité du référentiel	Répond au besoin	Répond partiellement au besoin	Répond partiellement au besoin	Répond partiellement au besoin
• Codification de tous les taxons	Répond au besoin	Ne répond pas au besoin	Ne répond pas au besoin	Ne répond pas au besoin
• Gestion des groupes de taxons	Répond au besoin	Répond partiellement au besoin	Répond partiellement au besoin	Ne répond pas au besoin
• Transcodage avec d'autres référentiels	Répond au besoin	Ne répond pas au besoin	Répond partiellement au besoin	Ne répond pas au besoin
• Identification unique des taxons	Répond au besoin	Répond au besoin	Répond au besoin	Répond au besoin
• Réactivité des mises à jour	Répond au besoin	Ne répond pas au besoin	Répond partiellement au besoin	Ne répond pas au besoin
• Gestion des identifiants	Répond au besoin	Ne répond pas au besoin	Répond partiellement au besoin	?

3.2. Conclusion et proposition du SAR pour le SIMM

Référentiel d'usage

Le référentiel ASFIS est trop limité au domaine d'utilisation que sont les statistiques de pêche pour répondre aux besoins du SIMM et des banques de données. Toutefois, les identifiants ASFIS sont indispensables et devront être intégrés au référentiel taxonomique du SIMM.

Le référentiel d'appellation taxonomique du SANDRE est le référentiel d'usage le plus adapté au SIMM. Il est le plus proche des besoins exprimés par le SIMM en termes de gouvernance et d'adaptabilité. Il est possible de l'utiliser dans sa forme actuelle car il propose déjà tous les services nécessaires au SIMM et à ses utilisateurs. De plus, il permet d'assurer l'interopérabilité avec le SIE qui est un des deux autres Systèmes d'Informations fédérateurs. Le groupe de pilotage du SANDRE doit évidemment valider le fait que le SANDRE prenne en charge l'ouverture de son référentiel à d'autres parties de la classification et l'intégration des codes d'ASFIS en tant que codes alternatifs dans sa base de données, ainsi que de compléter son référentiel avec les taxons utiles au SIMM.

Référentiel scientifique

TAXREF est le référentiel scientifique le plus adapté aux besoins du SIMM et permettant également l'interopérabilité avec le SNP. Le référentiel taxonomique du SIMM doit donc reposer sur ce référentiel scientifique. Toutefois, il est également important que le référentiel retenu intègre les identifiants WoRMS afin d'assurer l'interopérabilité des données du SIMM avec l'international (dans le cadre du rapportage pour la DCSMM par exemple ou pour la prise des espèces invasives marines).

Le référentiel d'appellation taxonomique du SANDRE utilise TAXREF comme référence scientifique pour l'identification des taxons. A défaut, dans le cas d'un taxon marin non encore décrit dans TAXREF, le SANDRE se base alors sur WoRMS dont il intègre les codes dans sa base de données. Le SANDRE en se basant prioritairement sur l'expertise de TAXREF répond à ce dernier besoin du SIMM relatif à l'utilisation de TAXREF comme référentiel scientifique du SIMM.

Le Service d'Administration des Référentiels du SIMM propose donc au Groupe de Travail Langage Commun l'utilisation du référentiel d'appellation taxonomique du SANDRE comme référentiel taxonomique pour le SIMM.

Glossaire

AFB : Agence Française pour la Biodiversité

AMP : Aire Marine Protégée

API : Interface de programmation

ASFIS : Aquatic Sciences and Fisheries Information System

CNRS : Centre national de la recherche scientifique

CSITAPA : Classification Statistique Internationale Type des Animaux et des Plantes
Aquatiques de la FAO

DCE : Directive Cadre sur l'Eau

DCPEM : Cadre pour la Planification de l'Espace Maritime

DCSMM : Directive cadre Stratégie pour le Milieu Marin.

FAO : Organisation des Nations Unies pour l'alimentation et l'agriculture

IFREMER : Institut français de recherche pour l'exploitation de la mer

INPN : Inventaire National du Patrimoine Naturel

MNHN : Muséum national d'Histoire naturelle

OBIS : Ocean Biogeographic Information System

OSPAR : Convention Oslo-Paris (Convention pour la protection du milieu marin de
l'Atlantique du Nord-Est)

RESOMAR : Réseau des Stations et Observatoires Marins

SANDRE : Service d'administration nationale des données et référentiels sur l'eau

SAR : Service d'Administration des Référentiels

SGMAP : Secrétariat général pour la modernisation de l'action publique

SIB : Système d'Information sur la Biodiversité

SIE : Système d'Information sur l'Eau

SIH : Système d'Informations Halieutiques

SIMM : Système d'Information sur le Milieu Marin

SINP : Système d'Information sur la Nature et les Paysages

SNDMM : Schéma National des Données sur le Milieu Marin

TIC : Technologie de l'information et de la communication

UICN : Union internationale pour la conservation de la nature

UMS Patrimoine : Unité Mixte de Service Patrimoine naturel

VLIZ : Vlaams Instituut voor de Zee (Institut flamand de la mer)

WoRMS : World Register of Marine Species

Partenaires du SAR ayant apporté leur contribution à cet état des lieux

NOM	Prénom	Organisme	Courriel	Lieu	Thématiques, mots-clefs
BORREMANS	Catherine	Ifremer	catherine.borremans@ifremer.fr	Brest	environnements profonds, BIGOOD
COUDERCY	Laurent	AFB	laurent.coudercy@afbiodiversite.fr	Vincennes/ Orléans	
DE MAZIERES DE CHAMBON	Jeanne	UMS Patrinat	jdemazieres@mnhn.fr	Paris	Espèces marines
DUPONT	Pascal	UMS Patrinat	pascal.dupont@mnhn.fr	Paris	TAXREF, bases de connaissance, inventaires
GARGOMINY	Olivier	UMS Patrinat	gargo@mnhn.fr	Paris	TAXREF, bases de connaissance
GAUTHIER	Emilie	Ifremer	emilie.gauthier@ifremer.fr	Nantes	benthos, plancton, ressources aquacoles, microbio
GUERIN	Laurent	UMS Patrinat	laurent.guerin@mnhn.fr	Dinard	Espèces non indigènes (DCSMM)
HUGUET	Antoine	Ifremer	Antoine.Huguet@ifremer.fr	Nantes	Quadrige
LEVEUGLE	Florine	Office international de l'eau	f.leveugle@oieau.fr	Limoges	
MERRIEN	Claude	Ifremer	claud.merrien@ifremer.fr	Lorient	Halieutique
MEUNIER	Dimitri	Office international de l'eau	d.meunier@oieau.fr	Limoges	
ROTH	Benjamin	Office international de l'eau	b.roth@oieau.fr	Limoges	diatomées, macro-invertébrés, macrophytes
TERCERIE	Sandrine	UMS Patrinat	sandrine.tercerie@mnhn.fr		TAXREF, bases de connaissance
VAN ISEGHEM	Sylvie	Ifremer	sylvie.van.lseghem@ifremer.fr	Brest	environnements profonds, BIGOOD
VINCENT	Dorothee	AFB	dorothee.vincent@afbiodiversite.fr	Brest	Zooplancton et phytoplancton ; espèces zooplanctoniques exotiques envahissantes

