

VALPARAISO

Application de contrôle, validation et
qualification des données pour le Système
d'informations halieutiques

Guide Technique

A destination du développeur

R-Shiny



RBE/HISSEO
Thibault GAUDUCHON

Octobre 2022

Avant-propos

A qui s'adresse ce document

Ce document s'adresse au développeur en charge de la maintenance et de l'évolution de Valparaiso. Son objectif est de documenter les aspects techniques du fonctionnement de l'application, de restituer leur contexte de développement au moment de leur création ; de détailler les raisons de leur conception et les objectifs auxquels ils répondent ; d'exposer les contraintes rencontrées et identifier les faiblesses ou les points de vigilance à observer.

Plan et lecture du document

Certains concepts ou schéma sont peut-être difficiles d'accès en première lecture. Le but du présent document n'est pas d'être lu pour lui-même, mais d'aider à cerner les éléments et les concepts précis qui ont été développés ; il vise aussi à grader trace de tout ce qui, en temps utile (en seconde, troisième lecture...), sera nécessaire à la poursuite des développements. Il devra par conséquent être lu plusieurs fois et sans doute de façon non linéaire, avec le code source et les autres documents disponibles (guide utilisateur, présentations) pour prendre pleinement possession de l'application.

Introduction

Les objectifs principaux du projet Valparaiso étaient :

- Le contrôle et validation de la donnée brute des programmes de collecte et d'enquête du SIH
- La centralisation du processus de validation, avec :
 - Un projet et une application unique pour toutes les actions
 - Un accès centralisé et facilité aux données
 - Le partage d'une même technologie (R-Shiny) pour toutes les actions, qui toutes par ailleurs disposent d'outils de saisie ou de valorisation très différents
- Atteindre la transversalité des contrôles avec :
 - Des croisements nouveaux de données entre les actions ;
 - La mobilisation de toutes sources de données du SIH.
- Faire gagner en autonomie des saisisseurs et des validateurs pour le contrôle des données

Quatre actions de collecte de données du SIH ont été développées pour la création de Valparaiso :

- ACTIFLOT (Activité ou calendriers d'activité des navires de la flotte française)
- OBSDEB (Observation des marées au débarquement Outre-mer)
- OBSVENTES (Observations en criées)
- OBSBIO (Paramètres biologiques individuels et données d'âgeage)

Principaux aboutissements du projet pour sa création :

- Une chaîne de développement et de déploiement de l'application autonome, entièrement intégrée à l'infrastructure de l'Ifremer (Gitlab, plateforme de développement et serveur de production de l'application) ;
- Une application en accès Internet pour la production ;
- Une application directement connectée à la base de données Harmonie ;
- La création d'un schéma Harmonie dédié à la validation des données ;
- Les champs de validation de Valparaiso relayés jusqu'aux données de saisie brutes d'Harmonie ;
- Le respect des droits utilisateurs centralisés dans Harmonie (utilisateur Ifremer ou prestataire ; menus d'actions autorisées, portefeuille navires) ;
- La création concomitante d'un entrepôt de données pour le SIH ;

Maintenance

Le projet n'a pas la prétention d'être exempt de toute faille. La très grande diversité et complexité des données qui transitent par l'application est plutôt un facteur de fragilité.

Valparaiso dispose de garde-fous pour éviter de planter l'application en cas d'imprévus ou d'évolutions délétères, comme la vérification systématique de la présence des fichiers à charger dans l'entrepôt de données par exemple. Mais comme pour tout projet, des situations nouvelles, non anticipées, peuvent surgir des données et perturber la chaîne de traitements. Une intervention sur l'application sera alors nécessaire.

Perspectives de développement

Les principales évolutions envisagées et perspectives identifiées pour l'application Valparaiso sont :

- Augmenter le portefeuille d'actions avec OBSMER et pourquoi pas des flux régaliens ou des produits Harmonie.
- Poursuivre le développement de modules transversaux. Les perspectives de développement de chaque action, lorsque celles-ci ont pu être identifiées, sont l'objet d'une partie spécifique à la fin du chapitre dédié à l'action.
- Reprendre et étendre le processus de qualification expérimenté pour OBSVENTES à d'autres actions (SACROIS, Activité des navires...).
- Utiliser la base de connaissance comme base de suivi et de mise à jour des référentiels RTP du SIH ; dans Valparaiso ou une autre application.

Corpus complet de documentation

Le présent document

La vignette du package : documentation interne du package avec citations de code, et exemples d'utilisation des fonctions. Utile pour identifier à quel niveau du serveur ou d'autres fonctions R intervenir. « (cf. #Vignette) » renvoie le lecteur à ce document.

La vignette aurait mérité bien d'avantage de chapitres. Elle donne quelques pistes supplémentaires pour la prise en main et le paramétrage des modules de base de connaissance et sur les données géographiques intégrées au projet.

Aides des fonctions du package : documentation interne du package avec indexation des lignes de code.

Commentaires du code : « (cf. #Commentaires) » indique au lecteur qu'il trouvera dans les commentaires du fichier ou du script des commentaires suffisamment explicites et précis pour prendre le relai ou compléter ce document.

Sommaire

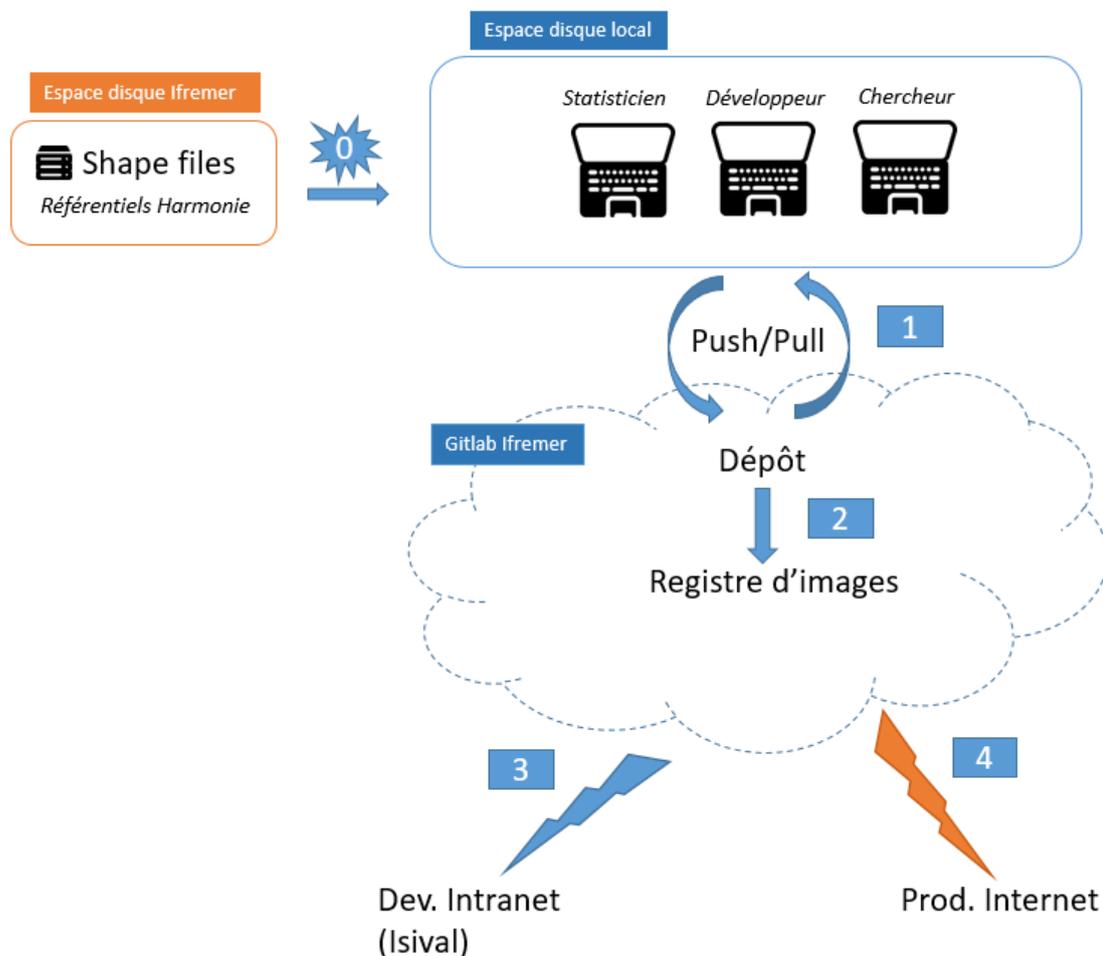
I. Valparaiso à l’Ifremer.....	8
I.1 Schéma général de développement et de déploiement de Valparaiso.....	8
I.2 Installation en condition de développement.....	9
I.3 Structure d’ensemble du projet.....	10
I.4 Intégration et mise à jour des couches géographiques de référence	11
• Requête des Ports – Criée – Quartier maritime (points) :	13
• Requête des zones en mer et toutes les autres géométries (polygones) :.....	13
I.5 Gestion des dépendances R : renv.....	13
I.6 Production d’une nouvelle version	14
I.7 Déploiement sur la plateforme de développements – accès Intranet.....	15
I.8 Déploiement en production – accès Internet	15
II. Base de connaissance et de qualification dans Harmonie.....	17
II.1 Usages de la base de connaissance.....	17
II.2 Création d’une nouvelle table dans la base de connaissance	18
II.3 Champs et groupes de colonnes	19
II.4 Principe d’historicité	21
II.4.1 Traitement et filtrage des tables de connaissances	22
II.5 Connexion Valparaiso - Ecriture dans la base de connaissance	23
III. Données de Valparaiso.....	24
III.1 Approvisionnement en données d’une session Valparaiso	24
III.2 Entrepôt de données du SIH	24
III.3 Système Harmonie	26
III.4 Gitlab Ifremer	26
III.5 Mise à jour des données	27
IV. Logique commune de chargement et de traitement des données	28
V. Navigation multi-actions - Modules et objets transversaux.....	29
V.1 Chargement de plusieurs actions.....	29
V.2 Module « Navire-armateur »	30
V.3 Module « Population » : filtrage transversal des données.....	31
V.4 Schéma général d’interaction entre le module de filtrage « Population », « Navire-armateur » et les modules spécifiques des actions	33

VI.	Activité des navires	34
VI.1	Jeux de données	34
VI.2	Contrôles réalisés	35
VI.3	Base connaissance pour l'enquête Activité	36
VI.4	Sélection d'un navire et navigation croisée OBSDEB	36
VI.5	Visualisation du calendrier d'activité complet.....	36
VI.6	Historique de l'activité d'un navire et changement d'armement	37
VI.7	Cartographie de l'activité des navires : l'indicateur « mois-navire ».....	37
VI.8	Cartographie de l'activité d'un navire	39
VI.9	Perspectives 48G30	39
VII.	OBSDEB	40
VII.1	Contexte	40
VII.2	Jeux de données	40
VII.3	Réalisation	40
VIII.	OBSVENTES	41
VIII.1	Préambule.....	41
VIII.2	Jeux de données	41
VIII.3	Traitements.....	43
VIII.3.1	Traitement du référentiel RTP.....	43
VIII.3.2	Transcodage des espèces	43
VIII.3.3	Contacts-espèce WAO & référentiel PETS.....	44
VIII.3.4	Données OBSVENTES.....	44
VIII.3.5	Mélange d'espèces	46
VIII.4	Contrôles des lots « Espèce-navire ».....	46
VIII.5	Qualification des lots échantillonnés	48
VIII.5.1	Schéma complet du cheminement des données	48
VIII.5.2	Base de connaissance et de qualification.....	50
VIII.5.3	Règle de qualification	51
VIII.5.4	Transfert des champs de qualification et de commentaire Programme vers les données brutes Harmonie	51
VIII.6	Contrôles et qualification des sorties	52
VIII.7	Perspectives d'évolution	52
IX.	OBSBIO	54
IX.1	Jeux de données	54

IX.2	Ciblage des espèces et échantillons « OBSBIO ».....	55
IX.3	Chargement et traitement des données.....	55
IX.4	Contrôles visuels des variable Taille, Poids, Âge.....	57
IX.5	Sélection et « suppression » d'échantillon	58
IX.6	Inter-calibration d'âges	58
X.	Annexes	59
X.1	Acronymes.....	59

I. Valparaiso à l’Ifremer

I.1 Schéma général de développement et de déploiement de Valparaiso



0 Intégration/Mise à jour des couches géographiques de référence du SIH au projet Valparaiso.

1 Synchronisation des développements du code et amendements aux tables versionnées au projet Gitlab.

2 Production d’une nouvelle image Docker à partir du dépôt Gitlab.

3 Déploiement d’une image Docker sur la plateforme de développement Isival (Intranet).

4 Déploiement d’une image Docker en production (Internet).

■ Périmètre du/des développeurs

■ Périmètre d’autres acteurs

I.2 Installation en condition de développement

Espaces disques réseau

Le développeur veillera à accéder, en lecture, aux espaces disque suivant :

- Entrepôt de données du SIH (production) :

« //iota1/sih_donnees/espace/valparaiso »

- Espace de stockage des Shapefiles de référence :

« //iota1/sih_donnees/referentiels/Shapes »

Accès au système Harmonie

Le développeur aura besoin d'un outil de gestion de base de données (type DB ou autre) afin de pouvoir élaborer de nouvelles requêtes SQL, créer une nouvelle table de connaissance sur le schéma SIH2_VALPARAISO_DBA du système Harmonie. Le schéma Valparaiso est présent sur les trois services d'Harmonie :

- HARMONIE : service de production ;
- HARMONIE_VAL : service de validation ;
- HARMONIE_TEST : service de tests et de développement.

Pour toute question ou assistance concernant l'installation et l'accès à l'ensemble des outils plateforme, espace disque et plateformes nécessaire au développement de Valparaiso, se rapprocher de l'équipe IRSI/ISI de l'Ifremer (G. PRIGENT).

Projets Gitlab

Valparaiso :

<https://gitlab.ifremer.fr/sih/sih-data-analysis/valparaiso>

Référentiels OBSDEB :

https://gitlab.ifremer.fr/sih_hisseo/reftabs_sihobsdeb

Plateforme de développement Isival (fichier de configuration) :

https://gitlab.ifremer.fr/-/ide/project/dev-ops/isival/isival_application/edit/master/-/group_vars/all/vars.yml

Autres outils et liens

Mantis (demande d'évolution de versions en production):

https://forge.ifremer.fr/mantis/view_all_bug_page.php

Marche à suivre

La marche à suivre pour tout nouveau déploiement en développement comme en production est documentée sur la page d'accueil Gitlab du projet Valparaiso ainsi que dans les sections suivantes de ce chapitre.

I.3 Structure d'ensemble du projet

Ce paragraphe décrit la structuration de l'ensemble des fichiers nécessaires au développement de l'application et à son déploiement.

Les spécificités de l'application seront détaillées ; mais les principes généraux d'utilisation de Docker, Gitlab ou même de R et ou de Shiny ne seront pas développés outre mesure.

Une application dans un package ; un package dans une image Docker

Pour pouvoir déployer une application R-Shiny sur l'infrastructure de l'Ifremer, celle-ci doit être intégrée à un package R standard qui lui-même pourra être installé au côté de R dans une image Docker. Ainsi : « déployer » l'application signifie dans le détail : 1. Tourner l'image Docker ; 2. Lancer R ; 3. Charger le package valparaiso ; 4. Lancer l'application.

Raccourci sémantique : application = package = image

Dans la mesure où le package R du projet n'a d'autres fins que pouvoir lancer l'application Shiny qu'il contient, on assimilera souvent les deux choses : le package et l'application. De même, on assimilera l'image Docker et l'application dans la mesure où celle-ci n'a pas d'autre intérêt et fonction que de pouvoir transporter et déployer l'application.

Dans le processus de développement, c'est bien la nouvelle image Docker produite sur Gitlab qui sera taguée avec le numéro de version. L'image taguée est consignée au registre du projet Valparaiso dans « Container Registry » ; elle pourra ainsi être identifiée et récupérée par l'équipe en charge du déploiement en production.

Afin de garantir cette unité sémantique (application, package, image) et d'éviter toute confusion dans le numéro des versions : on veillera, avant de produire une nouvelle image de Valparaiso, à bien faire correspondre la version du package dans le fichier DESCRIPTION et le tag de la version de l'image sur Gitlab (cf. Production d'une nouvelle version).

L'Architecture d'un package R

La racine du projet Valparaiso est conforme à l'architecture d'un package R classique ; on y trouve les standards :

- Dossier R : contenant l'ensemble des fonctions R;
- Fichier DESCRIPTION du package, avec notamment les dépendances nécessaires;
- Fichier LICENCE.

