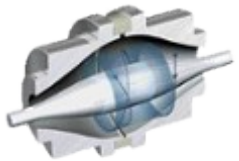
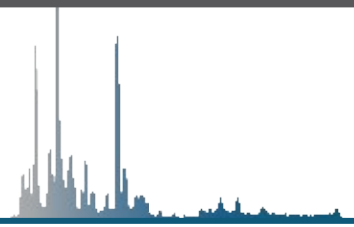
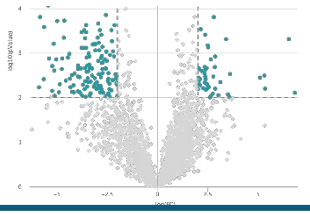
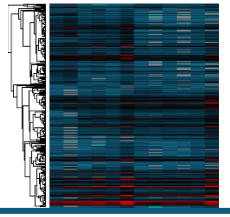
Retour d'expérience

Gestion des données de protéomique

PPM
Pôle Protéome
de Montpellier

Oana Vigy

oana.vigy@fpp.cnrs.fr

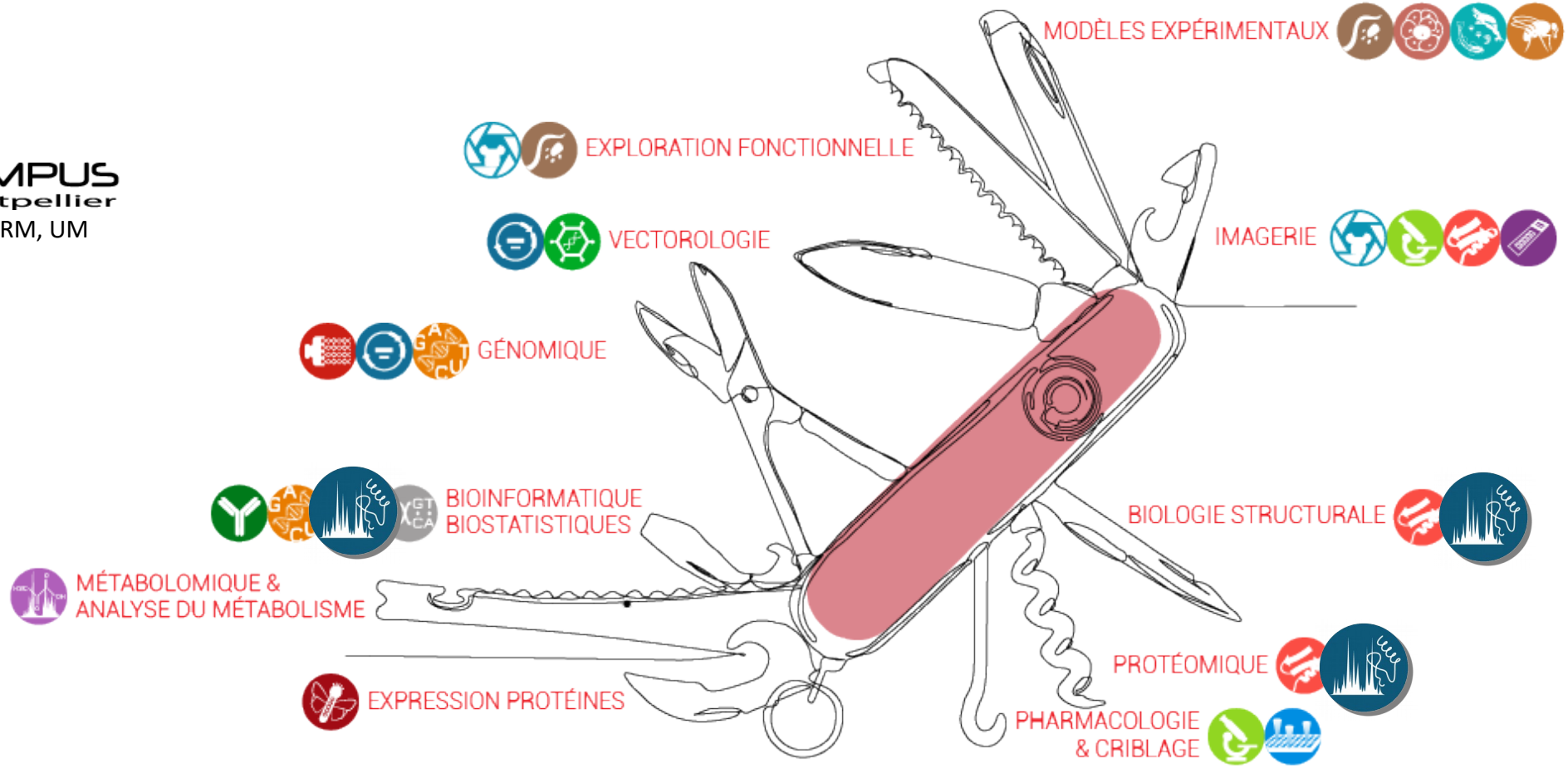
CC BY SA

PPM : la plateforme de protéomique de l'UAR BioCampus

18 plateformes de recherche et 11 familles technologiques

BioCampus Montpellier

UAR CNRS, INSERM, UM

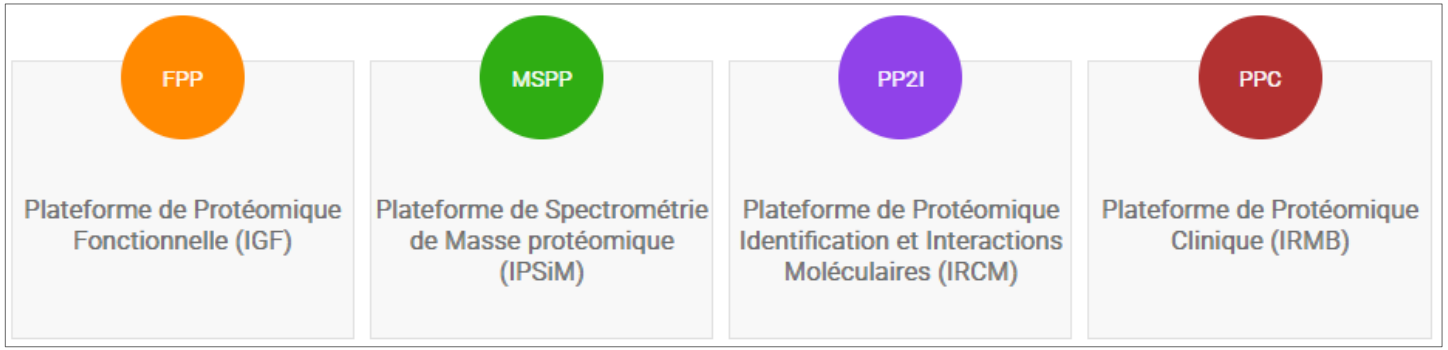


PPM : multi-sites, multi-tutelles et prestations multiples

https://ppm.cnrs.fr/

Structuration régionale des capacités technologiques de l'analyse protéomique

- 4 sites à Montpellier
- 27 ingénieurs et chercheurs (~ 20 ETP)
- Moyenne de 130 projets traités sur les 3 dernières années
- Acquisitions et analyses des données réalisées par le personnel plateforme



