



INRAE



Le management de la donnée au Centre International de Ressources Microbiennes

Jonathan MINEAU-CESARI
CIRM

Journées du PEPI IBIS @ Cité de l'Espace, Toulouse
16, 17 et 18 Novembre 2021



➤ Data-management au CIRM

- **Le CIRM**

- Un réseau national de CRB microbiens
- Cœur de métier
- Un producteur de données hétérogènes
- Des partenaires nationaux et européens

- **Data-management**

- Organisation, réseaux et communautés
- Système d'Information
- PGD
- RGPD
- FAIR
- Communication WEB

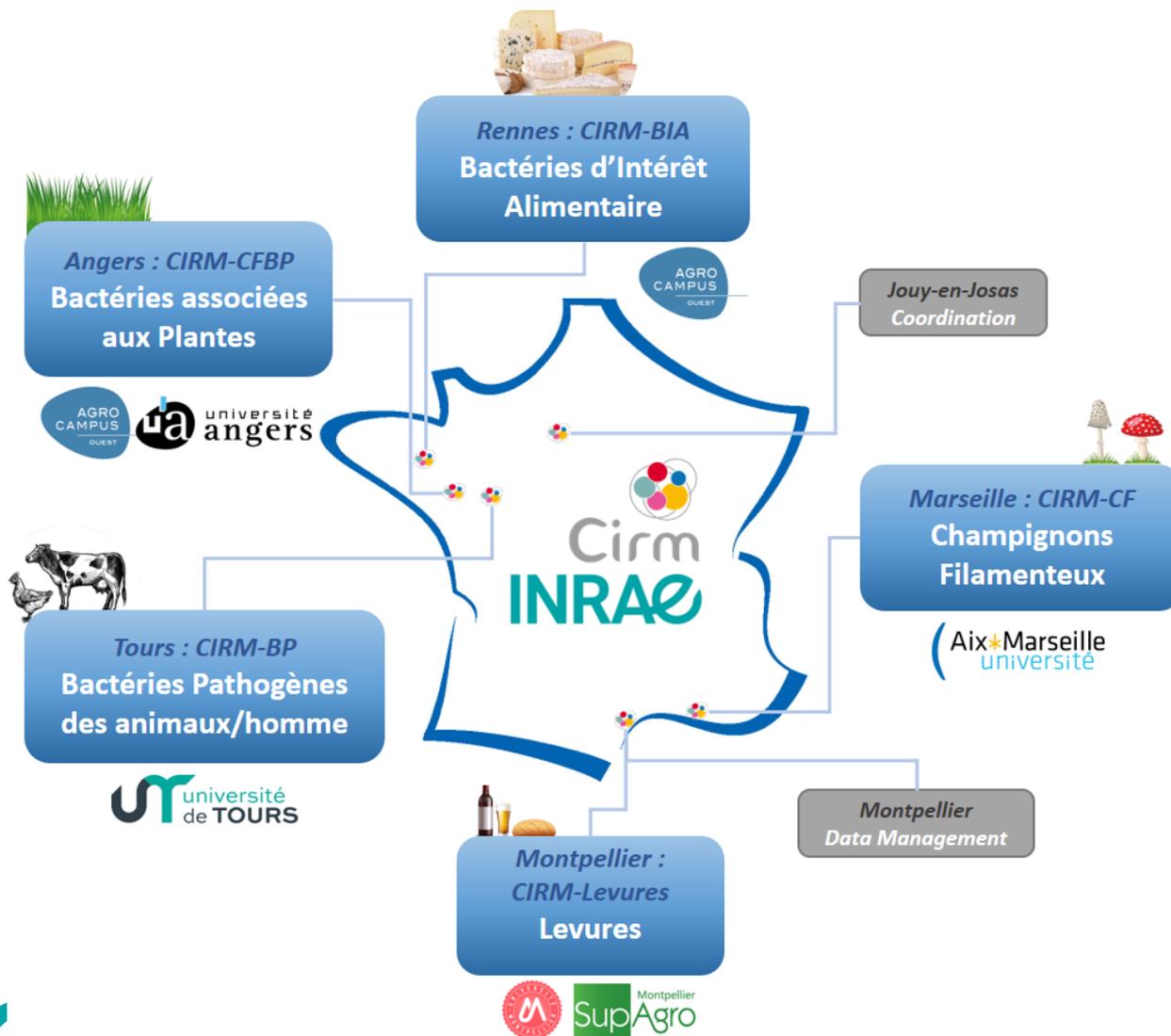
- **Bilan**



> Data-management au CIRM

- **Le CIRM** : Un réseau national de CRB microbiens

➤ Le CIRM : un réseau de CRB microbiens



Le CIRM en Bref et en Chiffres

- **Statu** : Groupement d'intérêt scientifique (GIS)
- **Nombre de CRB** : 5 (répartis sur le territoire Français)
- **Date de création** : 2004
- **Département INRAE** : Microbiologie et Chaîne alimentaire (MICA)
- **Objectif** : préserver et valoriser des collections de micro-organismes
- **Collectif** : 20 chercheurs et ITA
- **Nombres de souches inscrites** : 15000 (dont 3000 diffusées/an)
- **Organisation** :
 - Un coordinateur national (IDF)
 - Un data-manager national (Occitanie-Montpellier)
 - CRB accueillis dans des UMR (INRAE, écoles AGRO et universités)
 - Un/une responsable par CRB, tous membres du CoPil et CoDir
- **Qualité** : certification ISO 9001
- **Documentation** :
 - [Site web](#)
 - [Flyer de présentation](#)
 - [Charte](#)

> Data-management au CIRM