Dans cette architecture, l'application Shiny correspond à un ensemble de fonctions R spécifiques, en particulier « server » et « ui ». Le projet Valparaiso a été initialisé à partir du package R golem, dont on trouvera trace dans l'application.

Supplément Données internes

Le dossier R de Valparaiso contient un fichier de données internes standardisé : « sysdata.rda », dans lequel se trouvent les couches géographiques de référence du SIH au format R (voir plus loin : « Intégration et mise à jour des couches géographiques de référence »).

Supplément : Données brutes

Le dossier « ./data-raw » dossier contient des tables de pilotage pour certaines actions (Cf. OBSVENTES, Référentiels externes). Ce dossier ne participe pas à la construction du package et sert simplement de dépôt, d'emplacement pour ces tables.

Les tables de référence externes de ce dossier sont lues en direct par l'application via une **connexion au projet Gitlab** de l'application.

Supplément Fichiers installés

Le dossier « ./inst » contient les autres langages, notamment dans le cadre de Valparaiso, un fichier nécessaire à la connexion OJDBC Valparaiso-Harmonie.

Les autres fichiers/dossiers structurants à la racine du projet

- `renv.lock` : configuration des packages R à installer pour l'image Docker
- `renv/` : environnement local séparé pour le développement du projet ; librairies R utilisées
- `NAMESPACE` : index des imports package nécessaires et des fonctions propres du package à exporter
- `README.md` : Informations et instructions générales relatives au projet, visibles directement sur la page d'accueil Gitlab
- `Dockerfile` : fichier de commandes pour la constitution d'une image Docker
- `gitlab-ci.yml` : configuration de l'intégration continue « CI » pour le projet ; en particulier : commande machine de construction d'une nouvelle image, lancées automatiquement avec la création d'un nouveau « Tag » au projet (avec : `tag = version`)
- `valparaiso.Rproj` : fichier à ouvrir pour lancer le projet dans RStudio
- `.dockerignore` : répertorier les fichiers/dossiers à ignorer pour la construction d'une image
- `Vignette` : dossier pour la vignette (documentation) du package

I.4 Intégration et mise à jour des couches géographiques de référence

Les couches géographiques de référence du SIH sont directement intégrées au package sous forme de **données internes**.

Avantages/inconvénients Ceci permet de travailler à façon les modules cartographiques directement dans R. Les données internes sont directement disponibles dans les sessions des utilisateurs sans avoir à être importées à chaque ouverture de session. L'intégration sous forme de données internes est par contre nécessairement manuelle et ponctuelle. Il s'agit cependant de données relativement stables dans le temps.

Pour en savoir plus sur la mise à jour des géométries de référence, se rapprocher de la cellule d'Exploitation Harmonie et de la personne en charge des référentiels (A. ROUYER).

Marche à suivre actuelle pour mettre à jour les géométries dans Valparaiso

1. Les fichiers d'origine (shapefiles) sont dans un premier temps passer au format de données R.

Emplacement des fichiers d'origine :

```
\\iota1\sih\donnees\referentiel\Shapes
```

Les fichiers de données géographique au format R sont produits par le script suivant (cf. #Commentaire) :

```
\\iota1\sih\_donnees\espace\valparaiso\REFERENTIELS\SHAPES\shp2rda.R
```

Une copie de ce script est présente dans le dossier « ./dev » du projet (par sécurité).

Les fichiers produits sont déposés dans le même dépôt « SHAPES » de l'entrepôt de données SIH, dans le dossiers RDATA.

2. Les fichiers au format R sont dans un second temps chargés et compilés dans un fichier unique du projet Valparaiso : « ./R/sysdata.rda », par le script du projet « ./data-raw/raw-to-internal.R ».

3. Le fichier de données internes produit, positionné automatiquement dans le dossier « ./R », doit être poussé sur Gitlab pour être intégré dans l'image Docker.

Attention : Evolution à venir de la gestion des géométries

A court terme, les fichiers d'origine susmentionnés ne seront plus mis à jour : les géométries seront directement mises à jour dans le système Harmonie par la cellule d'exploitation.

Plusieurs pistes restent à explorer pour adapter au mieux le projet Valparaiso à ce changement :

- Requêter les géométries directement depuis Harmonie, en remplacement de la lecture sur disque réseau actuelle, puis intégrer des données au package valparaiso comme actuellement.
- Demander un service d'export régulier des géométries dans l'Entrepôt de données ; lire et traiter ces fichiers pour les intégrer au package ensuite.
- Sinon demander des exports ponctuels des géométries.
- A moyen/long terme : profiter du service de cartographie en ligne de SEXTANT (pas encore en place ; mais peut-être plus contraignant que de la cartographie à façon directement dans R et donc peu adapté au projet).

Géométries dans le système Harmonie

Les géométries sont donc également présentes dans le système Harmonie et requêtables au format WKT grâce aux instructions suivantes :

- **Requête des Ports – Criée – Quartier maritime (points) :**

```
select
location.label,
location.name,
location.location_level_fk,
location_area.update_date,
sdo_util.to_wktgeometry(location_area.position) as wkt
from
location,
location_area
where location.id = location_area.id;
```

- **Requête des zones en mer et toutes les autres géométries (polygones) :**

```
select
location.label,
location.name,
location.location_level_fk,
location_point.update_date,
sdo_util.to_wktgeometry(location_point.position) as wkt
from
location,
location_point
where location.id = location_point.id ;
```

A toutes fins utiles pour exporter les géométries directement depuis le système Harmonie. Le package {rgeos} de R et sa fonction « readWKT() » permet de passer aux formats de données spatiales classique du paquet R {sp}.

I.5 Gestion des dépendances R : renv

Localement (chez le développeur) le projet Valparaiso utilise un environnement spécifique de développement « isolé » de l'environnement global de R où sont installés par défaut les packages. L'inconvénient de l'environnement par défaut est de mal maîtriser la version de chacun des packages. L'avantage d'un environnement spécifique et de pouvoir tester la compatibilité des mises à jour proposées des dépendances avec le projet avant de les inscrire (ou non) comme nouvelles versions à installer.

Le package R de constitution et de gestion de l'environnement de développement est : {renv}. L'utilisation de ce package conduit à la création à la racine du projet R, d'une part

du dossier « ./renv », qui contient les packages utilisés dans le cadre du projet ; et d'autre part du fichier « ./renv.lock » qui en fixe les versions dans un format de configuration.

L'ensemble des opérations classiques d'installation et mise à jour de packages dans l'environnement de développement s'effectue via des commandes spécifiques de renv (similaires aux commandes classiques) écrites dans le script « ./dev/run_renv.R » (cf. #Commentaires).

Le fichier « renv.lock » sera ensuite utilisé pour l'installation des dépendances de Valparaiso lors de la constitution d'une nouvelle image Docker (cf. Production d'une nouvelle version).

Seul le fichier « renv.lock » est utile (et même indispensable) sur le Gitlab du projet. Le dossier « ./renv » n'est utile que localement et ne doit pas être pousser sur Gitlab.

1.6 Production d'une nouvelle version

La constitution d'une nouvelle image Docker de Valparaiso est automatiquement déclenchée par la création d'un nouveau « Tag » dans le projet Gitlab. Le nom à donner à ce tag est le numéro de la version.

Création et maintenance de l'intégration continue (CI) pour Valparaiso : E. BODERE, G. PRIGENT ; Ifremer/IRSI.

Les instructions pour la constitution de l'image, ainsi que de la commande de lancement de l'application à partir de celle-ci, sont inscrites dans le fichier « Dockerfile ». Classiquement, ce fichier demande : l'installation de R, des dépendances de Valparaiso et enfin du package valparaiso lui-même. Il définit enfin la commande de lancement de R et de l'application contenue dans le package {valparaiso} installé dans l'image.

Des variables d'environnement sont également déclarées par défaut. En fonction du contexte (dev/prod) celle-ci peuvent être réécrites dans le fichier de configuration au moment du déploiement. Pour le détail et l'utilité de ces variables : cf. #Commentaires.

Instructions La marche à suivre pas à pas pour la constitution d'une nouvelle version de Valparaiso est également décrite dans le fichier README.md du projet (cf. Page d'accueil du projet sur Gitlab).

I.7 Déploiement sur la plateforme de développements – accès Intranet

Modifier le numéro de version du fichier de configuration de la plateforme Isival¹.

Les paramètres supplémentaires relatifs au déploiement de l'application peuvent également faire l'objet de modifications si nécessaire, en particulier :

- Variables d'environnement ;
- Nouveaux volumes ;
- Ressources allouées ;
- Autorisation de flux entrants.

La procédure de déploiement d'une nouvelle version sur Isival est rappelée dans le fichier README.md (cf. Page d'accueil du projet sur Gitlab).

I.8 Déploiement en production – accès Internet

Le déploiement en production d'une nouvelle version de Valparaiso (préalablement testée sur Isival) est toujours opéré par RIC.

Réaliser une demande de Mise en production

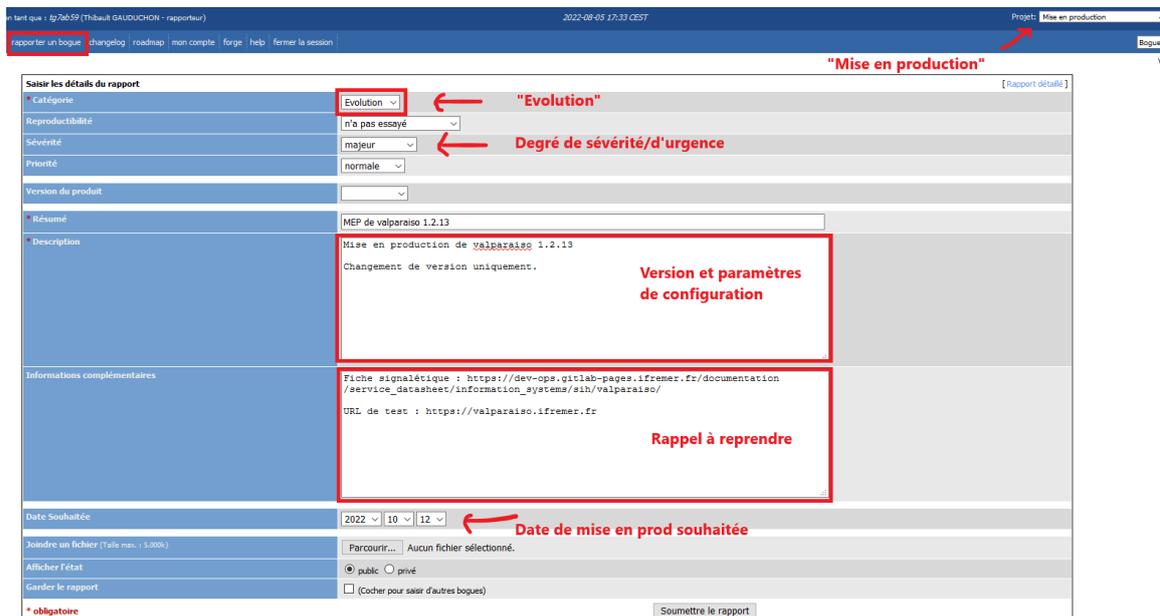
La demande comprend :

- Un rapport d'évolution dans Mantis ;
- Un mail de demande de « Mise en production » de la nouvelle version adressé à : [assistance@ifremer](mailto:assistance@ifremer.fr), webric@ifremer.fr
 - Joindre le lien vers le rapport Mantis ;
 - Préciser le numéro de version ;
 - Préciser s'il s'agit uniquement d'un changement de version ou si des paramètres de configuration sont en jeu.

¹ https://gitlab.ifremer.fr/-/ide/project/dev-ops/isival/isival_application/edit/master/-/group_vars/all/vars.yml

Fiche Mantis

Nouveau rapport : https://forge.ifremer.fr/mantis/bug_report_page.php



en tant que : [Ajouter un bogue](#) | [change log](#) | [roadmap](#) | [mon compte](#) | [forge](#) | [help](#) | [fermer la session](#) | 2022-06-05 17:33 CEST | [Mise en production](#) | [Rapport #](#) | [Voter](#)

"Mise en production"

Saisir les détails du rapport [rapport détaillé]

* Catégorie: Evolution ← "Evolution"

Reproductibilité: n'a pas essayé

Sévérité: majeur ← Degré de sévérité/d'urgence

Priorité: normale

Version du produit:

* Résumé: MEP de valparaiso 1.2.13

* Description: Mise en production de valparaiso 1.2.13
Changement de version uniquement. **Version et paramètres de configuration**

Informations complémentaires: Fiche signalétique : https://dev-ops.gitlab-pages.ifremer.fr/documentation/service_dashboards/information_systems/sih/valparaiso/
URL de test : <https://valparaiso.ifremer.fr> **Rappel à reprendre**

Date Souhaitée: 2022 10 12 ← Date de mise en prod souhaitée

Joindre un fichier (Taille max. : 5,000k): Parcourir... Aucun fichier sélectionné.

Afficher l'état: public privé

Garder le rapport: (Cocher pour saisir d'autres bogues)

* obligatoire

Dans la partie Description de la fiche :

- Préciser « Mise en production » de la version souhaitée ;
- Toujours rappeler le lien vers la fiche signalétique de l'application et l'URL de production ;
- Le cas échéant, penser à bien spécifier les paramètres de configuration qui seraient à ajouter ou à modifier (repris des essais fait préalablement sur Isival).

Pour plus de précision concernant la soumission de rapports Mantis, se rapprocher de l'assistance.

II. Base de connaissance et de qualification dans Harmonie

Contexte

La base dite de connaissance de Valparaiso est un nouveau **schéma du système Harmonie**, composé de **tables indépendantes** les unes des autres mais toutes organisées de la même manière.

L'usage de cette base de connaissance a évolué au fil des développements et des actions prises en charge.

La base de connaissance fut créée afin de garder trace de certaines informations ou commentaires laissés par les utilisateurs dans Valparaiso concernant les anomalies détectées dans les données (ex. : des espèces jugées « exceptionnelles » dans les marées OBSDEB).

L'action OBSVENTES utilise désormais ce schéma pour qualifier les lots d'espèces dans les données brutes d'Harmonie, par le biais d'un pont, ou transfert de champs, entre les tables Harmonie.

Documentation du module

On trouvera, en complément de la documentation suivante, un chapitre dédié au paramétrage et au fonctionnement des modules Shiny de base de connaissance dans une vignette de l'application (cf. librairie valparaiso).

II.1 Usages de la base de connaissance

Dans le schéma de la base de connaissance, on trouve en générale : **1 table par contrôle** dont on souhaite garder trace, pour chaque action.

Comme il est dit plus haut, dans le cadre particulier d'OBSVENTES, le contrôle des lots (couples espèce-navires) débouche finalement sur leur qualification à usage scientifique ; mais le principe de création et la mécanique d'interaction avec la base de connaissance restent les mêmes ; seule la philosophie d'usage et les instructions aux utilisateurs changent (cf. OBSVENTES).

Des entités ou individus statistiques contrôlés très variables

Le module Shiny d'interaction avec les tables du schéma de connaissance est générique : il a été conçu pour être réutilisable pour toutes les actions. L'entité contrôlée (c'est-à-dire : 1 ligne d'une table de connaissance) est spécifique de chaque contrôle et très variable. Quelques exemples :

- Activité des navires : contrôle du couple « mois-navire » au regard de la présence de recettes dans SACROIS ;
- Activité des navires : contrôle de chaque « mois-navire-métier » au regard du référentiel de contrôle Obsdeb ;
- OBSDEB : contrôle des maillages des métiers : couple « sortie-métier ».
- Etc.

Chacun des exemples cités ci-dessus fait l'objet d'une table indépendante dans le schéma de connaissance.

II.2 Création d'une nouvelle table dans la base de connaissance

La création d'une nouvelle table dans la base de connaissance revient à créer une table dans le schéma VALPARAISO_DBA du système Harmonie.

La création d'une nouvelle table de connaissance devra sans doute rester la prérogative du développeur de l'application, en lien avec les thématiciens et référents statistiques de l'action.

Nom de table standardisé

Les tables de la base de connaissance sont nommées selon le patron suivant : « VALPO_[ACTION]_[CONTROLE] ». Exemple : « VALPO_OBSVENTES_NAVIRE_ESPECE » pour le contrôle et la qualification du couple espèce-navire dans OBSVENTES.