IBiSA *depuis 2009*

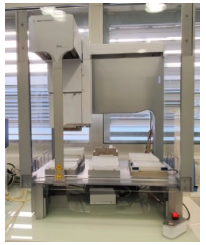
ISO 9001 : 2015 *depuis 2009*

Recherche, expertise et prestations en protéomique dans les domaines de la biologie et de la santé

- Intégration de PROFI-PA (oct. 2022) Plateformes Associées à l'infrastructure nationale en protéomique

PPM : technologies disponibles

Équipements dédiés



Bravo Robot System

NanoLC
HPLC

Chromatographie liquide

Orbitrap

Q-Tof

Préparation d'échantillons

Spectrométrie de masse

Triple Quadrupole

MESO scale discovery QuickPlex SQ 120

Immunodosages

Bioinformatique

Simoa Quanterix HD-X

Bio-Plex

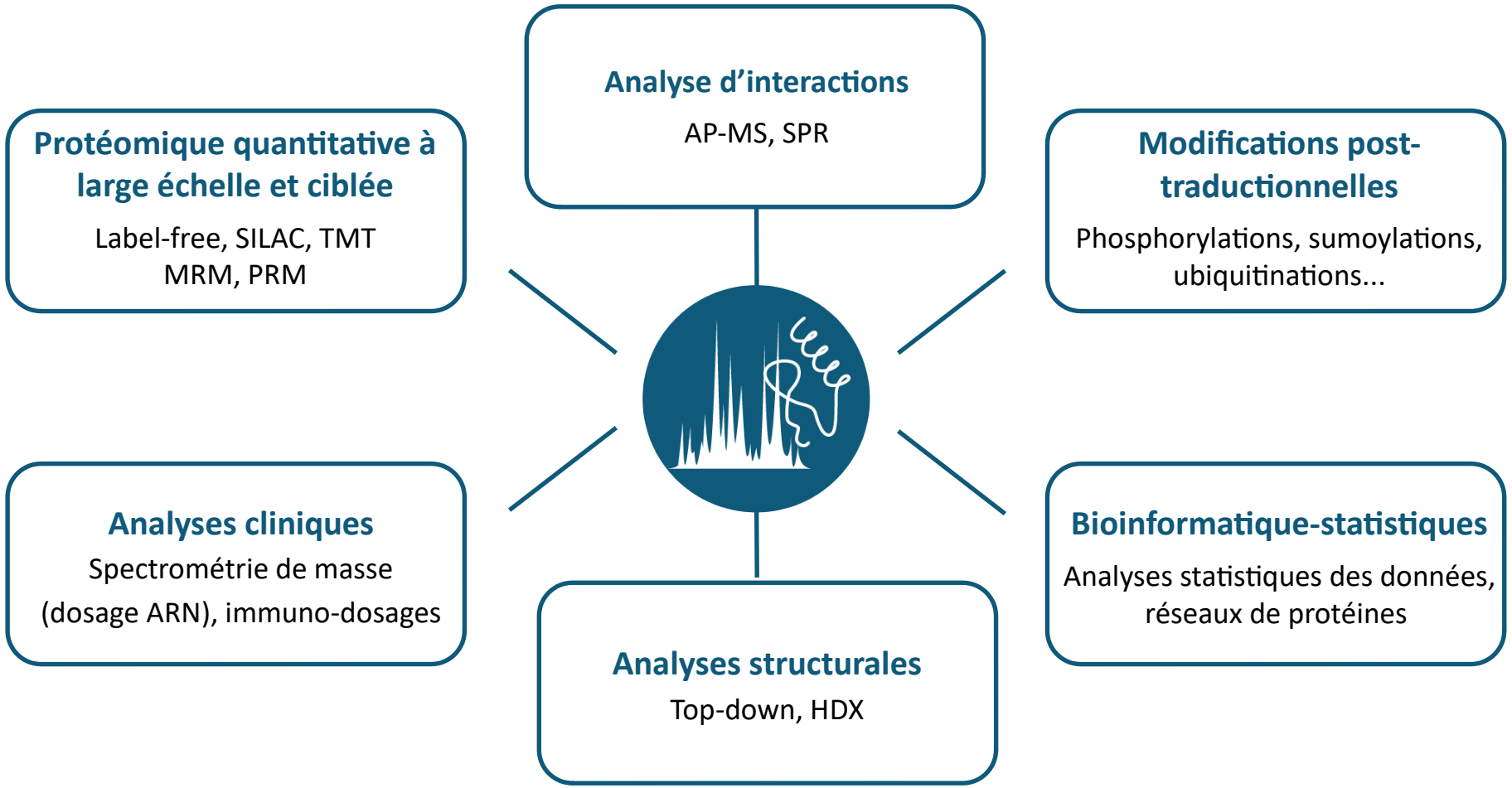
Résonance Plasmonique de surface

NanoTemper

Biacore T200

PPM : principales prestations offertes

Porfolio technique très large



Problématiques / Besoins → Objectif LIMS 2010

Gestion des données

- Projet, échantillon, analyse et résultats
- CR réunions, mails échangés, fiches échantillons, factures, .raw, .dat, .msf, etc.
- Collaborateurs
- Protocoles standards et version utilisée pour les analyses
- Gestion des communications liées aux projets

Génération de rapports

Gestion du matériel

- Machines (révision, maintenance, étalonnage, contact fournisseur, version du logiciel, anomalies)
- Parc informatique (ordinateurs, logiciels, versions installées)

Gestion des stocks et des commandes

Planification

Problématiques / Besoins → Objectif LIMS 2010

Gestion des données

- Projet, échantillon, analyse et résultats
- CR réunions, mails échangés, fiches échantillons, factures, .raw, .dat, .msf, etc.
- Collaborateurs
- Protocoles standards et version utilisée pour les analyses
- Gestion des communications liées aux projets

Génération de rapports

Gestion du matériel

- Machines (révision, maintenance, étalonnage, contact fournisseur, version du logiciel, anomalies)
- Parc informatique (ordinateurs, logiciels, versions installées)

Gestion des stocks et des commandes

Planification

Cahier des charges

- Centraliser les informations *via* un outil unique
- Partager l'information au sein de la plateforme
- Harmoniser la saisie et la présentation des données
- Permettre une exhaustivité et une non-redondance des données
- Faciliter les activités support par l'édition de rapports

Considérations pratiques

- Évolution constante des outils d'analyse (pas de chaîne de traitement / pas d'intégration des données MS dans le LIMS)
- Non- redondance de stockage
- Non- redondance de saisie
- Sécurité : restriction d'accès par projet

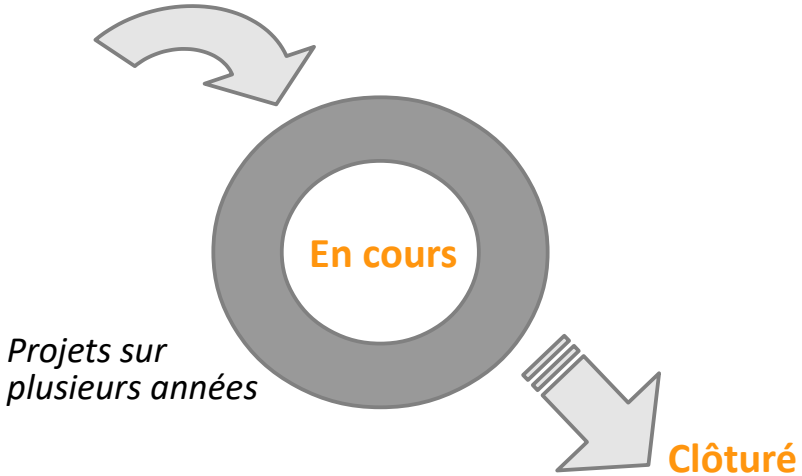
→ Mes missions !