- **Le CIRM : Cœur de métier**

> Le CIRM : cœur de métier



- **Collecter** : Le CIRM se doit de continuellement **enrichir ses collections pour maintenir une offre** en adéquation avec l'avancée des découvertes en microbiologie. Cela est réalisé à travers des **projets scientifiques**, des **collectes** sur le terrain ou le rapatriement de collections INRAE ou extérieures de **fort intérêt patrimonial ou scientifique**.
- **Identifier** : Les CRB ont des activités de **microbiologie traditionnelle** pour cultiver et conserver les microorganismes qui leur sont confiés. La caractérisation des entrées dans les collections consiste à minima à **assurer la position taxonomique de la souche**, le plus souvent par des techniques moléculaires (ARN16S, ITS, ...).
- **Conserver** : Selon les espèces les souches sont préservées par les techniques de **cryoconservation** (-80°, Azote liquide) et de **lyophilisation**. Une règle d'or est que chaque souche est **conservée physiquement en double**.
- **Caractériser** : A travers les projets de recherche, le CIRM entreprend la **caractérisation génomique et phénotypique** de ses collections.
- **Distribuer** : La **mise à disposition des ressources** microbiologiques présentes dans les collections ouvertes du CIRM est un principe de fonctionnement. Les collections sont **visibles et accessible sur les portails web** de chaque CRB et permettent aux **utilisateurs scientifiques, industriels ou enseignants de commander les souches d'intérêt**.