Création des tables existante

Les commandes de création de toutes les tables de connaissance et de qualification existantes sont présentes dans le fichier :

« ./dev/CREATE_VALPO_TABLES.sql »

A titre d'exemple, on trouvera dans ce fichier la création de la table de qualification OBSVENTES des lots d'espèces :

```
CREATE TABLE "VALPO_OBSVENTES_NAVIRE_ESPECE"
(
    "CREATION_DATE" TIMESTAMP(0) DEFAULT SYSDATE NOT NULL,
    -- KEY TO USER --
    "USERNAME_EXTRANET" VARCHAR2(7),
    "LASTNAME" VARCHAR2(20),
    "FIRSTNAME" VARCHAR2(20),
    "DEPARTMENT_FK" NUMBER,
    "DEPARTMENT_CODE" VARCHAR2(40),
    "PRIVILEGE" NUMBER,
    -- VALIDATION FIELD « VALPO » --
    "VALIDE" VARCHAR2(10),
    "COMMENTAIRE" NVARCHAR2(512),
    -- KEY TO OBSERVATION --
    "PROGRAM_FK" VARCHAR2(20),
    "ANNEE" VARCHAR2(4),
    "PROGRAMME_CODE" NVARCHAR2(150),
    "PLAN_CODE" NVARCHAR2(150),
    "NOM_STRATE" NVARCHAR2(150),
    "LIEU_COD" NVARCHAR2(30),
    "DATE_SORTIE" NVARCHAR2(10),
    "VESSEL_FK" NVARCHAR2(15),
    "TRIP_NUMBER" NVARCHAR2(15),
    -- CHAMPS COMPLEMENTAIRE --
    "ESPF_COD_RIC" NVARCHAR2(100),
    -- CHAMPS A VALIDER/QUALIFIER --
    "ESPF_COD_OBSVENTES" NVARCHAR2(100)
)
```

On retrouve dans cette commande les différents groupes de colonnes (signalétique utilisateur, champs dits « Valpo », etc.) décrits dans le propos général sur la base de connaissance au chapitre suivant : « Champs et groupes de colonnes ».

II.3 Champs et groupes de colonnes

Les tables de connaissance répondent à un patron bien précis de construction, ce qui forge son caractère générique. Les champs d'une table de connaissance Valparaiso sont regroupés comme suit :

- a. Les champs signalétiques de l'utilisateur ;
- b. Les champs signalétiques de l'entité contrôlée (identifiants ou clé d'identification en combinant les colonnes nécessaires);
- c. Les valeurs de contrôle (référentiels de contrôle ou valeur calculée) ;
- d. Les valeurs contrôlées proprement dites ;
- e. Les champs dits « VALPO » : commentaires, champs de validation ou de qualification, c'est-à-dire les champs propres de la base de connaissance, renseignés par les utilisateurs via les modules déployés.

Champs signalétiques de l'utilisateur

Ces champs sont communs à l'ensemble des tables du schéma de connaissance Valparaiso et doivent être systématiquement repris afin d'identifier les auteurs.

Groupe de colonne	Nom du champs	Valeur.s	Description
USER	USERNAME_EXTRANET		ID Extranet de l'utilisateur, utilise pour l'identification de celui-ci dans Harmonie après connexion CAS
	LASTNAME		Nom de l'utilisateur
	FIRSTNAME		Prénom de l'utilisateur
	DEPARTMENT_FK		ID du département/société
	DEPARTMENT_CODE		Code du département/société
	PRIVILEGE	1,2,3	Privilège Valparaiso de l'utilisateur

Origine et calcul des « privilèges » Valparaiso

Un champ particulier de ce groupe de colonne USER est le champ : « PRIVILEGE ».

Ce champ est important car il est impliqué dans le traitement de la base de connaissance et dans le mécanisme de fonctionnement des modules lorsqu'on souhaite mettre à disposition des utilisateurs des colonnes réservées à leur statut (ou privilège précisément). On parle dans ce cas de « déclinaison » des champs selon le privilège.

Le **privilège Valparaiso** est **calculé par action** ; il s'agit d'une **combinaison de deux informations** signalétiques de l'utilisateur tirées d'Harmonie :

1° Le **statut** le plus élevé attribué à l'utilisateur pour cette action dans le système Harmonie (les statuts étant cumulables) ;

2° Son appartenance à **l'IFREMER** ou à une **société prestataire** pour la collecte de données.

Dans le système Harmonie, le ou les statuts associés à chaque action pour les personnes sont tirés de : « "DENORMALIZED_PROGRAM2PERSON » ; les modalités des statuts utilisés sont décrites dans : « PROGRAM_PRIVILEGE ».

C'est ce nom de « privilège » que Valparaiso a repris à son propre compte.

Description des modalités du champ « PRIVILEGE »

Afin de conserver le caractère transversal et générique des modules, les modalités et la règle de calcul du champs « PRIVILEGE » propre à Valparaiso sont homogènes (i.e. les mêmes) pour toutes les actions. Les trois privilèges Valparaiso sont :

1. « Saisisseur » (Ifremer ou prestataire) :

Règle : Droit max sur l'action = « Saisisseur »

2. « Valideur Société »

Règle : Droit sur l'action = « Valideur » (ou « Administrateur » ou « Consultation étendue ») et DEPARTMENT **différent de** « IFREMER »

3. « Valideur Programme »

Règle : Droit sur l'action = « Valideur » (ou « Administrateur » ou « Consultation étendue ») et DEPARTMENT = « IFREMER »

Attention à l'évolution éventuelle du champ « DEPARTMENT » Pour le personnel de l'Ifremer, il existe dans Harmonie d'autres modalités pour le champ « DEPARTMENT » notamment pour décrire le laboratoire ou l'unité d'appartenance des personnes. Il est important pour Valparaiso que, dans la table Harmonie de description des statuts sur les actions « PROGRAM2PERSON » ne soit utilisé que la modalité « DEPARTMENT_FK = 408 » i.e. « IFREMER ».

Clé d'identification - Entité contrôlée

Les champs d'identification de l'entité contrôlée dépendent à la fois de l'action et du contexte de contrôle. Ils doivent être composés avec rigueur afin de pouvoir établir à l'avenir des liens avec les données de départ ou des produits par les exploitants d'Harmonie.

Les entités contrôlées sont diverses. Exemple :

- « un lot d'espèce d'un navire échantillonné lors d'une sortie » dans le cadre d'OBSVENTES, ou couple « espèce-navire » ;
- « un maillage d'un métier, rapporté dans une marée OBSDEB » ;
- « la recette mensuelle d'un navire » ou mois-navire, dans les calendriers d'activité.

Champs de contrôle

Les champs de contrôle regroupent les valeurs numériques ou qualitatives, uniques ou multiples, permettant d'établir le contrôle. Exemple : les bornes min/max d'un maillage pour un métier donné, valable sur une façade donnée.

Origine des champs de contrôle Les champs de contrôle peuvent être issus d'un référentiel de contrôle géré par les thématiciens de l'action, en autonomie (ex. : référentiels de contrôles OBSDEB) ; ils peuvent également être directement issus de référentiels Harmonie (ex. : la régionalisation des espèces ou des métiers) ; ou encore issus d'une source de données de référence comme SACROIS (ex. : contrôle de l'activité des navires avec la recette mensuelle fournie par Sacrois pour le contrôle de l'inactivité d'un navire un mois donné).

Ces champs font partie de la clé d'identification d'une anomalie, et donc d'une ligne complète dans la table de connaissance. Ainsi, une modification de la valeur de contrôle (révision de bornes de contrôle par les thématiciens de l'action) sera susceptible de générer de nouvelle anomalie. C'est bien ce que l'on souhaite puisque l'évolution d'une borne de contrôle appelle naturellement un nouveau contrôle. Dans la base de connaissance, on doit aussi pouvoir tracer les valeurs de contrôle : « au moment où le contrôle a eu lieu, telle valeur a été jugée anormale au regard de telle autre » (cf. Principe d'Historicité).

Champs contrôlés

Les champs contrôlés peuvent être quantitatifs (une valeur de maillage) ou qualitatifs (une espèce), calculés (ex. : temps de marée rapporté à 24h dans OBSDEB) ou bruts et directement issus des données.

Champs « VALPO »

Ces champs sont le propre de la base de connaissance de Valparaiso. Il s'agit des champs de commentaires, de validation ou de qualification.

II.4 Principe d'historicité

Il s'agit d'un principe fondamental du fonctionnement du schéma de connaissance Valparaiso : un utilisateur **écrit toujours de nouvelles lignes** dans la table de connaissance depuis l'application Valparaiso et **la création de chaque nouvelle ligne est datée** automatiquement dans la colonne « CREATION_DATE ».

Groupe de colonne	Nom du champs	Valeur.s	Description
[CHRONO]	CREATION_DATE		Date-heure (à la seconde) de creation de la ligne dans la table de connaissance

Illusion de suppression S'il supprime un commentaire dans l'interface du module pour le laisser vide et clique sur « Sauvegarder », l'utilisateur aura l'impression de *supprimer* le commentaire précédent. Faisant cela, il aura en fait bel et bien écrit une nouvelle ligne dans la table de connaissance et n'aura rien *supprimé* du tout !

Mais sur cette nouvelle ligne créée : le champ COMMENTAIRE est effectivement « vide » ; les champs signalétiques de l'entité contrôlée/commentée sont les mêmes que précédemment ; les champs signalétiques de l'utilisateur sont les siens. La date de création de la ligne permettra simplement d'identifier qu'il s'agit de la dernière version des champs de commentaire et de validation laissée par un utilisateur pour l'entité en question.

Le caractère historique de ces tables permet de conserver plusieurs versions des champs « VALPO » (commentaires, validation, qualification) laissés par les utilisateurs.

II.4.1 Traitement et filtrage des tables de connaissances

Récupération de la dernière version d'une ligne de connaissance post-chargement

Comme nous venons de le voir plus haut, les tables du schéma de connaissance VALPARAISO sont des tables historiques, qui garde en mémoire ou « versionne », en quelque sorte, les champs de commentaire, de validation et de qualification.

Le traitement des tables consiste à ne restituer à l'utilisateur (dans l'interface) que la version la plus récente de chaque ligne.

Champs déclinés (ou non) par privilège utilisateur

Les modules de base de connaissance ont été conçu de manière à pouvoir « décliner » facilement des champs du groupe de colonnes « VALPO » (par exemple le champ « COMMENTAIRE ») selon les privilèges de Valparaiso.

L'objectif est de présenter les champs déclinés côte-à-côte dans l'interface, et de les réserver en écriture selon le privilège de l'utilisateur. Cette disposition permet aux utilisateurs de profils différents d'écrire sur la même ligne du tableau, dans des colonnes réservées à chacun, une information de même nature qui sera par ailleurs stockée en base dans la même colonne.

Plus concrètement, prenons l'exemple du module de qualification des lots OBSVENTES :

- Le champs COMMENTAIRE est décliné par privilège.
- Dans la base de connaissance, ce champ est unique.
- Dans le serveur du module, les données du champ seront préalablement basculées de ligne à colonnes selon les privilège 2 et 3 (voir plus haut « Description des modalités du champ « PRIVILEGE »).
- Dans le tableau de l'application, le champ COMMENTAIRE sera ainsi décliné : COMMENTAIRE_SOCIETE ou « Commentaire (Société) » et COMMENTAIRE_PROGRAMME ou « Commentaire (Programme) ».
- Le privilège Valparaiso associé à l'utilisateur déterminera la colonne dans laquelle il a le droit d'écrire.
- Avant d'être sauvegardées en base, les nouvelles lignes de commentaire renseignées dans l'interface seront rebasculées en lignes, conformément au format de la table de connaissance dans le schéma Harmonie, afin d'être inscrites en base.

Lorsque des champs « VALPO » sont demandés à être déclinés par privilège utilisateur, on retiendra la dernière version des champs POUR CHAQUE PRIVILEGE dans le traitement post chargement de la table de connaissance.

Tous les modules ne requièrent pas cette déclinaison ; les champs de commentaires et de validation peuvent être communs et ouverts à tous les profils sans distinction. Dans ce

cas, le traitement post chargement de la table retiendra simplement la dernière version des champs « VALPO » quelle que soit la valeur du « PRIVIEGE ».

Données transmises au module

Le choix des données transmises au module dépendra de son utilisation. Si l'on souhaite ne présenter à l'utilisateur *que* les anomalies détectées, on veillera à transmettre au module les données préalablement filtrées selon le contrôle, après traitement des données (voir plus loin : « Logique commune de traitement »).

C'est à ce jour le cas de l'ensemble des actions sauf OBSVENTES, dont la base de connaissance est utilisée comme base de qualification exhaustive de la donnée. Dans la mesure où l'ensemble des données (lots espèces-navires d'une part, sortie d'autre part) doivent pouvoir être qualifiées, c'est bien l'ensemble des données qui est transmis à ce module en particulier.

Filtre interactif des données de connaissance

Les modules « VALPO » de connaissance sont également prévus pour pouvoir filtrer l'**affichage** des données transmises **par Navire** ou **par Sortie** afin de ne pas surcharger l'interface.

Exemple 1 : le module de qualification Espèce-navire d'OBSVENTES filtre les données au navire.

Exemple 2 : l'action OBSDEB ne filtre aucun de ses modules de connaissance sur les anomalies détectées.

II.5 Connexion Valparaiso - Ecriture dans la base de connaissance

Dans l'interface de l'application, l'interaction entre l'utilisateur et la base de connaissance Harmonie est gérée **table par table**, grâce au module Shiny « rhandsonValop » (fichier : « ./R/module_valpo_db.R »).

On parle bien ici de **module au sens Shiny** : c'est-à-dire d'un couple de fonctions serveur et UI (comme toutes application Shiny) prévues pour être appelées depuis le serveur et l'UI principale de l'application.

Le Choix du paquet {rhandsonable}

La librairie R {rhandsonable} offre dans l'interface des modules un tableau en tout point comparable à Excel, donc très intuitif pour les utilisateurs et très souple. La librairie dispose des options nécessaires pour empêcher la modification de certaines colonnes et propose de types de colonnes intéressants (cases à cocher correspondant au TRUE/FALSE de R, liste déroulante de choix restreints, etc.)

Autres avantage côté serveur : les données entrées dans les tableaux par les utilisateurs sont **recupérables en totalité** pour être traitées.

Ce module est générique et paramétrable, c'est-à-dire qu'il fonctionne de la même manière avec l'ensemble des tableaux de base la de connaissance de toutes les actions ; mais dispose de nombreux paramètres pour s'adapter à des besoins différenciés (mise à

disposition de champs de connaissance en fonction du statut, ou privilège des personnes ; champs partagés ; champs illustratifs n'appartenant pas à la table de connaissance Harmonie, etc.).

Avantages de la **modularisation** (au sens Shiny) du code pour déployer des tableaux **rhandsontable** en lien avec la base de connaissance :

- Éviter l'inflation du code du serveur principal de l'application en raison des nombreuses tables existantes dans le schéma de connaissance
- Déployer de façon dynamique les modules nécessaires, en fonction des actions ouvertes par l'utilisateur.

III. Données de Valparaiso

L'objectif de ce chapitre est de rendre compte de l'ensemble des sources de données dont se nourrit l'application Valparaiso.

III.1 Approvisionnement en données d'une session Valparaiso

L'essentiel des données traitées dans une session de validation Valparaiso provient de l'entrepôt de données du SIH, mis en place pour ce projet. Ce canal est privilégié pour plusieurs raisons :

- Il répond au souci de ne pas surexploiter le système Harmonie ;
- Il permet une lecture rapide de tableaux de données parfois volumineux grâce au format parquet de ses fichiers, et son positionnement sur le réseau favorise un chargement rapide des données.

Les données peuvent également provenir de requêtes directes en base Harmonie lorsqu'il s'agit de données peu volumineuses (Activité des navires : Synthèse de l'activité historique d'un navire) ; ou que l'équipe thématique veut pouvoir effectuer des corrections de données ou saisir des commentaires dans l'outil de saisie tout de suite visibles dans Valparaiso (OBSBIO : données de l'année de validation).

III.2 Entrepôt de données du SIH

Idéalement, le développeur dispose d'une connaissance du modèle Harmonie pour pouvoir lire les requêtes existantes et produire ses propres requêtes pour des besoins futurs.

L'entrepôt de données dispose d'un espace de test et d'un espace de production.