- Maintenir et faire évoluer les outils bioinformatiques pour l'analyse de données de protéomique
- Maintenir et moderniser le système d'information

Cycle de vie des projets (et des données)

Gestion de projet *(qualité)*

- Suivi et planification des projets
- Accès
- Conservation/Destruction *(suppression des données 2 ans après la clôture du projet)*



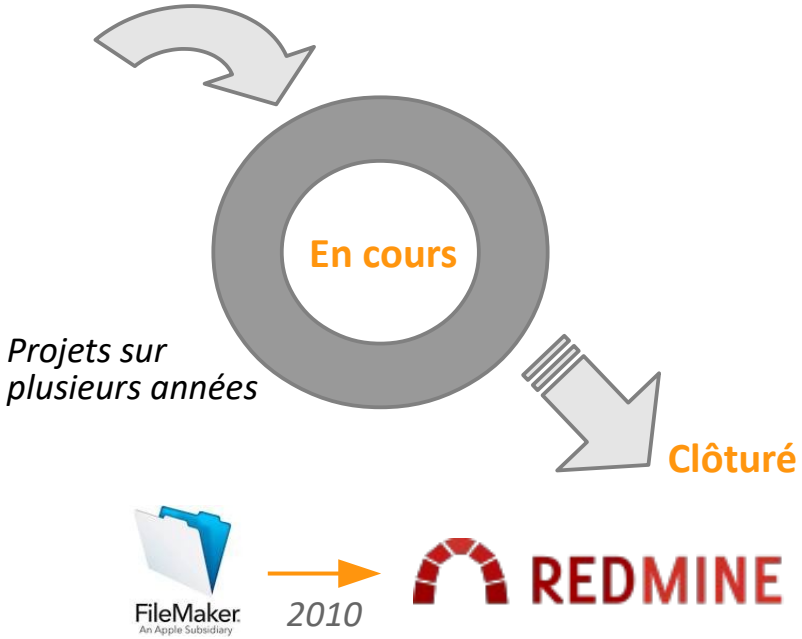
FileMaker → *2010* → REDMINE

Changement de logiciel et évolution régulière de la politique de gestion des données

Cycle de vie des projets (et des données)

Gestion de projet *(qualité)*

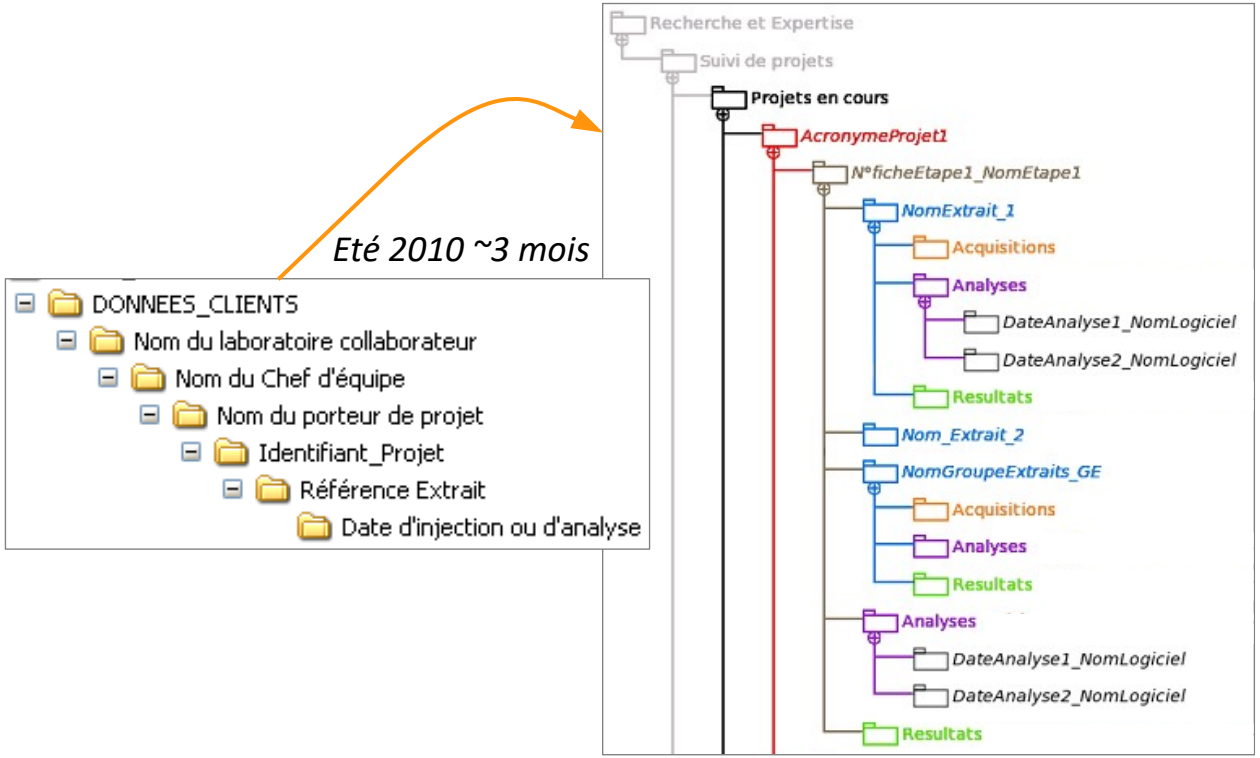
- Suivi et planification des projets
- Accès
- Conservation/Destruction *(suppression des données 2 ans après la clôture du projet)*



Changement de logiciel et évolution régulière de la politique de gestion des données

Réorganisation des données

- Distinguo entre les données R&E / support



Convention d'identification du matériel biologique et des étapes expérimentales

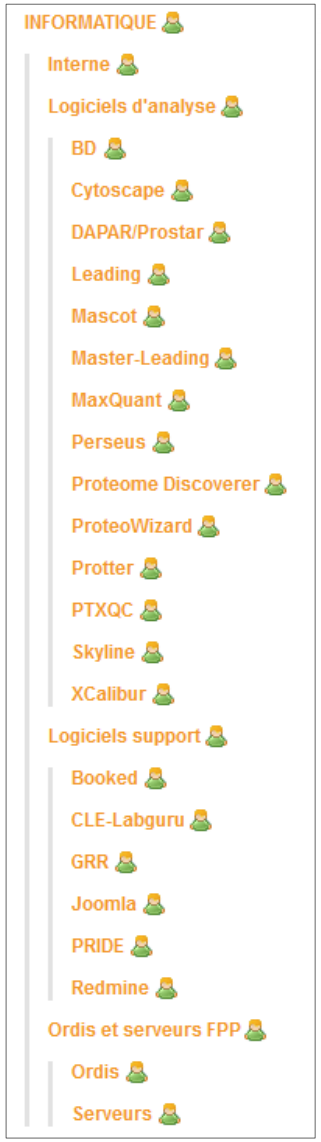
- ACRONYME_date

Gestion du matériel

Gestion du matériel *(qualité)*

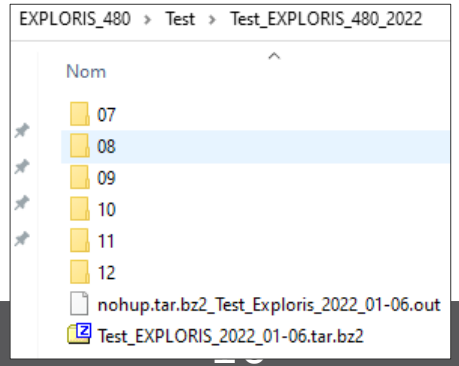
- Suivi des instruments (révision, maintenance, anomalies, etc.) et informatique (logiciels et versions installées)