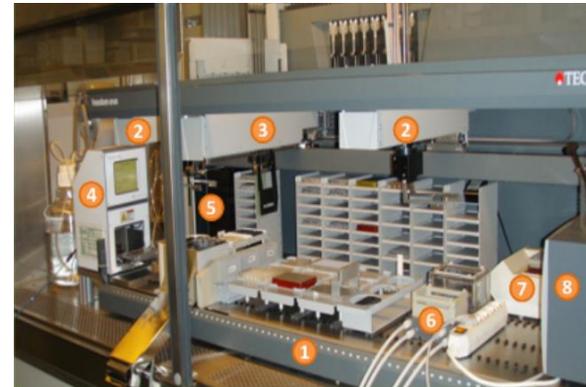
> Data-management au CIRM

- **Le CIRM** : Un producteur de données hétérogènes

➤ Le CIRM : un producteur de données hétérogènes

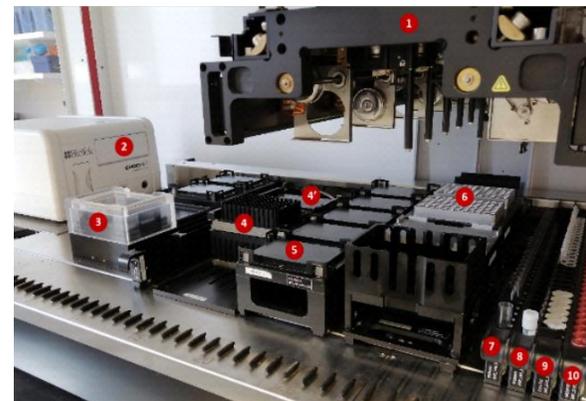
Le CIRM est engagé dans des projets de recherche pour évaluer la **biodiversité** de ses collections et leurs potentiels enzymatiques, métabolique ou biologiques. Ces objectifs requièrent un effort de **séquençage systématique** et de **caractérisation phénotypique** faisant appel à des dispositifs de **criblage automatisés**.

- **Taxonomie** : Identification / phylogénomique / recherche de gènes et potentiel enzymatique / profilage métabolique / dérégulation / Ensemble de souches de référence.
- **Métabolisme – activité enzymatique** : métabolisme des oligosaccharides production de vitamines , d'EPS, de composés aromatiques / activités cellulolytiques / dégradation de colorants.
- **Activité biologique** : Activité anti-fongique /production de bactériocines / profils de résistance aux antibiotiques selon les procédures EUCAST/CASFL / propriété anti-inflammatoires.



Plateforme robotisée - CIRM-CF

- 1: Automate de pipetage
- 2: Hotte à flux laminaire
- 3: Pipetage
- 4: Scellage de microplaques
- 5: Incubateur
- 6: Unité de filtration
- 7: Bloc chauffant
- 8: Spectrofluorimètre



Plateforme de criblage - CIRM-BIA

- 1: 8 canaux
- 2: Spectrophotomètre
- 3: Pompe à vide
- 4: blocs thermiques
- 5,6,7,8,9,10 : plaques, microplaques, ..

> Data-management au CIRM

- **Le CIRM : Des partenaires nationaux et européens**

> Le CIRM : des partenaires nationaux et européens



AgroBRC-RARe est une **Infrastructure de recherche nationale** inscrite sur la feuille de route du Ministère de la Recherche. Elle est pilotée par INRAE et implique trois autres instituts que sont le CIRAD, le CNRS et l'IRD. C'est un regroupement de réseaux de CRB constitué de **cinq piliers** : « Plantes », « Animaux », « Environnement », « Forêt » et « Micro-organismes ». Elle regroupe une quarantaine de CRB français multi-domaines et multi-instituts qui stockent des millions de ressources génétiques.

Le CIRM est **labellisé Infrastructure de Recherche Collective (ISC)** INRAE. Il coordonne le pilier microbien de l'infrastructure. Le **pilier microbien** d'AgroBRC-RARe est constitué du CIRM et de la plateforme SAMBO de MétaGénopolis.



MIRRI (Microbial Resources Research Infrastructure) est une infrastructure de recherche paneuropéenne dédiée aux ressources microbiennes. MIRRI propose un **catalogue** de 500,000 souches microbiennes (archae, bacteria, yeasts, fungi, algae, virus, ..) Il fédère 14 CRB microbiens de 10 pays Européens (ainsi que de la Russie.) L'objectif est d'exposer les données des infrastructures de recherche participantes et de rendre ces données interopérables pour les valoriser et construire l'Europe de la recherche. Avec le **CRBIP** de l'institut Pasteur, le CIRM est composante du nœud français MIRRI.