L'Espace de test

Configuration et données Test :

« \\iota1\isi-projets\sih\logiciels\int\tools\exportation\EXPORT_CSV_PARQUET\ »

Registre des requêtes en production : « ./requetes »

Fichier de configuration : « ./export_data.yaml »

L'Espace de production

Configuration de la Production :

« \\iota1\harmonie_exp\tools\exportation\EXPORT_CSV_PARQUET\ »

Registre des requêtes en production : « ./requetes »

Fichier de configuration : « ./export_data.yaml »

Emplacement de production des données : « \\iota1\sih_donnees\espace\valparaiso »

Ajouter un nouveau dépôt de données

L'ajout d'un nouveau dépôt ou service de fourniture de données peut être réalisé en autonomie, en fonction des connaissances du modèle de la base de données Harmonie.

Pour les besoins de connexion à des bases spécifiques (ex. : WAO, SMARTDOT) voir directement avec G. PRIGENT.

Sans connaissance du modèle, en accord avec E. LEBLOND, demander assistance prioritairement à S. LE BLOND ou G. PRIGENT.

Les étapes (d'abord dans l'espace de test):

1. Travailler la requête SQL qui extraira les données et la placer dans le registre ;
2. Choisir le cas échéant le ou les critères de ségrégation des données extraites (classiquement l'Année pour une majorité de dépôts ; ou encore l'Espèce scientifique pour OBSBIO).
3. Ajouter et régler la configuration du dépôt dans le fichier de configuration global de l'entrepôt (fréquence de mise à jour, nom des fichiers, chemin, paramètre de connexion etc.)
4. Tester le résultat du service en le lançant individuellement et manuellement.

Librairie de lecture Parquet : contexte et perspectives

Le chargement des données au format Parquet depuis l'entrepôt se fait actuellement fichier par fichier.

Les données d'action étaient jusqu'ici *séparées* en fonction du principal critère d'entrée dans les données (l'Année ou l'Espèce scientifique pour les données de paramètres biologiques).

En octobre 2022, le traitement informatique fournissant les données de l'entrepôt s'est amélioré et fournit désormais des fichiers proprement *partitionnés*, conformément au standard de stockage de ce format.

L'architecture n'ayant toutefois pas été remise en cause (l'emplacement, le nom et l'extension des fichiers importés), les anciens fichiers sont essentiellement devenus des dossiers Parquet, dans lesquels on trouve maintenant des fichiers plus petits.

Dans le même temps, la librairie R de lecture du format Parquet `arrow` a évolué, et permet aujourd'hui d'exploiter pleinement cette structure de stockage des données, grâce à la fonction `open_dataset`. Celle-ci permet de déclarer, non plus seulement des fichiers, mais également des dossiers Parquet comme porte d'entrée vers les données.

L'utilisation conjointe de `dplyr` permet alors de requêter la donnée de façon sélective au sein des fichiers sans avoir à tous les charger dans R.

Documentation arrow : <https://arrow.apache.org/docs/r/>

La philosophie et la méthode de chargement des données dans Valparaiso n'a pas pu être modifiée par manque de temps suite à cette amélioration de l'entrepôt, qui ouvre de belles perspectives quant à l'optimisation et la souplesse de l'importation des données.

III.3 Système Harmonie

L'application Valparaiso dispose également d'une connexion directe à la base de données Harmonie, utilisée pour :

- Requêter le profil de l'utilisateur au démarrage d'une nouvelle session et après authentification via le CAS, en particulier :
 - Sa société : Ifremer/Prestataire ;
 - Les actions de collecte auxquelles il est affilié ;
 - Ses droits/privilèges/statuts vis-à-vis de ces actions (Saisisseur, Validateur, Administrateur).
- Requêter la base de connaissance VALPARAISO d'Harmonie ;
- Spécifiquement pour OBSBIO : requêter en directe les données annuelles à valider (cf. Action OBSBIO) ;
- Spécifiquement pour ACTIFLOT : requêter les données d'historiques de l'activité d'un navire pour le module éponyme.

III.4 Gitlab Ifremer

Les données provenant de Gitlab sont à ce jour :

- Les référentiels de contrôles OBSDEB gérés par le statisticien référent (J. RODRIGUEZ) ;
- Les référentiels de transcodage des espèces OBSVENTES (élaborés spécifiquement pour les contrôles de cette action) ainsi que le référentiel « PETS » des espèces prioritaires à échantillonner en priorité pour les actions OBSMER et OBSVENTES.

Avantages

- Versionnement et centralisation des fichiers ;

- Ces tables peuvent être maintenues directement par les référents statistiques des actions ;
- Leur modification est simple et à effet immédiat, dans la mesure où ces tables sont importées depuis Gitlab, ce qui permet leur pilotage sans passer par un redéploiement de l'application.

Désavantages

- Gitlab n'est pas un lieu conçu pour le stockage de données ;
- Ajoute une connexion supplémentaire pour l'application, par ailleurs non optimale pour charger des données.

Le choix de positionner tel ou tel table de contrôle ou de référence sur Gitlab, sur un disque réseau de l'action ou dans le système Harmonie est en général un choix pragmatique, même s'il n'apparaît pas comme le plus naturel.

III.5 Mise à jour des données

En résumé, la mise à jour des données chargées dans Valparaiso est variable et dépend :

1. Du type d'approvisionnement (cf. Approvisionnement en données) ;
2. De la fréquence de mise à jour des données dans le système.

Exemple : Les données SACROIS de l'année en cours sont mises à jour dans le système Harmonie tous les 15 jours ; la configuration du dépôt procède à un export tous les mois.

IV. Logique commune de chargement et de traitement des données

Valparaiso fonctionne selon une logique commune de traitement appliquée à l'ensemble des actions.

Ces grandes étapes pourront servir de fil rouge dans la partie traitement/contrôle des données côté serveur.

Cette logique commune, reprise pour chaque action, a pour but de :

- Faciliter la prise en main et donc la maintenance de l'application ;
- Servir les modules transversaux : « Navire-armateur » et « Population », détaillés plus loin.

Etapas communes de chargement et de traitement d'une action dans Valparaiso

1. L'année est toujours le premier critère de chargement des données. Il est également le principal critère de ségrégation ou de partitionnement des données de l'Entrepôt de données du SIH.

L'année de validation proposées à l'utilisateur dépend de l'action choisie :

Programme	Année la plus ancienne proposée	Année par défaut (proposée automatiquement)
ACTIFLOT	2018	N-1
OBSDEB	2018	N
OBSVENTES	2020	N
OBSBIO	2022	N

2. La présence de tous les fichiers nécessaires au dépôt est systématiquement vérifiée. Un message d'alerte préviendra l'utilisateur en cas d'absence de fichiers, l'invitant à prévenir l'assistance Harmonie.
3. Les données annuelles de l'action sont chargées et filtrées selon le portefeuille navires de l'utilisateur *pour ce programme et cette année de validation*. Pour en savoir plus sur la requête du portefeuille annuel d'un utilisateur, voir le fichier : « ./R/query_vessel.R » du projet Valparaiso.
4. L'ensemble des données nécessaires pour contrôler les données de l'action est chargé (y compris les référentiels Harmonie, les référentiels externes gérés via Gitlab le cas échéant, etc.).
5. Les contrôles sont effectués sur l'ensemble des données chargées.
6. Les données adéquates (en erreur ou non selon les choix thématiques) peuvent ainsi être transmises aux modules de connaissance « VALPO » préalablement déployés (dès l'ouverture des menus des actions).
7. Les données des différentes tables de connaissance de l'action sont chargées et également passées aux modules.

8. Pour les programmes ACTIFLOT, OBSDEB et OBSVENTES, les navires de l'action sont concaténés à la liste globale des navires chargés dans la session de l'utilisateur.
9. Les combinaisons d'observations des critères de filtrage des données de l'action sont établies et concaténées à la table transversale de mixage pour le module population.
10. Les filtres des critères de filtrage sont déployés dans la barre de droite ; les onglets orange de la barre de filtrage doivent être cliqués au moins une fois afin d'être rafraîchis et pour que le module fonctionne correctement (i.e. que les filtres des différents onglets de la barre de droite fonctionnent bien entre eux) ; le module est pleinement opérationnel quand plus aucun onglet ne s'affiche en orange.

L'utilisateur dispose alors de toutes les données nécessaires et du résultat des contrôles pour pouvoir naviguer dans les différents onglets graphiques, cartographiques, ou encore dans les modules de base de connaissance et de qualification.

*Pour OBSBIO les données de mixage sont transmises à un module Shiny ad hoc (voir OBSBIO) qui est une reprise de la mécanique présente dans le serveur principal pour le module « Population » de filtrage transversal des données via la barre de droite.

Côté serveur (R), des listes d'objets transversaux

Tous les référentiels importés sont stockés dans la liste réactive (reactiveValues) « ref\$... ». Lorsqu'un référentiel est nécessaire pour traiter une action, on vérifie qu'il n'a pas déjà été chargé à l'occasion d'un traitement ou d'une autre action.

Les données annuelles des actions sont plutôt stockées dans la liste « an\$... » ; les données filtrées via le module « Population » (barre de droite) dans la liste « pop\$... » ; « navlist\$... » recueille les données pour le module Navire-armateur.

Pour les données OBSBIO, un changement de philosophie a été opéré, mais qui ne doit pas troubler le développeur outre mesure. Tous les tableaux de données nécessaires à OBSBIO sont stockés dans une liste unique « obsbio\$... ».

V. Navigation multi-actions - Modules et objets transversaux

V.1 Chargement de plusieurs actions

Plusieurs actions peuvent être chargées simultanément à l'ouverture de l'application.

Ceci peut être particulièrement utile pour les actions OBSDEB et Calendrier d'activité pour lesquels il existe des croisements de données et des contrôles communs.

Le chargement simultané d'OBSDEB est de l'Activité a pu être testé et est fonctionnel. Il a du sens à condition de **charger la même année** de validation dans les deux actions.

Le chargement simultané des actions pose le défi du déploiement des filtres dans le module « Population ». Cette question est traitée et détaillée plus loin dans : « Module 'Population' : filtrage transversal des données ».

Limites

On l'a dit, l'utilisateur est pour le moment libre de charger des années de validation différentes pour différentes actions, dans la même session Valparaiso.

Il est arrivé de planter l'application en chargeant plusieurs actions simultanément, sans que la cause de ce plantage ait pu être clairement identifiée par manque de temps et par souci de priorité de développement : les actions travaillent d'abord indépendamment.

En cas de difficulté, de complexité de navigation (trop de filtres et de navires par exemple) ou de plantage, un utilisateur qui dispose de plusieurs actions peut toujours ouvrir une action seule par session pour travailler plus simplement.

Perspectives et pistes d'amélioration

Des développements devraient être entrepris pour bien cadrer les possibilités offertes aux utilisateurs en chargeant plusieurs actions simultanément. Des options de filtrage transversal devraient sans doute alors être ajoutées.

Les coordinateurs d'actions et les usagers multi-action de Valparaiso devraient être consulté sur leurs besoins éventuels en cette matière.

Pistes et idées pour poursuivre et améliorer le chargement et la navigation entre plusieurs actions de collecte chargées simultanément :

- Ajouter un critère « Année » et « Programme » dans le filtre Population ?
- N'autoriser le chargement multi-action QUE pour la même année de validation ? et créer les contrôles au chargement des données pour cela.

V.2 Module « Navire-armateur »

Le module de sélection des navires et le premier module transversal de l'application. Le navire est l'entité commune à toutes les actions excepté OBSBIO.

L'immatriculation du navire est la manière la plus courante et la plus robuste pour identifier un navire. C'est sous cette forme que le navire est en générale identifié dans les données.

Cependant, la seule immatriculation n'est pas nécessairement le point d'entrée le plus rapide ni le plus intuitif pour rechercher un navire.

Nom du navire, Armateur et Lieu d'immatriculation

Les requêtes de l'entrepôt enrichissent les données de collecte avec des informations administratives complémentaires sur les navires, à la date de l'observation, c'est-à-dire :

- La date de sortie pour OBSVENTES ;
- La date de la marée pour OBSDEB ;
- Le dernier jour du mois pour l'Activité des navires.

Ces informations sont à minima : le **nom du navire**, son **lieu d'immatriculation (ou quartier maritime)** et le **nom de l'armateur**.

Les navires chargés dans les différents menus des actions sont placés dans une liste commune afin de nourrir le module de sélection des navires.

Les navires pouvant être observés plusieurs fois par une action de collecte, on garde systématiquement les dernières caractéristiques du navire observé (nom, armateur, etc.) pour constituer ces listes.

La liste déroulante indique dans quelles actions le navire a été observé.

Le navire sélectionné via ce module est dit « navire actif ». Le navire sélectionné est alors relayé à de nombreux modules.

V.3 Module « Population » : filtrage transversal des données

Le module dit « Population » est un module de filtrage transversal des données de toutes les actions chargées.

Les filtres à déployer sont définis par action (ex. : pas de filtre sur les espèces à déployer pour l'Activité).

Dans le cas où plusieurs actions sont déployées, des filtres peuvent être communs (c'est-à-dire fondés sur les mêmes variables en base) ou spécifique d'une action.

- Exemple de filtres communs : les agrégations métiers, déployées aussi bien pour OBSDEB que pour l'Activité ;
- Exemple de filtres spécifiques d'une action : les combinaisons d'espèces observées dans OBSVENTES ; de manière plus général, les filtres relatifs au contrôles d'une action sont spécifiques de celle-ci (Ex. : le filtre « Montant SACROIS » est évidemment spécifique de l'Activité, etc.).

Tous les traitements des actions ACTIVITE, OBSDEB et OBSVENTES aboutissent à la constitution d'une liste de toutes les combinaisons observées dans les données des modalités de filtrage demandées pour cette action.

Cette table de combinaisons est appelée « table de mixage ». C'est cette table qui permet de réduire les choix des listes déroulantes à mesure que des modalités sont sélectionnées par l'utilisateur.

Les filtres sur les données sont regroupés en trois thèmes :

- Métiers-Lieux : filtres liés aux métiers de pêche et leurs agrégats, à la localisation des observations ; mais aussi parfois au saisisseur de la donnée, à la taille des navires pour l'action Activité ;
- Espèces : filtres relatifs aux espèces échantillonnées (OBSVENTES et OBSDEB seulement) ;
- Contrôles : résultats des contrôles sur les données.

Déploiement et activation des filtres dans l'interface

Les onglets de filtrage de la barre de droite sont redéployés à chaque nouveau chargement et traitement de données d'une action.

Pour rendre pleinement fonctionnel et réactifs entre eux les filtres des 3 onglets de la barre de droite, l'utilisateur devra cliquer au moins une fois sur chacun avant de commencer sa sélection de critères : la couleur orange des icônes indique à l'utilisateur les onglets à activer. Lorsque toutes les icônes sont blanches, le module est pleinement fonctionnel, sinon : la bonne réactivité/actualisation des filtres entre eux, à mesure que des critères sont sélectionnés, n'est pas garantie.

Le module « Population » de la barre de droite est réinitialisé à chaque chargement et traitement de données d'une action : l'opération « d'activation » décrite ci-dessus devra alors être répétée le cas échéant.

Chargement multi-action

La table de mixage du module « Population » est une concaténation en colonne de tous les filtres déployés, et en ligne de toutes les combinaisons observées dans les différents jeux de données ; les colonnes des filtres spécifiques sont alors vides pour les autres actions. La table de mixage ajoute le paramètre PROGRAM_FK afin de garder trace de l'action dans le mix des observations, et agir spécifiquement au moment du filtrage proprement dit : c'est-à-dire lorsque l'utilisateur, satisfait de ses choix, clique sur « Filtrer les navires et sorties ».

Propagation des contrôles sur un navire dans toutes les actions chargées où le navire est présent

On l'a dit, les filtres des contrôles sont nécessairement spécifiques des actions. Ainsi, dès lors que l'on commence à utiliser ce type de critère dans le module de filtrage, seule les observations de l'action en question pourront être sélectionnées...

Or, on peut souhaiter cibler un ou plusieurs navires de façon transversal sur la base d'un contrôle (action-spécifique donc). Exemple : Filtrer les navires en erreur dans l'activité, et filtrer dans la foulée les données OBSDEB relatives à ce navire (à condition que celui-ci soit observé dans les deux actions évidemment).

Pour réaliser cela, on procède à la propagation du résultat du contrôle réalisé dans l'Activité dans la table de mixage, aux observations du même navire dans les lignes correspondant aux données du même navire, mais des autres actions. Autrement dit, on comble le vide des champs de contrôle de la table de mixage, par navire.

La liste des champs de contrôles « propagés » par navire est inscrite dans le serveur principal de l'application (« B. MODULE NAVIRES (En-tête) » -> « 1. Champs de contrôles thématiques à propager »).