REDMINE



Gestion des données « matériel » *(contrôle qualité)*

- Evolution de la politique de gestion des données (+2 To / -2 To chaque année, conservé n-2)



Gestion du matériel

Suivi de l'occupation machine

- Assurer la traçabilité de l'entretien des machines et du contrôle qualité des acquisitions
- Calculer le temps d'occupation des machines par type d'utilisation (manip, test, maintenance, indisponibilité, SAV)
- Recueillir des informations supplémentaires qui peuvent être utiles (nombre d'injection, de blanc, d'intervention de changement de colonne) pour faciliter le calcul d'indicateurs et le suivi des machines

2023

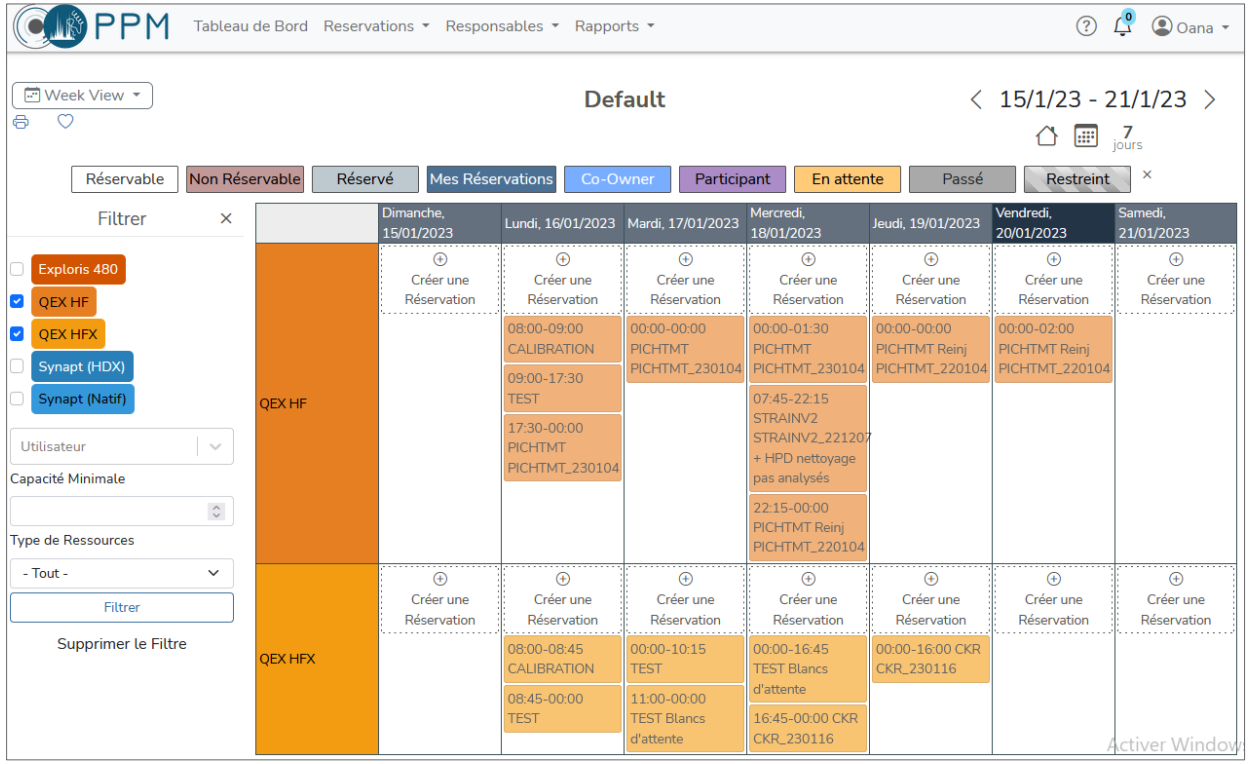
booked

→ Saisie à posteriori des créneaux d'utilisation des machines par les utilisateurs machine

			Manip					*Test*			*Maintenance*		
	Resource type	**Type**	*NB acquisition*	*NB réinjection*	*NB blanc*	*NB aiguille*	*NB biospin*	*Ref extrait*	*NB standard*	*QC NB*	*NB wash*	*Changement de colonne*	N° fiche Redmine
Exploris_480	BottomUp	[*Manip* / *Test* / *Maintenance* / Indisponibilité / SAV]	x	x	x			x	x	x		x	x
QEX_HF			x	x	x			x	x	x		x	x
QEX_HFX			x	x	x			x	x	x		x	x
Synapt_Natif	ProtStruct_Natif		x			x	x	x				x	x
Synapt_HDX	ProtStruct_HDX		x		x			x			x	x	x