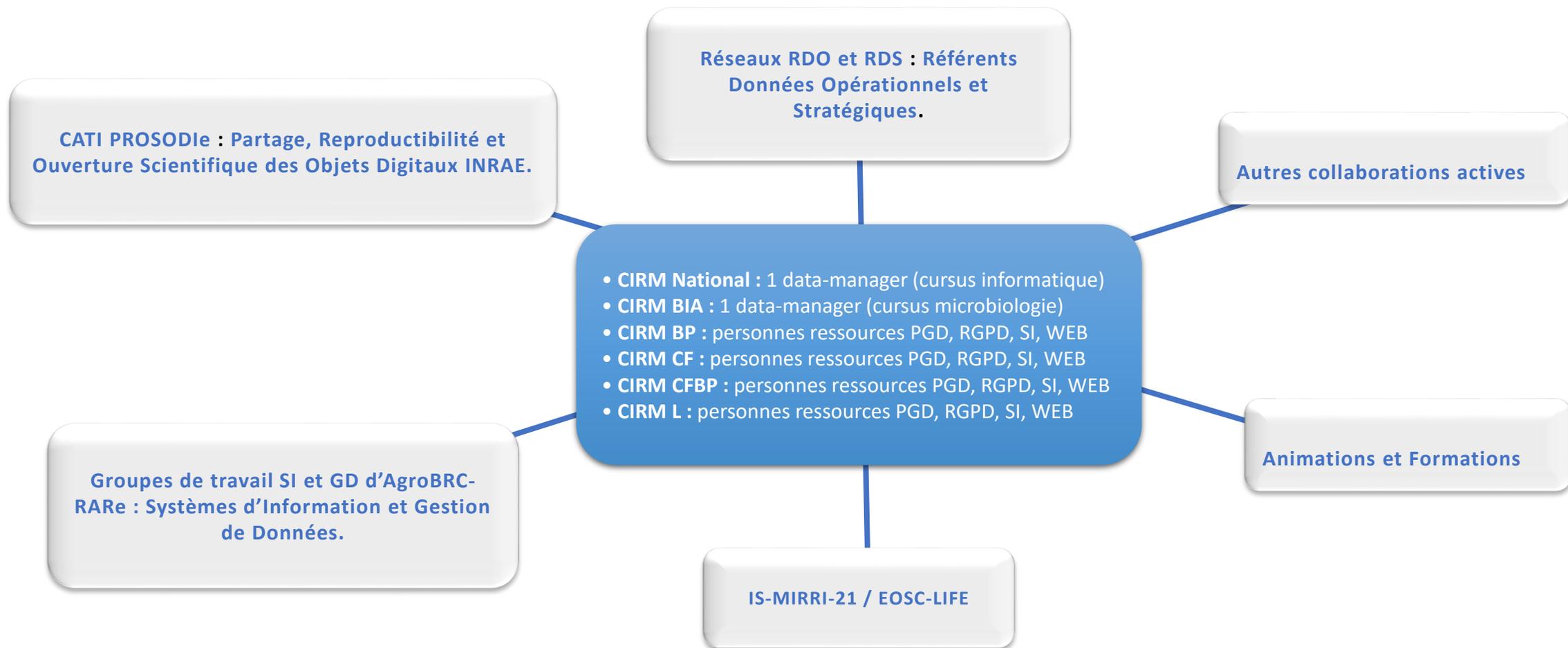
Autres Projets et collaborations

- Impliqué dans des projets (IS-MIRRI21 / EOSC-Life / On-Health EJP/ INFRA-SERV ...) pour le développement de l'offre de services techniques et scientifiques en microbiologie : Conservation des communautés microbiennes / Génomique / MALDITOF / AMR / ..
- Offre d'accès de services transnationaux (TNA) ex.: Test d'activités anti-infectieuses

> Data-management au CIRM

- **DM** : Organisation, réseaux et communautés

> DM: Organisation, réseaux et communautés



> DM: Organisation, réseaux et communautés

- CIRM National : 1 data-manager (cursus informatique)
- CIRM BIA : 1 data-manager (cursus microbiologie)
- CIRM BP : personnes ressources PGD, RGPD, SI, WEB
- CIRM CF : personnes ressources PGD, RGPD, SI, WEB
- CIRM CFBP : personnes ressources PGD, RGPD, SI, WEB
- CIRM L : personnes ressources PGD, RGPD, SI, WEB

Réseaux RDO et RDS : Référents Données Opérationnels et Stratégiques.

- Dans le cadre du **plan de gouvernances des données INRAE**, les **RDO** ont été identifiés pour accompagner au plus près les chercheurs dans la mise en œuvre concrète des bonnes pratiques de gestion des données. Les **RDS** ont été nommés par les chefs de département pour fournir une vision plus globale des données d'intérêt pour les thématiques du département.
- Les RDO du CIRM : les data-managers CIRM et CIRM-BIA et le directeur-adjoint du CIRM-CF

CATI PROSODIe : Partage, Reproductibilité et Ouverture Scientifique des Objets Digitaux INRAE.