Perspective OBSMER-OBSVENTES

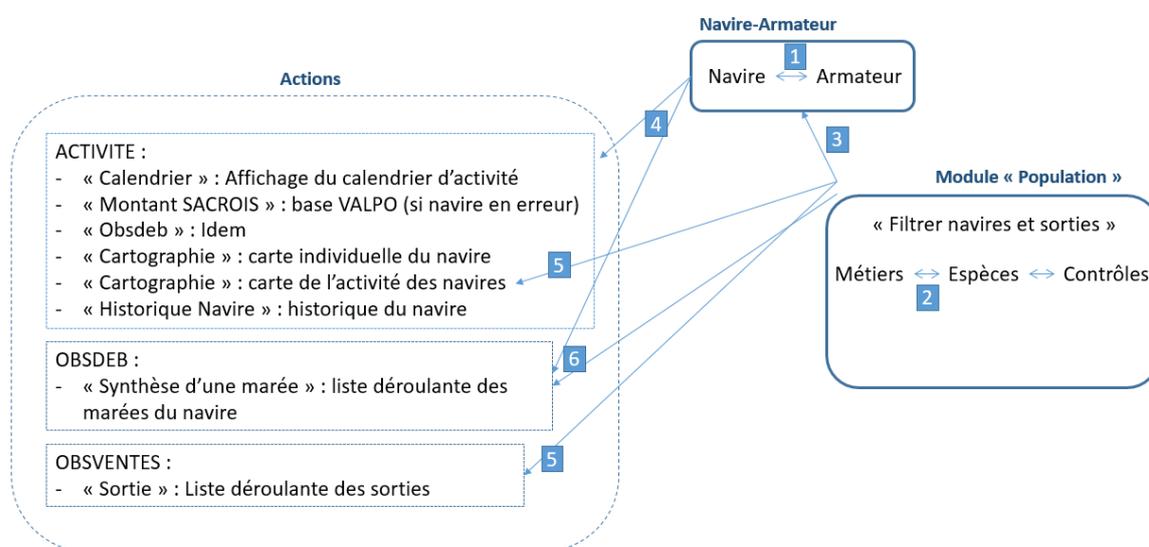
Jusqu'ici, la propagation a été réalisée pour les champs de contrôles Activité et OBSDEB.

Cette mécanique devrait être reprise à bon compte pour les actions OBSVENTES et OBSMER qui fonctionnent également en tandem.

V.4 Schéma général d'interaction entre le module de filtrage « Population », « Navire-armateur » et les modules spécifiques des actions

On l'a dit, le module « Population » a entre autre pour tâche de restreindre la liste des navires du module « Navire-armateur » selon des listes de critères. Mais ce module a également des répercussions directes sur d'autres filtres : en particulier les listes déroulantes des marées OBSDEB (du navire sélectionné, ou actif) et la liste déroulante des sorties OBSVENTES. Pour l'action Activité, le module « Population » est directement lié à la cartographie.

Le schéma complet des interactions entre les listes déroulantes des différents modules, transversaux et action spécifique, est détaillé si dessous :



- 1 Les filtres « Navire » et « Armateur » sont liés entre eux et s'actualisent mutuellement

- 2 Idem pour les filtres du module de filtrage transversal dit « Population »

- 3 Les filtres à droite* actualise et restreins la liste déroulante des navires

- 4 Le changement de navire se répercute sur les modules des actions listés

- 5 Les filtres à droite* actualisent la carte de l'activité des navires

- 6 Le changement de navire ET les filtres à droite* actualisent et restreignent la liste déroulante des marées OBSDEB

- 7 Les filtres à droite* actualisent et restreignent la liste déroulante des sorties OBSVENTES ET la disponibilité des navires pour le module de qualification des lots

*En cliquant sur le bouton « Filtrer les navires et sorties » du module

VI. Activité des navires

VI.1 Jeux de données

Jeux de données nécessaires à la validation des calendriers d'Activité.

Données d'Activité et OBSDEB :

- Calendrier d'activité : concaténation des informations administratives des navires et des données de l'enquête Activité (métiers pratiqués, jours de mer, etc.) ;

1 ligne = 1 navire-mois-métier

- Caractéristiques des engins pratiqués ;
- Liste des navires à enquêter pour recueillir leurs caractéristiques engins ;
- Données des métiers recueillis par l'action OBSDEB.

Données	Dossier ./ACTION/ACTIVITE
Calendrier d'activité	CALENDRIER_ACTIVITE
Caratéristiques engins	CARACTERISTIQUE_ENGIN
Enquête engin	ENQ_ENGIN
Métiers OBSDEB	METIER_OBSDEB

Source : Entrepôt de données SIH

Données SACROIS :

- Données SACROIS agrégées des recettes mensuelles des navires.

Données	Dossier ./SACROIS
Recettes mensuelles SACROIS	P08_SACROIS_NAVS_MOIS_MAREE

Source : Entrepôt de données SIH

Référentiels OBSDEB :

- Référentiels de contrôle OBSDEB Métier-Secteur ;
- Référentiels de contrôle OBSDEB Métier-Gradient.

Données	Dossier ./Control_repositories
Métier-secteur	OBSDEB_referentiel_controle_metier_secteur.csv
Métier-gradient	OBSDEB_referentiel_controle_metier_gradient.csv

Source : Gitlab Ifremer – Projet « SIH_HISSEO/ReferenceTabs_SIHOBSDEB »

Maintenance Les référentiels de contrôles OBSDEB sont maintenus à jour par le statisticien référent de l'action : J. RODRIGUEZ.

Référentiels Harmonie :

- Agrégations des métiers ;
- Régionalisation des métiers ;
- Hiérarchie des lieux et des secteurs géographiques.

Données	Dossier ./REFERENTIELS
Agrégations métiers	AGREGATION_METIER
Régionalisation métiers	REGIONALISATION_METIER

Hiérarchie	HIERARCHY
------------	-----------

Source : Entrepôt de données SIH

VI.2 Contrôles réalisés

Calendrier incomplet

Règle appliquée:

- Le Port d'exploitation principal du navire doit toujours être renseigné ;
- Si le navire est actif un mois donné, au moins un métier de pêche doit être saisi.

Régionalisation des métiers

Les métiers renseignés dans l'enquête Activité doivent être régionalisés selon les référentiels Harmonie. Tous les métiers de pêche ne peuvent normalement pas être déclarés partout.

Règle appliquée :

- Outre-mer : les métiers sont régionalisés « à terre », c'est-à-dire en fonction du lieu d'immatriculation, ou quartier maritime.
- En Atlantique et en Méditerranée : les métiers de pêche sont régionalisés « en mer » c'est-à-dire en fonction du secteur ou de la zone de pêche déclarée pour ce métier.

Hiérarchie des secteurs en mer Concernant la régionalisation des métiers en mer : les secteurs renseignés dans l'activité peuvent être plus précis que le ou les secteurs autorisés par le référentiel (ex. : Rectangle statistique vs Division CIEM). Dans le traitement des données d'activité, on fait appel à la hiérarchie des secteurs décrite dans le système Harmonie pour contrôler la bonne régionalisation des secteurs saisis dans l'enquête activité, et ce quel que soit le niveau saisi.

Action et référentiels OBSDEB « Métier-secteur » et « Métier-gradient »

- Pour les navires vus dans l'enquête OBSDEB, on vérifie que l'ensemble des métiers déclarés dans OBSDEB figurent dans le calendrier d'activité ;
- D'autre part, les métiers de pêche de l'activité outre-mer doivent être régionalisés selon les référentiels de contrôle OBSDEB « Métier-secteur » et « Métier-gradient ».

Recette mensuelle SACROIS

On vérifie que les navires déclarés « Inactif » un mois donné ne dépasse pas une recette de 1000 euros pour le mois en question dans le jeu de données consolidées SACROIS.

Caractéristiques engins à enquêter

Pour la liste des navires concernés par l'enquête sur les caractéristiques des engins pratiqués : on vérifie que chaque engin possède au moins 1 caractéristique saisie.

VI.3 Base connaissance pour l'enquête Activité

Contexte L'action Activité des navires n'a pu profiter que tardivement de la mise en place de la base de connaissance Valparaiso, mise en place après les développements réalisés pour cette action.

Une table de connaissance a été déployée pour les contrôles suivant :

- Recette mensuelle SACROIS >1000 euros et bateau déclaré inactif ;
- Combinaison Métier-secteur et Métier-gradient non présentes dans les référentiels de contrôle OBSDEB

Les modules de connaissance associés à ces contrôles ont été paramétrés très simplement :

- Affichage navire par navire, via le module « Navire-armateur » ;
- Affichage de navires en erreur seulement (mieux vaut donc les cibler d'abord via le module de filtrage à droite sur critères de contrôle) ;
- Les champs « Commentaire » et « Validation » sont ouverts à tous les privilèges ;
- Ces champs de connaissance ne sont pas « déclinés » par privilège (c'est-à-dire présentés en colonne avec une réservée pour chacun, voir OBSVENTES)

VI.4 Sélection d'un navire et navigation croisée OBSDEB

On sélectionne un navire dont on souhaite visualiser le calendrier d'activité de l'année à valider, l'historique de l'activité ou encore les tables de connaissance relatives aux contrôles, via le module transversal « Navire-Armateur » (voir plus haut « Module 'Navire-armateur' »).

Le navire peut également être ciblé et sélectionné depuis l'onglet cartographique (voir plus loin : « Cartographie de l'activité des ou du navire »).

L'idéal pour cibler un navire en erreur par exemple, est de restreindre la liste des navires du module « Navire-armateur » à l'aide du module de filtrage à droite dit « Population ».

Pour les utilisateurs multi-action Activité-OBSDEB qui ont ouvert dans la même session les deux actions, chargé et contrôlé les données de la même année à valider, ils ont la possibilité de pointer un navire présent dans les deux actions, et ainsi et de visualiser à la fois son calendrier dans : « Activité des navires -> Calendrier d'activité » et ses marées OBSDEB dans : « OBSDEB -> Synthèse d'une marée ».

VI.5 Visualisation du calendrier d'activité complet

Le calendrier d'activité du navire sélectionné dans la liste des navires disponibles dans le portefeuille de l'utilisateur s'affiche dans l'onglet « Calendrier d'activité ».

Mise en page du calendrier

Le calendrier d'activité est mis en page comme suit :

- En colonne : les 12 mois de l'année de validation, de janvier à décembre ;

- 3 groupes d'informations en ligne :
 - « Administratif » : Lieu d'immatriculation, Armateur et Présence du navire du FPC pour chaque mois de l'année.

Ces trois informations sont bancarisées dans Harmonie par période dans des tables comme VESSEL_OWNER_PERIOD pour la table des armateurs, avec une date de début et de fin d'armement (1 ligne armement). Pour constituer le calendrier d'activité, on requête cette information à date du **dernier jour de chacun mois**. Ainsi, lorsqu'un changement d'armateur survient un mois donné, le calendrier d'activité restituera bien le **dernier armateur du mois**, donc le changement qui a eu lieu par rapport aux mois précédents. De même pour une entrée ou une sortie du fichier de Flotte communautaire (FPC) et un changement de lieu d'immatriculation (ou quartier maritime).

- « Effort de pêche » :

La seconde partie du calendrier contient : 1° les informations de l'enquête activité suivantes : Le Port principal d'exploitation, l'Activité du navire (oui/non), le nombre de Jours de mer, de Pêche et d'Hommes à bord et 2° les informations agrégées tirées du jeu de données consolidé SACROIS sur le cumul des Quantités capturées et le cumul des recettes (sommes sur chaque mois).

- « Métiers » :

Cette dernière partie présente les principaux métiers pratiqués par le navire pour chaque mois ; le ou les Secteurs en mer principaux de pêche pour ces métiers, associés à un Gradient à la côte pour affiner la zone fréquentée.

Les métiers sont ordonnés par nombre de mois de pratique : les métiers les plus fréquemment pratiqués dans l'année figurent en tête.

VI.6 Historique de l'activité d'un navire et changement d'armement

La donnée concernant l'historique de l'activité d'un navire (le navire actif du module « Navire-armateur ») ne provient pas de l'Entrepôt de données SIH mais d'une requête directe effectuée sur la base Harmonie.

La table requêtée est un produit Harmonie : ASP_SYNTHESE_ACTIVITE.

Le module pointe également les changements d'armateur survenus dans l'histoire du navire.

VI.7 Cartographie de l'activité des navires : l'indicateur « mois-navire »

L'indicateur cartographié est un **cumul annuel de « mois-navire »** par secteur géographique. Autrement dit, pour un secteur géographique donné (ex. : un rectangle statistique) :

- Un navire ajoute 1 unité à l'indicateur « mois-navires » du secteur, chaque mois qu'il fréquente ce secteur.

- Donc, un navire qui fréquente 1 mois seulement dans l'année le secteur, contribuera *in fine* d'1 unité seulement dans le cumul annuel de l'activité de ce secteur.
- Ou encore, si le navire fréquente le secteur 2 mois dans l'année au total (que ce soit pour le même métier, pour un autre, ou pour plusieurs métiers ; n'importe le nombre de métiers) il contribuera alors de 2 unité au cumul des « mois-navires » du secteur.

Cartographie des secteurs « bruts »

Les secteurs en mer des enquêtes Activité sont potentiellement **multiples** et de **niveaux hiérarchiques différents** d'un calendrier à l'autre.

Lorsque l'utilisateur cartographie l'activité d'une population de navires, i.e. d'une flottille qu'il a lui-même définie à l'aide des critères des filtres à droite, les secteurs se recourent les uns les autres la plupart du temps.

La cartographie de Valparaiso mobilise tous les niveaux de secteur en mer possibles sans préjuger de ceux qui seront effectivement rencontrés dans les données : du quadrilatère à la plus petite unité statistique.

Le module permet à l'utilisateur de sélectionner ou de dé-sélectionner les niveaux de secteurs rencontrés.

Contrairement à l'ancien outil cartographique (actuellement en refondation) le traitement des données d'activité de Valparaiso ne procède pas à une réallocation de l'effort qui tienne compte du gradient à la côte.

En effet, l'objectif de l'application est de procéder au contrôle visuel « brut » des secteurs renseignés, et rendre compte des niveaux réellement présents dans les données.

Sans doute, dans certains cas, serait-il possible de réallouer l'activité à un niveau hiérarchique plus précis de manière certaine ; par exemple lorsque le niveau géographique renseigné est le rectangle statistique et que le gradient à la côte indique un sous-rectangle statistique parfaitement défini et identifiable.

Cette question devrait être creusée par l'équipe ISI en charge de la refonte de l'outil cartographique dédié à l'activité des navires. La cartographie de Valparaiso pourrait, si besoin, s'inspirer de ce travail pour proposer au utilisateur :

- D'autres représentations cartographiques de l'activité, avec réallocation ;
- Ou, un contrôle sur la précision du secteur renseigné combiné au gradient (indiquer par exemple s'il existe un type de secteur plus précis qui correspond parfaitement à la combinaison secteur-gradient renseignée).

Evolution de l'outil de saisie Ces besoins d'évolution dépendront des améliorations apportées à l'outil de saisie (Allegro) avec sa refonte, déjà annoncée. Plus l'outil de saisie contrôlera et contraindra en amont la saisie des secteurs aux niveaux pertinents souhaités, plus l'outil de validation pourra se concentrer sur d'autres aspects.

Lien avec le module « Population »

Pour les besoins de la cartographie et le bon calcul de l'indicateur de l'activité « mois-navires » décrit plus haut, lorsque l'utilisateur filtre la donnée d'Activité via le module à droite, l'individu statistique considéré est bien la combinaison « mois-navire » et non pas le « navire » tout entier.

En utilisant le module de la barre de droite, l'utilisateur filtre pour chaque navire : la fraction de son activité qui répond au jeu de critères choisis, que ce soit en termes de métiers pratiqués, de mois où des erreurs ont été détectées, de lieu d'immatriculation (car celui-ci peut avoir changé au cours d'une année comme on l'a vu plus haut), etc.

A l'issue de ce filtrage, une fois cliqué sur « Filtrer les navires les sorties », la liste de navires restant dans les données est actualisée dans le module « Navire-armateur », et l'indicateur mois-navire est calculé.

Exemple L'utilisateur choisit comme critères de filtrage : Département = « Charente-Maritime » et Métier de base = « Trémails à sole (GTXSOX) ».

Il reste alors dans les données cartographiées : les mois où les navires (i.e. les combinaisons « mois-navires ») étaient à la fois immatriculés en Charente-Maritime et ont déclaré avoir pratiqué du trémail ciblant principalement la sole.