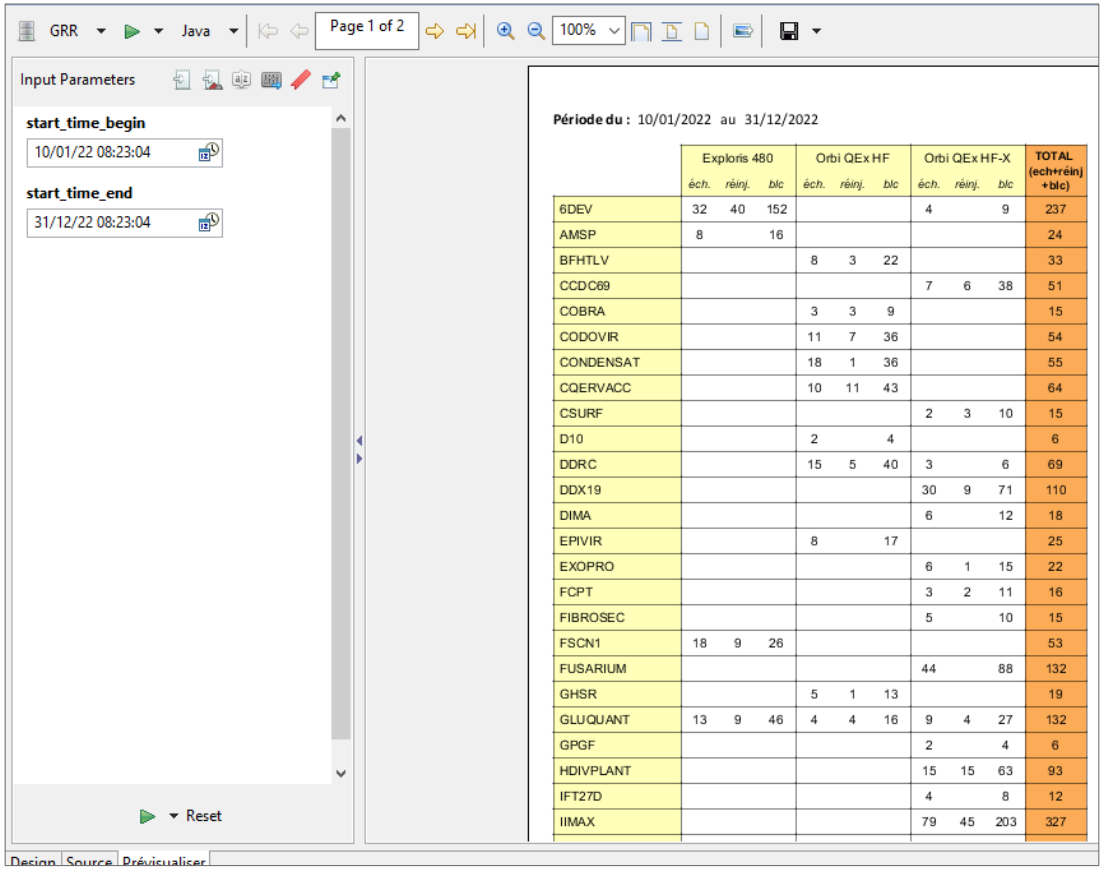
Gestion du matériel

Suivi de l'occupation machine → pour tous les sites ?



→ Migration des données GRR/Booked

Rapports *(db GRR/Redmine)*



	Exploris 480			Orbi QEx HF			Orbi QEx HF-X			TOTAL (ech+réinj +blc)
	éch.	réinj.	blc	éch.	réinj.	blc	éch.	réinj.	blc	
6DEV	32	40	152				4		9	237
AMSP	8		16							24
BFHTLV				8	3	22				33
CCDC89							7	6	38	51
COBRA				3	3	9				15
CODOVIR				11	7	36				54
CONDENSAT				18	1	36				55
CQERVACC				10	11	43				64
CSURF							2	3	10	15
D10				2		4				6
DDRC				15	5	40	3		6	69
DDX19							30	9	71	110
DIMA							6		12	18
EPIVIR				8		17				25
EXOPRO							6	1	15	22
FCPT							3	2	11	16
FIBROSEC							5		10	15
FSCN1	18	9	26							53
FUSARIUM							44		88	132
GHSR				5	1	13				19
GLUQUANT	13	9	46	4	4	16	9	4	27	132
GPGF							2		4	6
HDIVPLANT							15	15	63	93
IFT27D							4		8	12
IIMAX							79	45	203	327

Système d'information

Système d'information : ensemble organisé de ressources : matériel, logiciel, personnel, données, procédures permettant d'acquérir, traiter, stocker, communiquer des informations (sous forme de données, textes, images, sons, etc…) dans des organisations. *(Reix)*

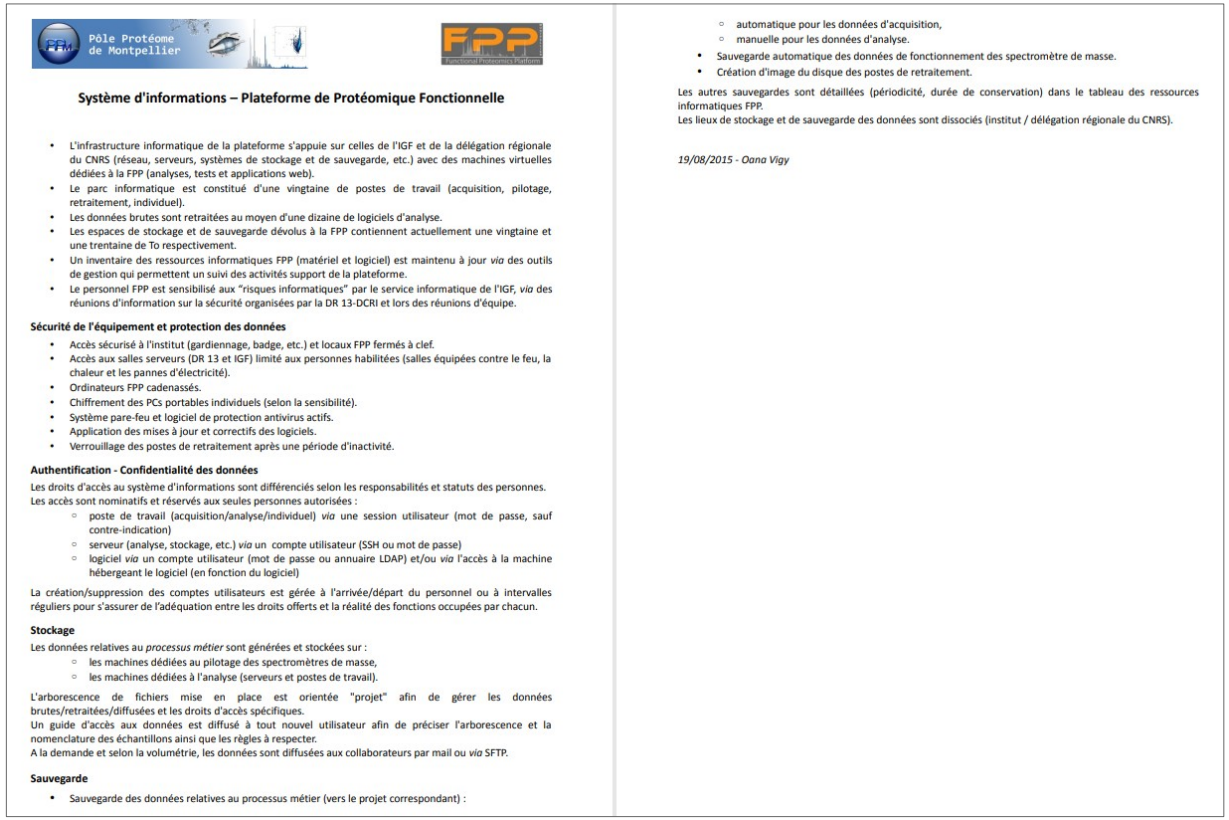
- Données brutes et retraitées
- Bases de données
- Applications métiers
- Outils de gestion/support
- Infrastructure réseau
- Serveurs d'applications (analyse, test)
- Systèmes de stockage/sauvegarde
- Dispositifs de sécurité

Système d'information

Système d'information : ensemble organisé de ressources : matériel, logiciel, personnel, données, procédures permettant d'acquérir, traiter, stocker, communiquer des informations (sous forme de données, textes, images, sons, etc…) dans des organisations. *(Reix)*

- Données brutes et retraitées
- Bases de données
- Applications métiers
- Outils de gestion/support
- Infrastructure réseau
- Serveurs d'applications (analyse, test)
- Systèmes de stockage/sauvegarde
- Dispositifs de sécurité

~ PGD de structure



Pôle Protéome de Montpellier

Système d'informations – Plateforme de Protéomique Fonctionnelle

- L'infrastructure informatique de la plateforme s'appuie sur celles de l'IGF et de la délégation régionale du CNRS (réseau, serveurs, systèmes de stockage et de sauvegarde, etc.) avec des machines virtuelles dédiées à la FPP (analyses, tests et applications web).
- Le parc informatique est constitué d'une vingtaine de postes de travail (acquisition, pilotage, retraitement, individuel).
- Les données brutes sont retraitées au moyen d'une dizaine de logiciels d'analyse.
- Les espaces de stockage et de sauvegarde dévolus à la FPP contiennent actuellement une vingtaine et une trentaine de To respectivement.
- Un inventaire des ressources informatiques FPP (matériel et logiciel) est maintenu à jour *via* des outils de gestion qui permettent un suivi des activités support de la plateforme.
- Le personnel FPP est sensibilisé aux "risques informatiques" par le service informatique de l'IGF, *via* des réunions d'information sur la sécurité organisées par la DR 13-DCRI et lors des réunions d'équipe.