- **CATI transversal** dont la vocation est de promouvoir, accompagner et valoriser la mise en œuvre de la gestion et du partage des données et des codes dans des objectifs de réutilisation et de reproductibilité de la science.
- Projets : PGD, RGPD, FAIR, DOI, Ontologies
- Implications : les data-managers CIRM et CIRM-BIA

➤ DM: Organisation, réseaux et communautés

- CIRM National : 1 data-manager (cursus informatique)
- CIRM BIA : 1 data-manager (cursus microbiologie)
- CIRM BP : personnes ressources PGD, RGPD, SI, WEB
- CIRM CF : personnes ressources PGD, RGPD, SI, WEB
- CIRM CFBP : personnes ressources PGD, RGPD, SI, WEB
- CIRM L : personnes ressources PGD, RGPD, SI, WEB

Groupes de travail SI et GD d'AgroBRC-RARe : Systèmes d'Information et Gestion de Données.

- **SI** : le data-manager du CIRM représente le pilier « microbien » pour les SI de RARe.
- **GD** : piloté et animé par 3 data-managers (RARe, CIRM et CIRM-BIA) pour accompagner les 40 CRB dans la mise en place du RGPD et des PGD.
- **SI et APA** : collaboration des deux groupes (data-managers CIRM et CIRM-BIA pour les SI et directrice CIRM-CFBP pour l'APA)

IS-MIRRI-21 / EOSC-LIFE

- Le CIRM est impliqué dans tous les work-packages (WP) d'**EOSC-LIFE** et d'**IS-MIRRI-21** dont il pilote le WP2 (« Research Strategy »). Ces projets ont pour but de développer et consolider MIRRI en tant qu'ERIC. Il a été demandé à IS-MIRRI-21 de répondre aux exigences d'EOSC sur la compatibilité de MIRRI avec les principes FAIR. Le CIRM pilote cette mission.
- Implications : le data-manager CIRM et un IE informaticien (EOSC-LIFE) basé au CIRM-Levures.

➤ DM: Organisation, réseaux et communautés

- CIRM National : 1 data-manager (cursus informatique)
- CIRM BIA : 1 data-manager (cursus microbiologie)
- CIRM BP : personnes ressources PGD, RGPD, SI, WEB
- CIRM CF : personnes ressources PGD, RGPD, SI, WEB
- CIRM CFBP : personnes ressources PGD, RGPD, SI, WEB
- CIRM L : personnes ressources PGD, RGPD, SI, WEB

Autres collaborations actives

- EIFI (E-infrastructure Fédérative INRAE)
- DIPSO (Direction Pour la Science Ouverte)
- BioAWARE (pour le SI BioLMICS)
- RDA (Research Data Alliance)

Animations et Formations

- **Animations et sensibilisation** et TP/TD auprès des CRB du CIRM autour du management des données.
- **Formation diplômante** des data-managers CIRM et CIRM-BIA, «DU Data Scientific Manager» (ISDM)

> Data-management au CIRM

- **DM** : Système d'information

➤ DM : Système d'Information BioloMICS

BioloMICS

- système de gestion de données biologiques utilisé par de nombreux acteurs dans les domaines de la biotechnologie, de l'industrie pharmaceutique, de l'alimentation, des services de santé publique, des institutions académiques, universités, collections et musées.
- développé par la société [BioAware](#) dont le siège est basé en Belgique. Les interlocuteurs sont francophones et anglophones.
- « software as a service » (SaaS) ou logiciel en tant que service
- licence par CRB annuelle et pour une seule instance mais la connexion peut être multi-utilisateurs.
- chaque CRB a une personne ressource pour la construction de la base, la maintenance et l'évolution des outils. Forte réactivité du SAV.
- le système a été choisi comme SI de l'IR européenne MIRRI
- suite d'outils dont les deux principaux sont BioloMICS.net et BioloMICS.web