Sélection d'un navire

En cliquant sur un secteur en

VI.8 Cartographie de l'activité d'un navire

L'activité d'un navire donné peut être mise en parallèle avec celle d'une population de navire (ou une flottille) ciblée par l'utilisateur, en sélectionnant un navire via la cartographie de gauche ou directement via le module Navire-armateur.

VI.9 Perspectives

L'information contenue dans les tables de connaissance de Valparaiso déployées dans le cadre de l'enquête activité pourraient être prise en compte dans les produits de type SACROIS (en cas d'enquête directe peut-être) si on considère qu'il peut y avoir là une source corrective intéressante pour ce produit très diffusé et très utilisé.

Des chercheurs ont également manifesté leur intérêt pour cette base de connaissance concernant l'activité des navires, notamment en tant qu'utilisateur de la donnée SACROIS ; ceux-ci peuvent avoir des questionnements sur ces données relativement à l'activité de certains navires.

La base de connaissance sur l'activité des navires pourrait peut-être faire l'objet d'une ouverture vers les chercheurs afin qu'ils puissent, via un privilège spécial, à créer sans doute, participer à l'effort de contextualisation et de commentaire pertinent sur la donnée.

VII. OBSDEB

VII.1 Contexte

Les contrôles de l'action OBSDEB sont intégralement repris d'un algorithme antérieur à Valparaiso développé par J. WEISS. Pour en savoir plus sur ces contrôles (historique, conception, justification), se rapprocher de l'équipe thématique OBSDEB et des statisticiens connaisseurs de cette action (en particulier S. DEMANECH, J. RODRIGUEZ).

Ces contrôles sont basés sur une série de référentiels de contrôle gérés par l'équipe OBSDEB depuis un dépôt Gitlab (J. RODRIGUEZ).

VII.2 Jeux de données

Référentiels de contrôles OBSDEB :

Données	Dossier ./Control_repositories
Métier-Secteur en mer	OBSDEB_referentiel_controle_metier_secteur.csv
Métier-Gradient à la côte	OBSDEB_referentiel_controle_metier_gradient.csv
Métier-Dimension de l'engin	OBSDEB_referentiel_controle_metier_dimension.csv
Métier-Durée de la marée	OBSDEB_referentiel_controle_metier_duree_maree.csv
Métier-Espèce	OBSDEB_referentiel_controle_metier_especes.csv
Métier-Maillage	OBSDEB_referentiel_controle_metier_maillage.csv
Métier-Quantité par marée	OBSDEB_referentiel_controle_metier_quantite_maree.csv
Métier-Prix-espèce	OBSDEB_referentiel_controle_metier_prix_espece_maree.csv

Source : Gitlab Ifremer – Projet « SIH_HISSEO/ReferenceTabs_SIHOBSEB »

Maintenance Les référentiels de contrôles OBSDEB sont maintenus à jour par le statisticien référent de l'action : J. RODRIGUEZ.

VII.3 Réalisation

Lors de la création de Valparaiso, le volet de développement OBSDEB fut consacré :

1. A la conception et à la mise en place de la base de connaissance VALPARAISO dans Harmonie pour recueillir des informations de terrain sur les anomalies ressorties des contrôles, et à terme, faire évoluer les référentiels à l'aune de ces retours;
2. Le déploiement des premiers modules de cette base de connaissance pour chaque contrôles OBSDEB existants ;
3. Mettre à disposition de nouvelles représentation des données :
 - a. La synthèse d'une marée OBSDEB et de ses anomalies potentielles ;
 - b. Cartographie des marées ;
 - c. Un contrôle visuel des poids des individus par espèce.

VIII. OBSVENTES

VIII.1 Préambule

Définitions du mot « lot »

Attention, ce terme possède de nombreuses significations.

Découpage territorial Le territoire métropolitain d'observation de l'action OBSVENTES est découpé en 4 « lots » d'observation : Méditerranée (lot n°4), Golfe de Gascogne (lot n°3), etc.

Caisses de vente Sur le terrain, on parlera de « lots » d'un navire, pour désigner des caisses de vente observées ou de débarque. Le terme est alors étroitement lié au contexte d'observation.

Lot informatique Dans le logiciel de saisie Allegro, le « lot » est un concept générique qui correspond aux niveaux hiérarchiques de l'arbre de saisie situés sous le navire, qui est le premier point d'entrée dans la saisie; autrement dit : après avoir créé dans cet outil un navire (et lui avoir associé une vente dans le cas d'OBSVENTES), on saisit des lots, ou des lots de lots, etc., à mesure que l'on précise la saisie en termes d'espèce commerciale, catégorie, sexe, espèce scientifique...

Principe de base des contrôles – périmètre d'observation métropolitain

Les contrôles actuels des données OBSVENTES sont principalement fondés sur le croisement avec les données de ventes du Réseau inter-cirées (RIC).

Les lots ouverts aux contrôles et à la validation sont les **lot 1 à 3** de l'Atlantique et le **lot 4** de Méditerranée.

En l'absence de données de référence sur les ventes, les données OBSVENTES collectées outre-mer ou dans le cadre d'autres programmes ne rentrent pas (encore) dans le périmètre de validation de l'application.

VIII.2 Jeux de données

Liste des jeux de données nécessaires à la validation des données OBSVENTES dans Valparaiso.

Données OBSVENTES :

- Ventes observées OBSVENTES ;
- Table des captures (poids) agrégées au taxon de vente (espèce principale de la caisse de vente) ;
- Table des mensurations : mesures individuelles.

Données	Dossier ./ACTION/OBSVENTE
Ventes observées	OBSVENTE_SALE_NAV_STRATA
Captures	OBSVENTE_SL_CAISSSE
Mesures individuelles	OBSVENTE_HL

Source : Entrepôt de données SIH

Données WAO (plan d'échantillonnage) :

- Contacts-espèces : description des sorties, des espèces à échantillonner, des contacts navire
- Suivi du plan : réalisation des lignes de plan, par trimestre, en termes de sorties

Données	Dossier ./WAO
Contacts-espèces	CONTACT
Suivi du plan	LIGNE_PLAN

Source : Entrepôt de données SIH

Référentiels Harmonie :

- Référentiel RTP
- Référentiel des lieux : ports, criées et quartiers
- Référentiel espèces

Données	Dossier ./REFERENTIELS
RTP	RELATION_TAILLE_POIDS
Lieux, ports, quartier	LIEU_PORT_CRIEE_QUARTIER
Espèces	ESPECE_FAO

Source : Entrepôt de données SIH

Données de ventes :

- Ventes en criées

Données	Dossier ./VENTE
Ventes en criée	VENTE_CRIEE

Source : Entrepôt de données SIH

Référentiels OBSVENTES :

- PETS : référentiels des espèces prioritaires et systématiques ;
- Transcodage des espèces entre WAO, RIC et OBSVENTES;
- Paramètres : paramètres divers de pilotage thématique.

Données	Dossier ./raw-data
PETS	./valpo_obsventes_PETS.csv
Paramètres divers	./valpo_obsventes_param.csv
Transcodage	./valpo_obsventes_taxon_transcode.csv

Source : Gitlab Ifremer – Projet Valparaiso

Base de connaissance/qualification :

- Table de qualification des lots d'espèces-navires ;
- Table de qualification des sorties.

Données	Schéma SIH2_VALPARAISO_DBA
Espèces-navires	VALPO_OBSVENTES_NAVIRE_ESPECE
Sorties	VALPO_OBSVENTES_SORTIE

Source : Système Harmonie – Service HARMONIE

VIII.3 Traitements

Les données OBSVENTES sont traités par lot et par trimestre, entre afin de réduire le jeu de vente conservé en mémoire.

VIII.3.1 Traitement du référentiel RTP

Les coefficients RTP utilisés ont une clé : Espèce + Zone (Atlantique/Méditerranée) + Trimestre + Sexe.

On utilise les coefficients les plus récents disponibles dans le référentiel.

La clé trimestre est utilisée lorsque les 4 trimestres sont présents pour l'espèce dans le référentiel (ex. d'espèces : clé RTP annuelle pour *Dicentrarchus labrax* ; clé trimestrielle disponible pour *Solea solea*).

La zone de pêche (division CIEM) n'étant pas disponible dans les données OBSVENTES extraites du système Harmonie pour la validation (voir TODO) la seule clé Zone possible est la façade : Atlantique/Méditerranée ; les coefficients détaillés aux divisions CIEM, s'ils existent, sont donc moyennés par façade.

De même, le type ou paramètre de longueur (LENGTH_TOTAL etc.) n'est pas disponible dans les données exploitées pour la validation, et ne fait donc pas parti de la clé de sélection des coefficients de la RTP. Dans le référentiel RTP les longueurs utilisées par défaut sont : LENGTH_CARAPACE, LENGTH_MANTLE, LENGTH_TOTAL.

VIII.3.2 Transcodage des espèces

Espèces WAO, RIC, OBS

Pour pouvoir joindre les données du plan d'échantillonnage, les données de contrôle du RIC et les données observées OBSVENTES il est nécessaire d'établir une correspondance entre :

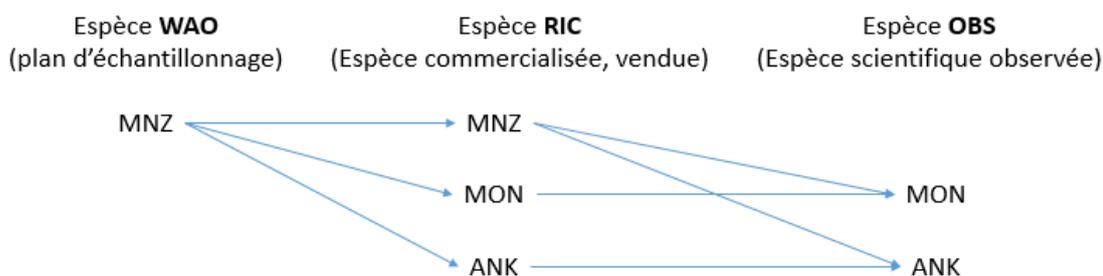
- L'espèce du plan d'échantillonnage, parfois plus large (famille d'espèce) que l'espèce scientifique ;
- L'espèce commerciale vendue (le taxon de la caisse de vente enregistré dans le RIC) ;
- L'espèce scientifique réellement échantillonnée.

Attention : Code RIC = Taxon caisse OBSVENTES La règle donnée aux observateurs est de saisir comme espèce commerciale (avant l'espèce scientifique) l'espèce telle qu'elle est vendue (étiquette de la caisse) et donc, telle qu'elle devrait apparaître ensuite dans le RIC. C'est pourquoi dans Valparaiso, on suppose Espèce RIC = Espèce taxon caisse dans les données de OBSVENTES_SL_CAISSE.

L'espèce du plan et l'espèce commerciale peuvent être un regroupement d'espèces scientifiques. Ex. : les baudroies (MNZ) regroupent les baudroies rousse et commune (MON et ANK). Si MNZ est demandée par le plan d'échantillonnage, il faudra bien lui associer MON ou ANK comme espèce scientifique observée ou comme espèce vendue, bien que les baudroies soient généralement vendues indistinctement sous MNZ. Mais le

plan peut (rien ne l'en empêche) préciser dès le départ une espèce scientifique précise (MON ou ANK).

Le diagramme complet de transcodage si MNZ est demandée par le plan d'échantillonnage est donc (toutes les combinaisons n'étant pas aussi fréquentes les unes les autres) :



D'autres espèces sont concernées par cette problématique de transcodage. Le référentiel de transcodage peut être amendé directement via le projet Gitlab de Valparaiso sans avoir à redéployer l'application par le statisticien référent ou un chercheur accrédité connaissant de l'action OBSVENTES (salut Norbert).

VIII.3.3 Contacts-espèce WAO & référentiel PETS

Le jeu de données « contacts-espèces » de l'outil de gestion du plan d'échantillonnage OBSMER/OBSVENTES WAO contient, par sortie, les espèces à observer par ordre de priorité, ainsi que les contacts des navires enquêtés.

Le plan d'échantillonnage fourni par l'outil WAO doit être augmenté du référentiel PETS ainsi que des espèces transcodées afin d'obtenir la liste complète des codes espèces à rechercher dans le RIC et dans les données observées.

VIII.3.4 Données OBSVENTES

Contexte

En 2022, le format « COST » est encore le format d'usage des données OBSVENTES produites et mises à disposition par l'Iremer.

Pour plus d'information sur l'usage et histoire du format COST : A.-S. CORNOU, T. CLOATRE, J. VIGNEAU.

Ce format fut vite jugé insuffisant pour la validation des données au démarrage du volet OBSVENTES de Valparaiso ; bien qu'il fût le seul disponible.

Les tables « tr » (trips) et « sl » de ce format ont été retravaillées. En remplacement de ces deux tables, Valparaiso utilise respectivement les tables « SALE_NAV_STRATA » et « sl_caisse ». Les deux nouveaux jeux de données sont le fruit d'une reprise des requêtes SQL de production des agrégats COST « tr » et « sl » en collaboration avec S. LE BLOND.

« SALE_NAV_STRATA » (pour « Sales, navires et stratification ») correspond aux ventes observées pour l'action « SIH-OBSVENTES » avec des informations complémentaires utiles sur les navires et absentes du format COST (nom du navire, armateur, etc.) ; des informations complètes sur les strates d'échantillonnage provenant directement d'Harmonie ; et également la date-heure de la vente saisie, plutôt que la simple date dans le format COST.

La table « SL_CAISSE » contient les mêmes informations que la table classique « sl » du format COST mais agrégées à la caisse de vente, c'est-à-dire au taxon principal (le premier dans l'arbre de saisie) du lot observé, plutôt qu'à l'espèce scientifique échantillonnée. Cette nouvelle extraction permet notamment de distinguer les caisses de mélange d'espèces, information indispensable notamment pour la confrontation des poids renseignés dans le RIC ainsi que la qualification des espèces.

Le format de données COST sera par ailleurs, à terme, remplacé par un nouveau format de données européen RDBES.

Attention : Correspondance de date WAO-OBSVENTES

Il n'existe hélas dans les données OBSVENTES (c'est-à-dire la saisie Allegro) **aucun lien explicite** qui associe les lots observés avec la sortie à laquelle ils correspondent dans le plan d'échantillonnage. Aucune combinaison de variables ne permet de réaliser cette association de manière exacte et certaine, puisque la date renseignée dans l'outil de saisie et la (les) date de WAO sont différentes :

- WAO décrit la sortie avec **une date de début et de fin d'observation** ;
- Allegro renseigne la **date de vente des lots du navires** (et non d'observation)

Règle de jointure WAO-OBSVENTES actuelle :

- Dans l'algorithme de traitement des données OBSVENTES, la jointure entre les deux dates est faite sur la base d'une plage calendaire définie par un jour min et un jour max;
- La plage actuelle est :
 - Jour min = jour de début d'observation ;
 - Jour max = jour de fin d'observation INCLUS.
- La plage actuelle choisie s'étend sur maximum de 2 jours, dans le cas où la sortie démarre avant minuit et finit après.

Cette règle de décision semble satisfaire l'ensemble des sorties jusqu'ici.

Message d'alerte En cas de un message alertera l'utilisateur en cas de lot observé OBSVENTES orphelin (non raccordé à une sortie du plan d'échantillonnage).

Cas limite encore non rencontré : le jour de vente renseignée dans Allegro (outil de saisie) est postérieur au jour de vente de fin d'observation renseigné dans WAO. Si de tels cas étaient rencontrés, il faudrait adapter la règle de jointure en augmentant le Jour max de la règle (à J + 1, 2... ?).

VIII.3.5 Mélange d'espèces

Les caisses échantillonnées contiennent parfois plusieurs espèces scientifiques différentes. On parle alors de caisses de mélange. Les caisses de mélange sont vendues avec une seule étiquette d'espèce : celle qu'on trouvera normalement dans les données RIC.

Dans les données observées OBSVENTES, la table « sl_caisse » donnera, pour chaque espèce ou taxon de caisse, le détail des espèces scientifiques sous-jacentes.

La prise en compte des caisses de mélange est importante pour le contrôle des lots. Dans un cas de mélange, le RIC associera le poids total de la caisse à l'espèce renseignée sur son étiquette ; le poids échantillonné de cette espèce est alors nécessairement inférieur au poids total de la caisse, qui contient donc plusieurs espèces. L'usage de la table « sl_caisse » permet, dans les données observées, de rattacher les espèces secondaires des mélanges au taxon de la caisse, et ainsi de faire correspondre le poids du RIC et le poids observé.