Sécurité de l'équipement et protection des données

- Accès sécurisé à l'institut (gardiennage, badge, etc.) et locaux FPP fermés à clef.
- Accès aux salles serveurs (DR 13 et IGF) limité aux personnes habilitées (salles équipées contre le feu, la chaleur et les pannes d'électricité).
- Ordinateurs FPP cadenassés.
- Chiffrement des PCs portables individuels (selon la sensibilité).
- Système pare-feu et logiciel de protection antivirus actifs.
- Application des mises à jour et correctifs des logiciels.
- Verrouillage des postes de retraitement après une période d'inactivité.

Authentification - Confidentialité des données

Les droits d'accès au système d'informations sont différenciés selon les responsabilités et statuts des personnes.
Les accès sont nominatifs et réservés aux seules personnes autorisées :

- poste de travail (acquisition/analyse/individuel) *via* une session utilisateur (mot de passe, sauf contre-indication)
- serveur (analyse, stockage, etc.) *via* un compte utilisateur (SSH ou mot de passe)
- logiciel *via* un compte utilisateur (mot de passe ou annuaire LDAP) et/ou *via* l'accès à la machine hébergeant le logiciel (en fonction du logiciel)

La création/suppression des comptes utilisateurs est gérée à l'arrivée/départ du personnel ou à intervalles réguliers pour s'assurer de l'adéquation entre les droits offerts et la réalité des fonctions occupées par chacun.

Stockage

Les données relatives au *processus métier* sont générées et stockées sur :

- les machines dédiées au pilotage des spectromètres de masse,
- les machines dédiées à l'analyse (serveurs et postes de travail).

L'arborescence de fichiers mise en place est orientée "projet" afin de gérer les données brutes/retraitées/diffusées et les droits d'accès spécifiques.
Un guide d'accès aux données est diffusé à tout nouvel utilisateur afin de préciser l'arborescence et la nomenclature des échantillons ainsi que les règles à respecter.
A la demande et selon la volumétrie, les données sont diffusées aux collaborateurs par mail ou *via* SFTP.

Sauvegarde

- Sauvegarde des données relatives au processus métier (vers le projet correspondant) :
  - automatique pour les données d'acquisition,
  - manuelle pour les données d'analyse.
- Sauvegarde automatique des données de fonctionnement des spectromètre de masse.
- Création d'image du disque des postes de retraitement.

Les autres sauvegardes sont détaillées (périodicité, durée de conservation) dans le tableau des ressources informatiques FPP.
Les lieux de stockage et de sauvegarde des données sont dissociés (institut / délégation régionale du CNRS).

19/08/2015 - Oana Vigy

Avec le support du service Informatique de l'IGF !

Evolution du système d'information FPP jusqu'en 2023

Samples

Acquisition

Proteomics data analysis sofwares *(locally installed)*

PROTTER

Skyline

DynamX

PLGS

MassLynx

Be tested / under watch

No more used

MATRIX SCIENCE

myProMS

Progenesis QI

Tested

ProFI PROTEOMICS Proline 1.1

Datasources

UniProt-GOA

UniProt

STRING

BioGRID

Tools

proteowizard

PTXQC

Data sharing

PRIDE

PRoteomics IDEntifications Database

24 dépôts depuis 2017

Evolution du système d'information FPP jusqu'en 2023

FPP Functional Proteomics Platform

Samples

Acquisition

Proteomics data analysis sofwares *(locally installed)*

MQ

LEADING

Skyline

DynamX

PLGS

PROTTER

Pd

Master Leading

MassLynx

No more used

MATRIX SCIENCE · myProMS · Progenesis QI

Tested

ProFI PROTEOMICS · Proline 1.1

Be tested / under watch

Prostar · script... · R

Datasources

UniProt-GOA

UniProt

IntAct

BioGRID

STRING

Tools

proteowizard

PTXQC

Data sharing

PRIDE

PRoteomics IDEntifications Database

24 dépôts depuis 2017

IT infrastructure

Save

Synchro

Acquisition (4)

55 To

x2 sur les 5 dernières années

Virtual Machines

Supporting tools *(locally installed)*

REDMINE

Joomla!

GLPI

booked

2023

JASPERSOFT

Zotero

Migration done or no more used

FileMaker

LimeSurvey

CITADEL TEAM

FileSender Accès

Coming soon ?!

Labguru

Mise en place d'un cahier de laboratoire électronique

Réalisé → définir et intégrer

- l'extrait biologique dans CLE *(via collection)*
- les protocoles *(standards et manuel technique)*

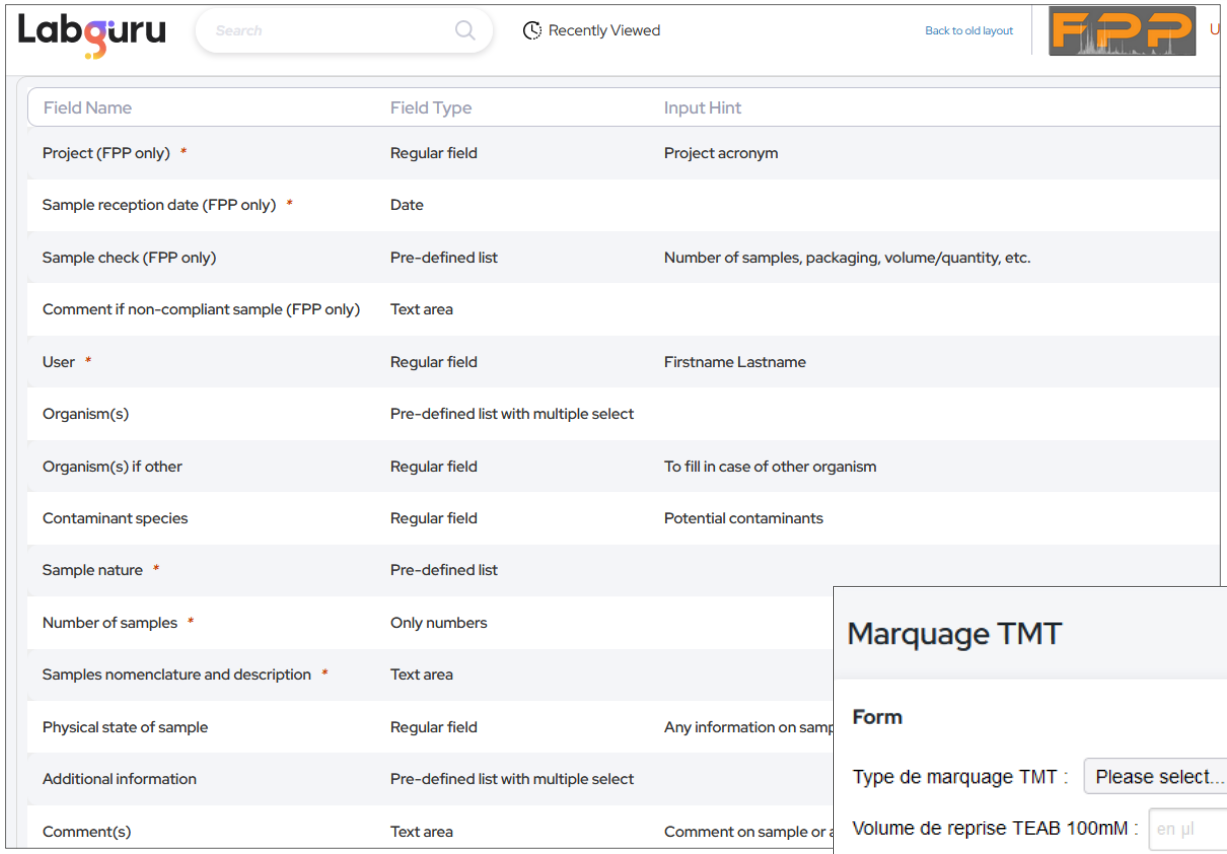
En cours → définir et « formater » 1 expérience