Date d'obtention du système par les CIRM

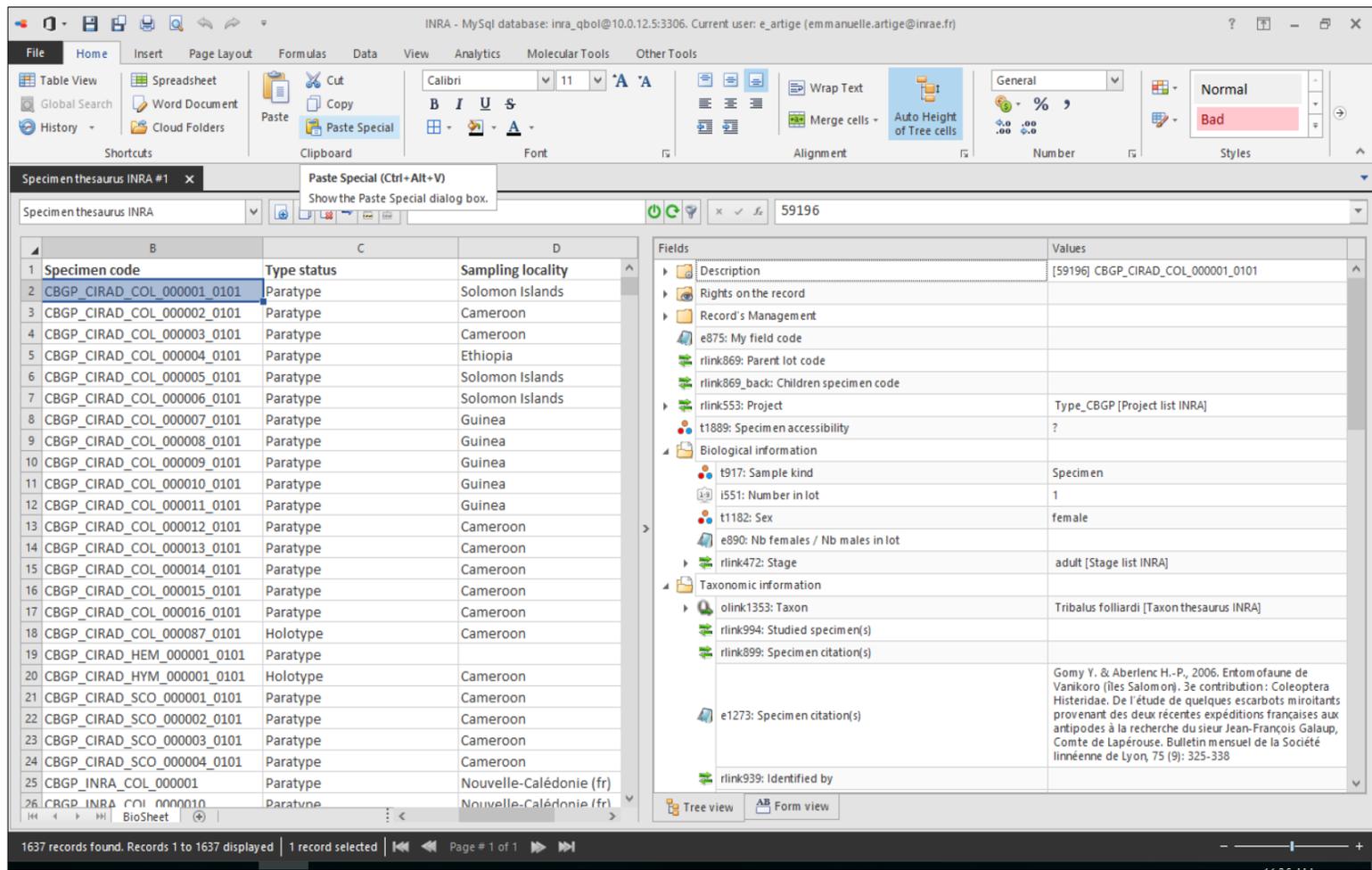
- CIRM-Levures : **2017**
- CRIM-BIA : **2018**
- CIRM-CFBP : **2019**
- CIRM-CF : **2021**
- CIRM-BP : **2022** → **uniformisation totale des SI du CIRM**

Prochainement : parution d'un article dans les « Cahiers des Techniques » INRAE dédié à l'utilisation de BioloMICS dans les CRB de RARe.