Présentation finale des espèces : notion d'espèce principale

Les cas de transcodage et de mélange d'espèces scientifiques pouvant se combiner, Valparaiso dégage la notion d'espèce principale pour s'y retrouver dans ce puzzle. **L'espèce principale** est l'espèce demandée par le plan (exactement) ou bien l'une de ses déclinaisons selon le référentiel de transcodage. Pour un lot donné, on peut donc avoir plusieurs espèce scientifiques principales et plusieurs espèces de mélange.

Dans l'onglet « Espèce-navire » les espèces demandées par le plan sont finalement présentées comme suit, dans la colonne « Espèce obsventes » : [Espèce.s principale.s] ([Espèce.s secondaire.s de mélange]).

Tableau : Espèces du navire

	Espèce WAO	Espèce RIC	Nom	Espèces obsVentés	Poids RIC	Cat. RIC	Poids obsVentés	Cat. obsVentés	Contrôle
▶	JOD	JOD	Saint Pierre	JOD	5.7 kg	10 (5.7 kg)	5.7 kg	10 (5.7 kg)	cas_01
▶	MNZ	MNZ	Baudroies nca	ANK, MON (BRB, GUU)	70.272 kg	10 (17.202 kg) / 20 (6.954 kg) / 30 (14.03 kg) / 40 (15.25 kg) / 50 (16.836 kg)	70.275 kg	10 (17.202 kg) / 20 (6.954 kg) / 30 (14.031 kg) / 40 (15.251 kg) / 50 (16.837 kg)	cas_01
				<i>Espèces principales (déclinaison de MNZ)</i>				<i>Espèces secondaires de mélange</i>	
▶	MUR	MUR	Rouget de roche	MUR (BBS)	14.144 kg	20 (5.512 kg) / 30 (8.632 kg)	14.145 kg	20 (5.512 kg) / 30 (8.633 kg)	cas_01
▶	NEP	NEP	Langoustine	NEP	107.7 kg	30 (3.8 kg) / 40 (103.9 kg)	105.008 kg	30 (3.801 kg) / 40 (101.207 kg)	cas_01
▶	POL	POL	Lieu jaune	POL	8.073 kg	30 (8.073 kg)	8.073 kg	30 (8.073 kg)	cas_01

VIII.4 Contrôles des lots « Espèce-navire »

Contrôles des poids et des catégories

Les lots d'espèces sont contrôlés sur la base du RIC, bien que celui-ci soit imparfait ou puisse ne pas rendre compte de la réalité de terrain rencontrée par l'observateur (ex. : lots non exposés sous la criée bien que déclarés et présents dans le flux de ventes).

Ce contrôle compare l'observé et le RIC en terme de :

- Le nombre de catégories (toutes les catégories d'un navire échantillonné doivent être observées) ;

- Le poids total par espèce ;
- Les poids de chaque catégorie ;

avec une marge d'erreur autorisée de 5%. La marge d'erreur pourra être rajustée par l'équipe OBSVENTES via Gitlab².

La comparaison des poids observés dans OBSVENTES et des poids du RIC permettent de dégager plusieurs cas de figure ou « cas-contrôle » de 1 à 17.

Ceci sont pleinement détaillés dans le document thématique dédié (le demander à l'équipe OBSVENTES). La marche à suivre pour les correspondants de façade et pour les prestataires et encore en cours d'élaboration.

On soulignera simplement ici les cas les plus notables, 1 et 17 :

- « 1 » est le cas sans anomalie particulière : toutes les catégories présentes sous créée (selon le RIC) ont été observées ; tous les poids (totaux et par catégorie) correspondent selon une marge d'erreur autorisée ;
- « 17 » est celui où les lots observés ne correspondent à aucun des lots du RIC en première approche. Il peut alors s'agir d'un mauvais rattachement de la sortie et du RIC (immatriculation mal renseignée, non concordance de dates, etc.) détectable facilement en consultant les données du RIC via le module de recherche mis à disposition des correspondants de façade (voir Module de recherche dans le RIC), ou encore de cas de figure plus complexes.

Contrôle des tailles des individus par RTP

Pour contrôler les tailles d'individus mesurés : on compare le poids échantillonné total avec le poids total reconstitué via la RTP à partir des tailles des individus mesurés. Un écart trop important entre le poids reconstitué RTP et le poids échantillonné total peut indiquer des erreurs sur la taille des individus.

Contrôle du nombre d'individus et autres indicateurs

La synthèse des lots d'espèces par bateau met à disposition d'autres indicateurs comme les tailles min et max du lot échantillonné, le poids échantillonné et le poids observé, ainsi que le nombre d'individus mesurés.

Les observateurs devant réaliser un nombre de mesures suffisant par lot échantillonné pour des raisons de représentativité (voir protocole d'échantillonnage OBSVENTES), particulièrement en cas de sous-échantillonnage du lot observé ; le coordinateur est invité à vérifier le nombre absolu de mesures réalisées et juger s'il est suffisant.

Pour cette première mouture, seul le contrôle sur le poids reconstitué par RTP donne lieu à une alerte automatique dans Valparaiso (règle rouge à gauche du tableau).

Une première mouture tournée vers le contrôle des poids Ainsi, cette première version de Valparaiso-OBSVENTES s'est-elle d'avantage tournée (dans le temps imparti) vers le

² https://gitlab.ifremer.fr/sih/sih-data-analysis/valparaiso/-/blob/master/data-raw/valpo_obsventes_param.csv

développement des contrôles des poids et des catégories via le RIC. Un second volet de développement permettrait :

1° D'approfondir l'utilisation du référentiel RTP Harmonie dans OBSVENTES ;

2° De prévoir une fouille plus fine des tailles mesurées et peut-être de dégager une règle de contrôle systématique (sans doute spécifiques pour certaines espèces).

Le retour d'expérience sur les premiers trimestres validés devrait aider à préciser ces opportunités de développement pour, à l'avenir, détecter plus finement les anomalies et réduire le temps des correspondants alloué aux contrôles.

VIII.5 Qualification des lots échantillonnés

Une fois l'ensemble des données OBSVENTES traitées et contrôlées, l'utilisateur dispose de toutes les informations nécessaires pour la qualification des lots.

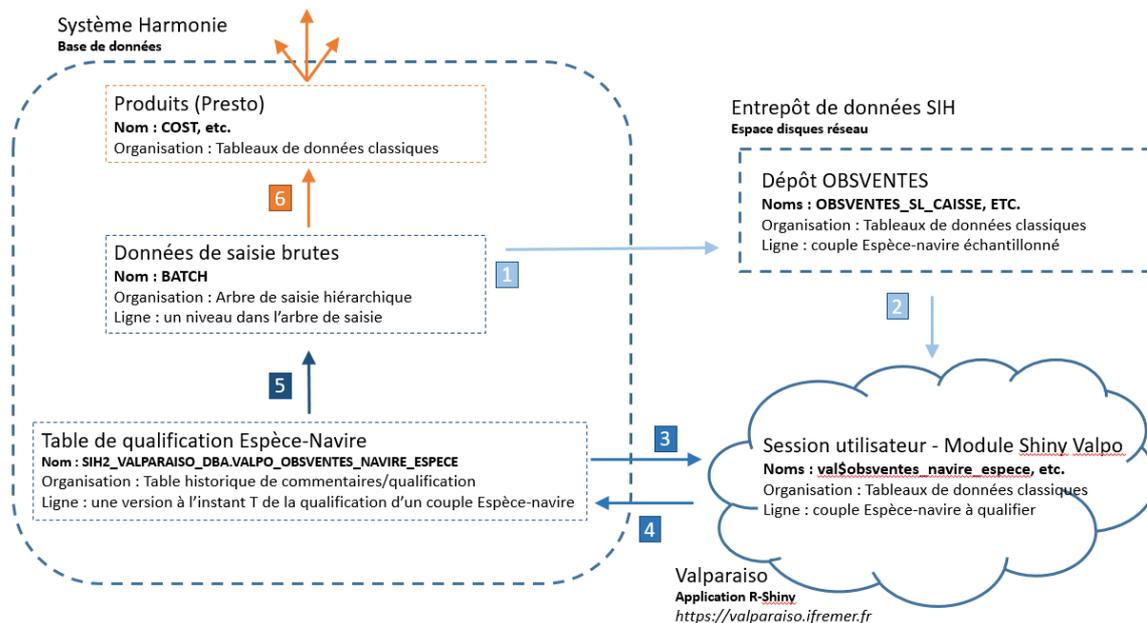
L'action OBSVENTES dispose du cycle de contrôle et de qualification de la donnée le plus complet à ce jour, puisque les champs de qualification des lots renseignés dans la base de connaissance de Valparaiso sont directement transférés aux données brutes de départ.

Ces informations pourront ensuite être intégrées aux produits Harmonie mis à disposition des utilisateurs finaux de la donnée.

VIII.5.1 Schéma complet du cheminement des données

Le schéma suivant propose au lecteur une vision globale du **cheminement des données OBSVENTES** au travers du cycle de qualification via Valparaiso. L'accent sera donc mis sur les données de caisses de ventes observées. L'objectif du schéma n'est pas de tracer toutes les données impliquées dans la validation (voir plus haut : « Jeux de données »).

L'objectif est plutôt d'indiquer les principaux fichiers/jeux de données/tables et bases de données impliqués dans ce cycle, ainsi que la manière dont sont organisées les données des lots dans ces différentes entités.



- 1 Service d'extraction en routine du système Harmonie vers l'Entrepôt de données SIH
- 2 Chargement des données de l'Entrepôt dans l'application Valparaiso
- 3 Lecture de la Base de connaissance - Ici la table de qualification des lots Espèce-navire
- 4 Ecriture dans la Base de connaissance - Ajout de nouvelles lignes à la table
- 5 Transfert des champs de qualification vers les données de départ
- 6 Remontée des champs de qualification vers les produits Harmonie à destination des utilisateurs finaux

Commentaire du schéma

La donnée OBSVENTES circule dans 3 grands « compartiments » :

- Le système Harmonie : **base de données** Oracle ; lieu de bancarisation des données halieutiques, dans laquelle on trouve différents schémas et tables de données.
- L'Entrepôt de données du SIH : un **espace disque réseau** de l'Ifremer, organisé en plusieurs dépôts (dossier lié à un service d'extraction), où des fichiers de données extraites d'Harmonie sont déposés régulièrement.
- Valparaiso : **application R-Shiny** dans laquelle l'ensemble des données nécessaires à la validation sont chargées depuis différentes sources ; les modules Shiny d'interaction avec la base de connaissance forment des sous-environnements dans la session de l'utilisateur ; les données enregistrées en base de connaissance passent par ces modules.

1. Les données OBSVENTES sont extraites du système Harmonie et mises au format classique de tableaux de données.

2. Les données sont chargées dans la session Valparaiso de l'utilisateur depuis l'Entrepôt de données pour y subir les traitements et contrôles nécessaires.
3. Les données de la base de connaissance et de qualification sont également chargées et traitées dans Valparaiso*.
4. Les nouveaux commentaires et champs de qualification des lots sont envoyés à la base de connaissance*.
5. Toute modification de la table de qualification des lots Espèce-navire OBSVENTES est automatiquement détectée et déclenche le transfert des champs de commentaire et de qualification nouveaux vers les données de départ. Ce point sera détaillé plus loin dans la partie : « Transfert des champs de qualification ».
6. Les traitements qui permettent de constituer les produits Harmonie à destination des utilisateurs finaux peuvent ainsi récupérer ou tenir compte de ces champs ; soit en filtrant la donnée à introduire dans le produit, soit simplement en restituant le champ de qualification et le commentaire associé.

* Pour en savoir plus sur le mécanisme de lecture/écriture dans la base de connaissance et sur le fonctionnement de ses tables, se référer au chapitre : Base de connaissance -> Principe d'historicité.

VIII.5.2 Base de connaissance et de qualification

La table de connaissance destinée à la qualification des lots d'espèces OBSVENTES est : SIH2.VALPARAISO_DBA.VALPO_OBSVENTES_NAVIRE_ESPECE.

Pour une description générale du fonctionnement de de la mise en place de la base de connaissance Valparaiso et des groupes de colonnes récurrents, voir le chapitre [Base de connaissance VALPARAISO](#).

Description des champs de qualification et des lots observés OBSVENTES

Groupe	Nom du champs	Valeur.s	Description
[USER]	[...]	[...]	[...]
VALPO	VALIDE	"Bonne", "Douteuse", "Fausse", NULL	Champ de qualification de l'espèce scientifique observée ; la saisie de ce champ est contrainte dans l'application et est reserve aux seuls "Validateur" IFREMER
	COMMENTAIRE	Chaîne de caractères libres	
ENTITE CONTROLEE	PROGRAM_FK	"SIH-OBSVENTE"	Code Harmonie de l'action OBSVENTES
	ANNEE		Année d'observation
	PROGRAMME_CODE		Description de la ligne de plan et des strates d'observation
	PLAN_CODE		
	NOM_STRATE		
	LIEU_COD		Code lieu (Criée-Point de débarquement) de l'observation
	DATE_SORTIE		Jour de la vente observée, soit SAMPLE_DATE de la table SALE arrondie au jour* (voir Remarque 2)
	VESSEL_FK		ID du navire
TRIP_NUMBER		ID Harmonie de la vente observée (SAMPLE_FK); le terme trompeur de "trop" est hérité du format de données COST.	
ESPF_COD_RIC		Code FAO Espèce commerciale, ou taxon principal du lot observé	
ENTITE FINALE CONTROLEE	ESPF_COD_OBSVENTES		Code FAO Espèce scientifique observée* (voir Remarque 2)

Remarque 1 La table de connaissance ne garde pas trace du résultat du contrôle (ou « cas-contrôle » : cas 1, 2 etc.) dans la mesure où :

1. Ceux-ci restent évolutifs;
2. La qualification ne se fait pas uniquement via ce critère de contrôle, mais également sur la reconstitution des poids via la RTP (bien qu'aujourd'hui imparfaite), et de nouveaux contrôles sur les lots pourraient à l'avenir entrer en ligne de compte.

* **Remarque 2** Pour consolider d'avantage cette table, il eût mieux valu conserver la date-heure complète de la vente (SAMPLE_DATE) : réflexe de travail hérité du format de données COST... ; également : conserver le nom complet de l'espèce scientifique.

VIII.5.3 Règle de qualification

La qualification des lots d'espèces observées est réalisée par les correspondants de façade de l'Ifremer selon des règles thématiques qui font encore l'objet de débat concernant les écarts constatés avec le RIC.

L'expertise terrain des correspondants, la synthèse de tous les éléments de contexte et de toutes les données disponibles pour résoudre les « cas-contrôle » les plus complexes, ainsi que le dialogue avec les prestataires ayant réalisés l'échantillonnage, le surgissement de cas de figure nouveaux ; tout ceci empêche de qualifier *in fine* de manière automatique les lots OBSVENTES.

Les commentaires laissés dans Valparaiso par les prestataires et les correspondants de façade fourniront de précieux éléments de contexte pour l'évolution future de l'application, l'élaboration de contrôle plus fins et détaillés selon les cas-contrôles, ainsi que l'affinage des règles de décision pour la qualification des lots.

Pour en savoir plus sur ces règles de décision, se rapprocher de l'équipe OBSVENTES.

Droits et champs réservés

Le champ de qualification « Qualification (programme) » est ouvert aux personnes de l'Institut qui ont un profil « Validateur » associé au programme SIH-OBSVENTE dans le système Harmonie. En pratique, les personnes en charge de cette qualification sont les correspondants de façade.

Les prestataires qui disposent du profil « Validateur » pour SIH-OBSVENTE auront droit en écriture dans le champs « Commentaire (Société) » uniquement.

VIII.5.4 Transfert des champs de qualification et de commentaire Programme vers les données brutes Harmonie

L'objectif de la qualification des lots OBSVENTES dans Valparaiso est double

1. Quantifier la qualité du travail réalisé pour le suivi administratif ;
2. Transmettre l'information sur la qualité des données aux utilisateur finaux (scientifiques)

VIII.6 Contrôles et qualification des sorties

Couverture d'observation des catégories

L'ensemble des catégories commerciales présentes sous criée doivent idéalement être observées pour des questions représentativité et d'élévation future des données.

Ce souci doit guider les observateurs dans le choix des navires qu'ils vont effectivement échantillonner.

Pour juger de la bonne couverture des catégories présentes sous criée le jour de la sortie, des statistiques sont mises à disposition dans l'onglet « Sortie » en dépliant les différentes espèces demandées.

Qualification des sorties OBSVENTES

Une base connaissance VALPO servant à la qualification des sorties (valide/invalid) est d'ores et déjà déployée.