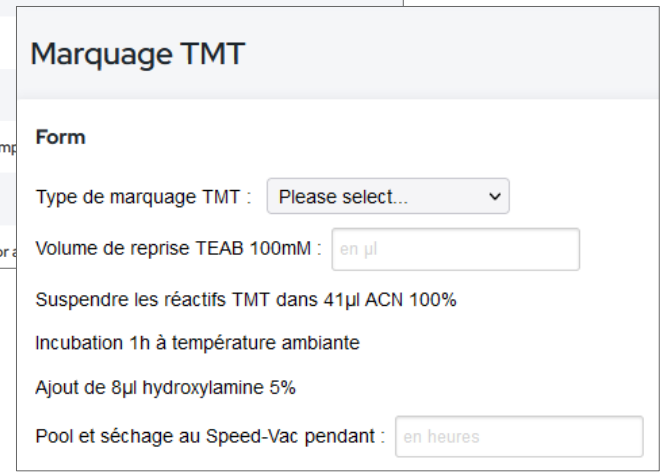
Ensuite : migration du suivi des projets...

Labguru

***cf.* nos besoins exprimés en 2010**

- « Meilleure traçabilité sur les étapes expérimentales
- Pouvoir retrouver rapidement toutes les informations expérimentales
- Générer un rapport d'analyse complet et détaillé
- Partager l'information entre les acteurs d'un projet"





Management intégré basé sur le SI

Logiciels d'analyse

Logiciels support

booked

Zotero

REDMINE

Labguru

Acronyme du projet

Sans redondance

Espace de stockage

- Acquisitions
- Analyses
- Résultats

Supports papier

- Cahier de laboratoire
- Planning machine

PGD de structure

Focus sur la gestion des données!

Plan de gestion des données

Groupe de travail – Plateformes BioCampus

- PF de production/analyse de données
- PF de production/transformation de produits biologiques (métadonnées)

Soutien

Groupe de travail – PPM *(multi-sites, multi-SI, multi-activités)*

- Difficultés à définir des produits de recherche pour une intégration réfléchie dans un PGD de projet (*ma-dmp*)
- Choix de l'outil ? Du modèle ? (modèles AgroParisTech/INRAE), évolution du modèle → évolution du PGD ?
- 2 PGDs de structure en protéomique publics (DMP Opidor) : PPM-PPC, PAPPSO (PIXANIM)