➤ DM : Système d'Information BioloMICS

BioloMICS.NET

- Interface utilisateur «user-friendly» inspirée de l'environnement Microsoft
- Pas de compétences spécifiques à l'informatique pour s'approprier le logiciel et construire une base de données.
- Ergonomique, flexible et modulable qui peut être **utilisé à différentes finalités** (stockage, gestion, analyse) pour les ressources biologiques, les expériences de recherche ou les données biologiques de tous types différents
- Outils mis à disposition **pour gérer et analyser les données** de sa base, par exemple (LIMS, analyses spectrales, distribution géographiques, génération de graphs, ..)



The screenshot displays the BioloMICS.NET interface within a web browser window. The main window title is "INRA - MySql database: inra_qbol@10.0.12.5:3306. Current user: e_artige (emmanuelle.artige@inrae.fr)". The interface includes a menu bar (File, Home, Insert, Page Layout, Formulas, Data, View, Analytics, Molecular Tools, Other Tools) and a ribbon with various tools like Table View, Spreadsheet, Word Document, and Cloud Folders. A data table is visible with columns for Specimen code, Type status, and Sampling locality. A detailed record view for specimen 59196 is shown on the right, including fields for Description, Rights, Record's Management, Biological information, and Taxonomic information.

Specimen code	Type status	Sampling locality
1	Paratype	Solomon Islands
2	Paratype	Cameroon
3	Paratype	Cameroon
4	Paratype	Ethiopia
5	Paratype	Solomon Islands
6	Paratype	Solomon Islands
7	Paratype	Guinea
8	Paratype	Guinea
9	Paratype	Guinea
10	Paratype	Guinea
11	Paratype	Guinea
12	Paratype	Guinea
13	Paratype	Cameroon
14	Paratype	Cameroon
15	Paratype	Cameroon
16	Paratype	Cameroon
17	Paratype	Cameroon
18	Holotype	Cameroon
19	Paratype	Cameroon
20	Holotype	Cameroon
21	Paratype	Cameroon
22	Paratype	Cameroon
23	Paratype	Cameroon
24	Paratype	Cameroon
25	Paratype	Nouvelle-Calédonie (fr)
26	Paratype	Nouvelle-Calédonie (fr)

> DM : Système d'Information BioloMICS

BioloMICS.WEB

- Déploiement d'une interface web à partir de la base de données.
- **Sélection** et mise à disposition en ligne de ressources souhaitées au client d'une manière sécurisée et à jour.
- Diffusion des données contenues dans BioloMICS.net avec **une interface en WYSIWYG**
- **Plusieurs sites web** peuvent être générés à partir d'une même base de données.



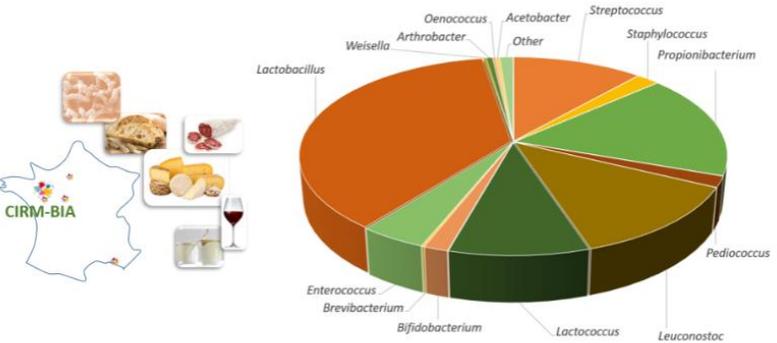
Home Catalog Order Price Service Deposit Quality Cite CIRM-BIA GDPR-RGPD Contact us Unknown user

The International Centre for Microbial Resources

Managed by the French National Research Institute for Agriculture, Food and the Environment (INRAE) the CIRM is a Biological Resource Centre dedicated to preserving, enriching and promoting the microbial diversity of its collections.
For more details, visit the CIRM website.