Cependant, les règles et les indicateurs nécessaire à celle-ci sont encore en cours d'élaboration (% de lots qualifiés bons, degré quantifié de respect du plan, etc.).

VIII.7 Perspectives d'évolution

Utilisation plus précise du référentiel RTP pour le contrôles des tailles

Contexte Le référentiel RTP détaille, selon les espèces, la RTP au niveau de la Zone (en mer) ; les coefficients de la RTP peuvent varier de façon significative avec la zone. L'information de la zone de pêche est présente dans le logiciel de saisie Allegro et est utilisée dans ce dernier pour proposer au saisisseur une clé taille-poids précise.

En revanche, les données extraites d'Harmonie pour la validation ne dispose pas de cette information. En conséquence, les coefficients RTP mobilisés pour le contrôle RTP dans l'application sont moyennés sur les zones. Seule la clé : Espèce, Sexe, Trimestre est utilisée.

Solution Pour utiliser le référentiel RTP à l'échelle de la Zone, la zone de pêche doit être extraite des données exportées du logiciel de saisie dans le système Harmonie et ajouté à la table « sl caisse ». Cette variable devra être identifié dans le système et la requête retravaillée.

Calcul des indicateurs de qualification en direct

Faire évoluer les modules de qualification pour analyser en directe les nouveaux champs renseignés et recalculer à la volée les indicateurs de suivi de la qualification.

Transfert des champs de qualification vers des produits Harmonie

Les champs de qualification de la base de connaissance VALPO_OBSVENTES_ESPECE_NAVIRE ont pour vocation d'être mis à disposition des utilisateurs finaux de la donnée.

Ce chantier doit être ouvert avec l'équipe informatique, avec laquelle la discussion est déjà engagée sur ce sujet (G. PRIGENT).

La table de connaissance contient normalement les clés nécessaires pour effectuer ce transfert. Reste la règle de qualification par défaut comme « Bonne » les lots d'espèces qui ne figurent pas dans la base de connaissance (ou laissé vide) par souci d'économie de temps consacré à la qualification pour les correspondants de façade.

Contraintes de chaque partie

- Equipe informatique : souhaiterait que tous les lots soient qualifiés pour ne pas valider des lots qui n'auraient pas encore été vu/corrigés ; plus simple et plus propre.
- Equipe OBSVENTES, correspondants de façade: souhaite qualifier les lots anormaux seulement et considérer « bons par défaut » les autres.
- Valparaiso : le module générique de base de connaissance de Valparaiso est conçu pour n'enregistrer des informations que manuellement, pour des raisons de parcimonie et d'ergonomie ; il n'est donc pour l'instant pas question de remplir automatiquement des champs et de les envoyer également automatiquement en base.

Compromis dans la date Le processus de contrôle/révision des lots anormaux prenant un temps incompressible de vérification et d'échange avec le prestataire, le transfert des champs de qualification ne peut, de toutes les manières, être envisagé qu'après une date butoir à définir (trimestre, semestre, année...). Charge à l'équipe thématique de s'engager à ce que tous les lots « anormaux » aient été passé en revue et qualifié. Un indicateur du nombre de lots « Anormaux sans qualification », donc à qualifier impérativement, existe dans Valparaiso afin de contrôler qu'aucun lot en erreur ne soit laissé de côté.

Gestion des lots orphelins

Les données du RIC et d'Allegro ne correspondent pas toujours. La clé de jointure complète de ces deux jeux de données opérée pour les contrôles est : Ligne de plan + Navire (immatriculation) + Lien de vente + Date de vente + Espèce (celle de la caisse).

Remarque : Les catégories commerciales ne font pas partie de la clé de jointure entre le RIC avec l'OBS. L'objectif est dans un premier temps de rattacher la vente et l'observé au niveau de l'espèce. Les catégories commerciales interviennent dans un second temps à des fins de contrôle des poids vis-à-vis du RIC et de l'étendu de l'échantillonnage. On conserve ainsi pour les contrôles, toutes les catégories du RIC (même celle non observées) et réciproquement l'ensemble des catégories observées, même celle qu'on ne retrouve pas dans le RIC.

L'absence de correspondance entre le RIC et Allegro peut être due à un ou plusieurs termes de cette clé, en particulier :

- Le Navire : l'immatriculation d'un navire peut être mal renseignée dans l'outil de saisie, notamment lorsque deux navires ont des noms très proches ;
- Le lieu : il arrive que des lots soient (ex. des ventes à Noirmoutier relocalisées dans les données aux Sables-d'Olonne) ;

- Date de vente : il arrive que les lots d'un navire soient vendus de façon fractionné ; c'est le cas le plus délicat à gérer, qui mériterait un algorithme de recherche dédié.

Module de recherche dans les ventes

Les correspondants de façade disposent d'un module de recherche libre dans les ventes qui leur permet de s'affranchir de cette clé et souvent de retrouver dans le RIC des lots orphelins.

Les retours d'expérience des correspondants devraient permettre d'affiner la jointure entre les données du RIC et d'observation en prenant en compte les spécificités locales des criées.

Filtrer les cas contrôles pour les prestataires

Le travail actuel d'élaboration des règles de qualification des lots pour les correspondants, et le recul des premières campagnes de qualification, devrait permettre d'identifier les « cas-contrôles » qui concernent plus directement les prestataires.

Grace au profil de connexion de Valparaiso, il serait alors possible :

1. De ne leur présenter que les cas qui les concernent directement ;
2. De leur présenter des indicateurs sur le « reste à faire » mieux adapté (les indicateurs présentés actuellement dans l'onglet « Synthèse » sont d'avantage dédiés à l'avancement de la qualification à destination des correspondants de façade).

Remarque : en méditerranée, les cas 17 et 16 sont par ailleurs passer « cas 01 » c'est-à-dire bons automatiquement.

IX. OBSBIO

IX.1 Jeux de données

Données OBSBIO :

- Paramètres biologiques : ensemble des données de paramètres biologiques, tous programmes de collecte ou campagnes halieutiques confondus ;
- Estimation d'âge : extraction complète de la base de données SMARTDOT (différente d'Harmonie) contenant toutes les estimations d'âge réalisées, y compris celles qui n'ont pas été retenue comme âge de référence transféré dans Harmonie ;
- Plan d'échantillonnage : suivi des lignes de plan OBSBIO avec taux de réalisation des échantillons, description des paramètres biologiques à mesurer, secteur en mer à cibler, etc.

Données	Dossier ./OBSBIO
Paramètres biologiques	PARAMETRES_BIOLOGIQUES
Estimation d'âge	ESTIMATION_AGE
Plan d'échantillonnage	SUIVI

Source : Entrepôt de données SIH

Référentiels Harmonie :

- Référentiel Espèce : référentiel espèce FAO ;
- Ports, criées, quartiers : description des lieux à terre : criées, port-points de débarquement et quartier maritime ;
- Relation Taille-Poids : référentiel des coefficients a et b des relation taille-poids d'Harmonie ;
- PSFM : extraction de tous les PSFM, combinaison de « Paramètre, s..., fraction, méthode » des mesures ;
- Hiérarchie : hiérarchie des secteurs en mer.

Données	Dossier ./REFERENTIELS
Référentiel espèces	ESPECE_FAO
Ports, criées Quartier maritime	LIEU_PORT_CRIEE_QUARTIER
PSFM	PSFM
Hiérarchie	HIERARCHY

Source : Entrepôt de données SIH

IX.2 Ciblage des espèces et échantillons « OBSBIO »

La liste des espèces proposées au chargement est issue d'une requête directe en base Harmonie.

Les espèces listées sont objectivement celles qui disposent d'échantillons tombant dans le périmètre de validation des données OBSBIO, à savoir :

- Les échantillons du programme « SIH-OBSBIO » (échantillonnage en criée, achat de poisson, etc.) ;
- Les échantillons de campagnes halieutiques *avec estimation d'âge uniquement*.

La requête est relancée chaque fois que l'utilisateur change d'année de validation dans la barre du haut.

IX.3 Chargement et traitement des données

Requêtes et chargement au dépôt de données

Les données de l'année de validation N sont toujours requêtées directement en base Harmonie. De cette manière, une modification des données via les outils de saisie se répercute directement dans l'application après re-chargement des données.

Le plan d'échantillonnage (description et réalisation) est chargé depuis l'entrepôt de données.

Un jeu de données antérieures à l'année de validation N (jusqu'à N-3) est chargé depuis l'entrepôt de données du SIH, rafraîchi en moyenne toute les 12 heures.

Traitement des variables Taille et Poids

L'application ne préjuge pas du ou des paramètres de Taille des échantillons présents dans les données.

Si plusieurs types de mesure de Taille sont présentes, le choix est donné à l'utilisateur d'utiliser tel ou tel type de mesure pour les onglets graphiques.



Exemple : Deux types de longueur ont été utilisés pour le bar en 2022 : « Longueur totale » et « Longueur standard ».

Homogénéisation des unités de mesure de Taille et de Poids L'application ne préjuge pas non plus de l'unité de mesure de la Taille employée (cm ou mm) ni de celle du Poids (g ou kg) ; cela dépendra notamment de l'espèce. En outre, pour une même espèce, différentes unités de mesure sont parfois utilisées.

Les variables Taille et Poids sont donc homogénéisées, respectivement en mm et cm, et en g et kg, afin de toujours offrir à l'utilisateur une représentation complète des échantillons sur les graphiques (onglets « Classes de taille », « Relation Taille-poids » et « Relation Taille-âge »).

L'homogénéisation des unités de mesure est appliquée et pour chaque paramètre de taille présent dans les données (cf. Exemple précédent : les deux types de longueurs sont proposées à l'utilisateur en cm et en mm).

L'unité originale (saisie) du poids et de la taille est fournie dans le tableau de donnée qui s'affiche sous le graphique en cliquant sur un échantillon en particulier.

Contrôles implémentés

Seules les données de l'année N (année de validation) sont contrôlées, au regard du plan d'échantillonnage. Les données antérieures sont simplement illustratives et destinées à enrichir les nuages de points des graphiques de RTP et RTA avec une dimension historique.

Paramètres à mesurer

Plusieurs types de mesure (pour la taille par exemple) sont parfois autorisées par la ligne de plan (ex. : Exemple précédent, bar 2022). Dans ce cas, au moins un paramètre de taille doit être renseigné pour passer ce contrôle.

Les résultats de ce contrôle figurent dans l'onglet « Paramètre à mesurer ».

Régionalisation des échantillons - Secteurs en mer

Les zones d'observations renseignées doivent être autorisées par le plan.

Pour une même ligne de plan, plusieurs zones d'observation de différents niveaux géographiques peuvent être autorisées (division CIEM, rectangle statistiques, etc.). De

même, plusieurs zones de différents niveaux géographiques peuvent être associées à un échantillon dans l'outil de saisie.

Chaque secteur saisi est contrôlé individuellement ; on vérifie, grâce à la table de hiérarchie Harmonie, que chacune des zones renseignées pour un échantillon appartient à au moins une des zones autorisées par la ligne de plan rattachée à l'échantillon.

Règle de contrôle : On considère comme anormal un échantillon dont au moins une des zones associées n'appartient à aucune zone décrite par la ligne de plan ; autrement dit, toutes les zones renseignées, quel que soit leur niveau géographique et leur nombre, doivent être comprises dans le périmètre décrit par les zones de la ligne de plan.

Les résultats de ce contrôle figurent dans l'onglet « Zones d'échantillonnage ».

IX.4 Contrôles visuels des variable Taille, Poids, Âge

Le contrôle des paramètres : **Taille**, **Poids** et **Âge** se fait graphiquement au moyen des nuages de points projetés dans les onglets : « Relation Taille/poids », « Relation Taille/âge ».

Une fois les données chargées et mises en forme, l'utilisateur dispose de l'ensemble des échantillons disponibles de l'année N de validation et d'un jeu de données historiques à titre illustratif.

Il revient à l'utilisateur, à l'aide des options mises à sa disposition, de filtrer les données et d'illustrer les nuages de points convenablement afin d'obtenir une représentation qu'il jugera pertinente pour le contrôle visuel des échantillons.

Filtres à gauche

Les filtres à gauche permettent de cibler les données et de ne pas représenter tous les échantillons à la fois sur le même graphique. On sélectionne un sous-ensemble des échantillons.

Les filtres à gauche s'appliquent sur l'ensemble des données de l'année N et des données antérieures.

Pour désactiver les filtres, désélectionner l'ensemble des critères et cliquer de nouveau sur « Filtrer les échantillons ».

Les critères actifs appliqués aux échantillons observés sont rappelés pour mémoire à l'utilisateur dans le titre de la fenêtre en haut à gauche, après le nom de l'espèce et l'année de validation.

Les filtres appliqués se répercutent sur tous les onglets graphiques.

Options d'illustration/comparaison des échantillons

Les options en haut à droite permettent de :

- Comparer les échantillons (ou le sous-ensemble d'échantillons si des filtres ont été appliqués au préalable) selon un critère donné (Trimestre, Sexe, etc.) ;

- Changer d'unité pour les variables de Taille (cm/mm) et de Poids (g/kg) ;
- Changer de type de mesure.

Exemple : filtrer les données par sexe (ex. : Femelle uniquement) pour ne pas les mélanger sur le même graphique ; puis, illustrer sur le graphique les différences entre trimestre ou autre pour pouvoir les comparer.

IX.5 Sélection et « suppression » d'échantillon

Les options de suppression (retrait) d'échantillon ou de sélection sont transversales, c'est-à-dire valable pour les trois onglets graphiques « RTP », « RTA » et « Inter-calibration ».

Sélection

En cliquant sur un échantillon, l'utilisateur le mettra en surbrillance dans tous les onglets graphiques et affichera les informations détaillées sur les mesures réalisées sous forme de tableau en dessous des graphes.

Suppression d'échantillon

Cette option est utile lorsque l'on souhaite retirer du nuage de point un échantillon qui gêne la représentation globale, qu'il soit extrême et manifestement faux, ou qu'on souhaite évaluer son influence sur la RTP par exemple.

L'échantillon ne sera à aucun moment supprimer de la base de données.

Pour retirer un échantillon de la représentation graphique : cliquer sur celui-ci, puis sur « Retirer l'échantillon » ; son identifiant sera ajouté à la liste des échantillons retirés. Pour le réintégrer aux graphiques : le sectionner dans cette liste, puis cliquer sur « Réintégrer les échantillons ».

IX.6 Inter-calibration d'âges

La base complète des données d'âgeage de l'outil SMARTDOT permet de confronter l'âge donné par différents lecteurs pour le même échantillon. Cet exercice de confrontation des lecteurs d'âge est appelé : « inter-calibration » d'âges.

Droits restreints

Le module d'inter-calibration d'âge est ouvert aux utilisateurs « Validateur » pour le programme SIH-OBSBIO de l'IFREMER. Dans la pratique ce profil n'a été accordé qu'aux coordinateurs et développeurs du projet Valparaiso ainsi qu'aux personnes du pôle de scléro-chronologie de Boulogne-sur-Mer concernées.

Graphique

L'utilisateur est invité à choisir lui-même le lecteur de référence ; car l'application ou le système Harmonie n'ont aucun moyen de le connaître.

L'âge des abscisses (x) est l'âge donné par le lecteur de référence ; en ordonnées (y) sont projetés l'ensemble des âges donnés par les autres lecteurs.

Les échantillons sont légèrement « secoués » selon l'axe x, c'est-à-dire horizontalement, afin qu'ils ne se retrouvent pas tous superposés, et que l'utilisateur puisse les distinguer pour éventuellement les sélectionner.

Tableau de synthèse

Le tableau de synthèse situé en dessous du graphique contient la même information que le graphique mais sous forme de matrice de dénombrement des échantillons, agrémentée du pourcentage d'accord entre les lecteurs, par âge et pour l'ensemble des données. Le tableau est exportable en CSV.

X. Annexes

X.1 Acronymes

ISI : Unité du département IRSI de l'Ifremer en chargée entre autre de l'assistance au développement des outils.

RIC (1) : Réseau inter-créées ; les données du RIC permettent le contrôle des lots d'espèces échantillonnées par l'action OBSVENTES.

RIC (2) : Unité du département IRSI de l'Ifremer en charge de l'infrastructure informatique, en particulier du déploiement des outils.

RTA : Relation taille-âge.

RTP : Relation taille-poids.

SACROIS³ : Jeu de données consolidé de « croisement [...] des marées VMS, des marées déclarées et des ventes en créées ».

SIH : Système d'informations halieutiques.

³ Documents et extraction de SACROIS : <https://sih.ifremer.fr/prive/Acces-aux-donnees/Extractions/Sacrois>