Groupe de travail – MuDis4LS

GT PGD MuDis4LS

DSW
DATA STEWARDSHIP WIZARD

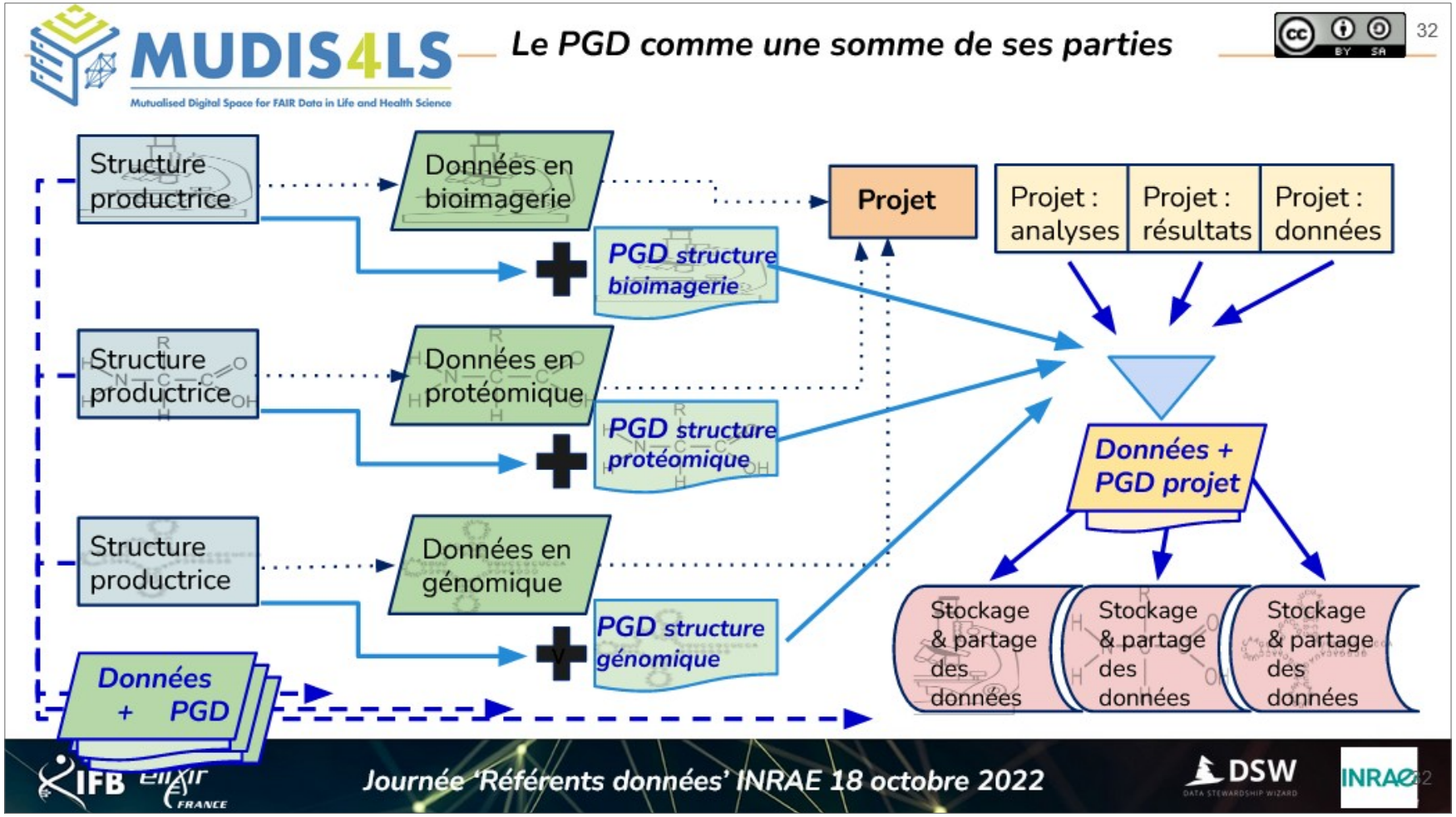
Modèle de PGD de structure

- Multi-omique
- En cours de dévelopement
- Retours utilisateur

Interconnexion DSW-DMP Opidor à venir

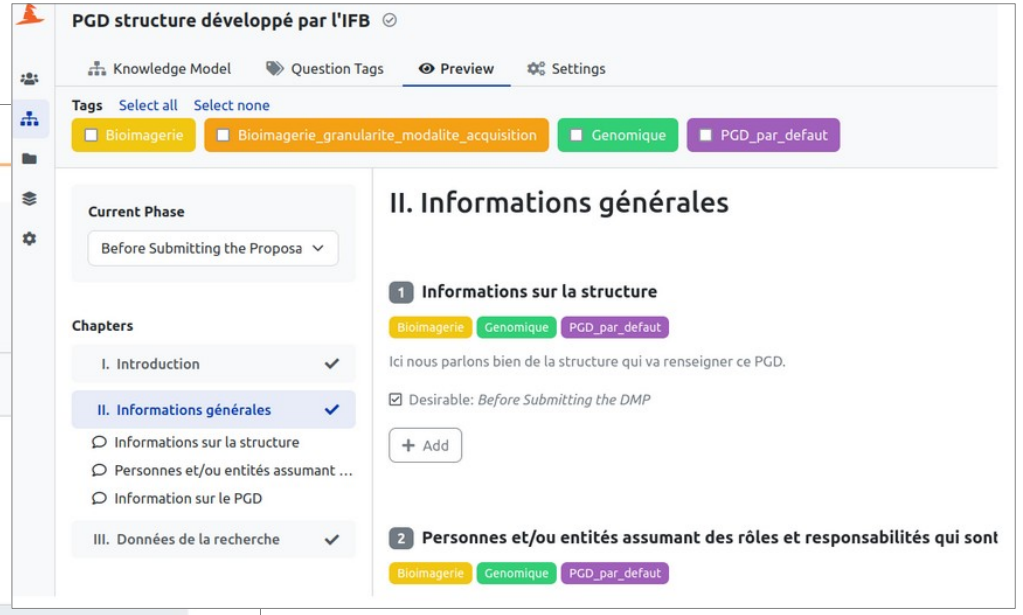
1 autre GT prévu sur les cas d'usage

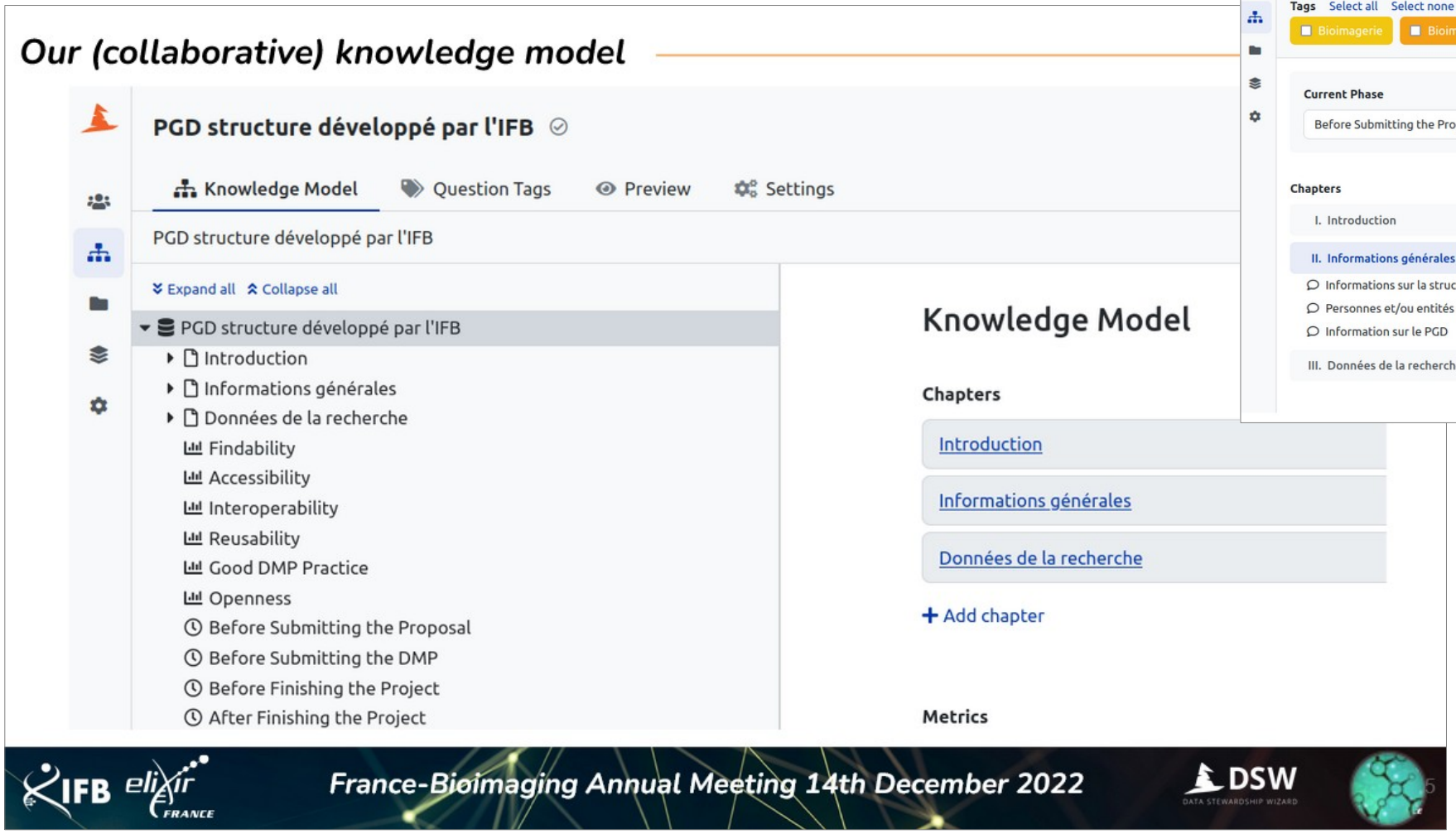
GT PGD MuDis4LS



Restructuration importante fin 2022 avec v2

Granularité possible





Frédéric Brau &
Paulette Lieby

Faisal Bekkouche
Audrey Bihouee
Christophe Bruley
Sophie Desset
Frédéric de Lamotte
Jean-François Dufayard
Nadia Goué
Hamid Meziane
Perrine Paul-Gilloteaux
Jean Salamero
Fanny Sebire
Marek Suchánek
Raluca Teusan
Oana Vigy