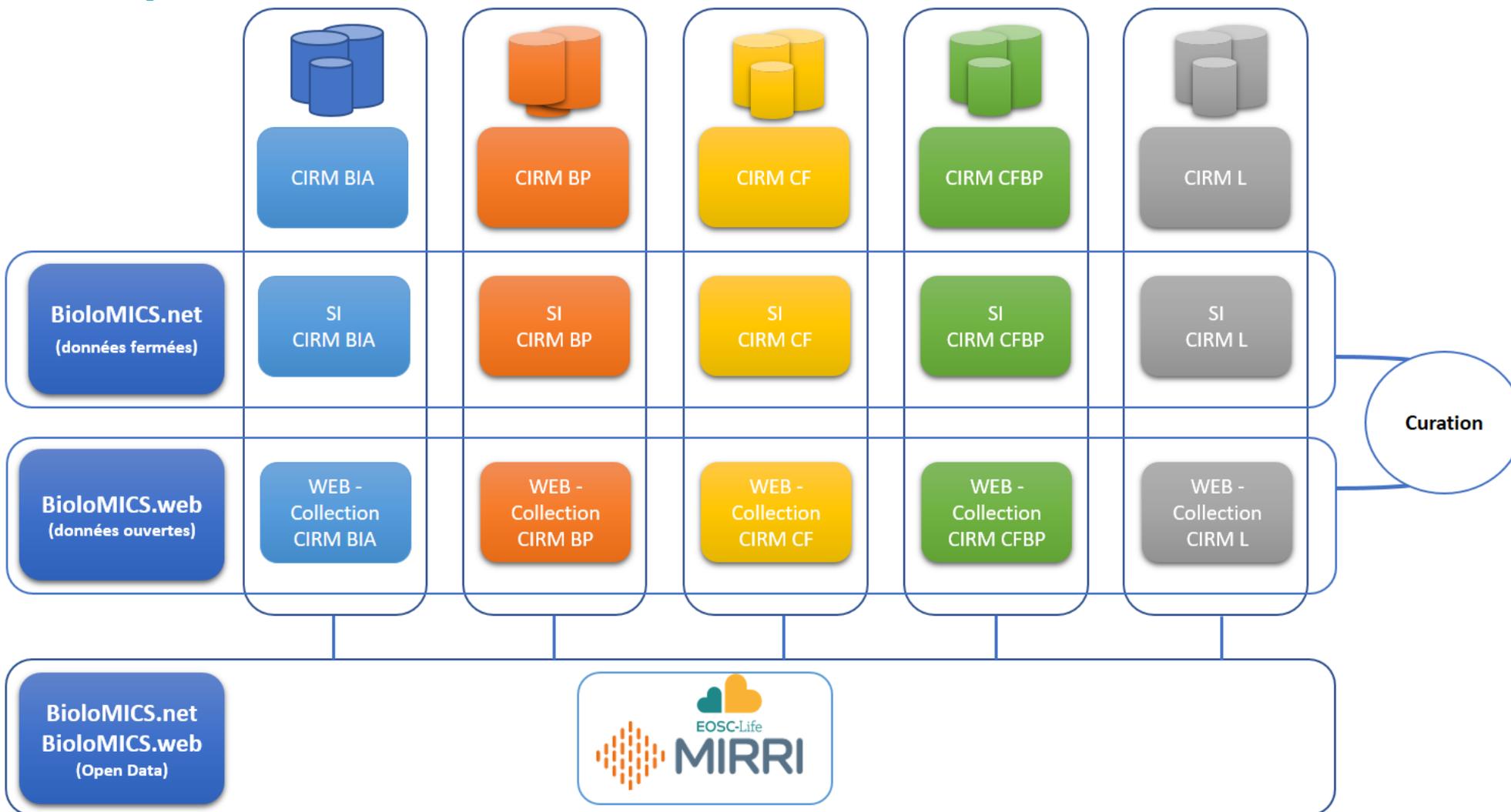
CIRM-BIA dedicated to food-associated bacteria

The Centre host more than 4,000 strains of bacteria related to fermented foods and bio-preservation which are as diverse as possible (160 species, 30 genera). These numbers are growing constantly through targeted collection from various biotopes.
The strains are selected in order to offer the greatest possible diversity, from both the taxonomic and functional points of view, so as to ensure their usefulness to the scientific and industrial communities.



The pie chart illustrates the distribution of bacterial genera in the CIRM-BIA collection. The largest categories are Lactobacillus, Enterococcus, and Bifidobacterium. Other significant genera include Lactococcus, Leuconostoc, and Propionibacterium. Smaller categories include Weisella, Arthrobacter, Oenococcus, Acetobacter, Streptococcus, and Staphylococcus. An 'Other' category is also present.

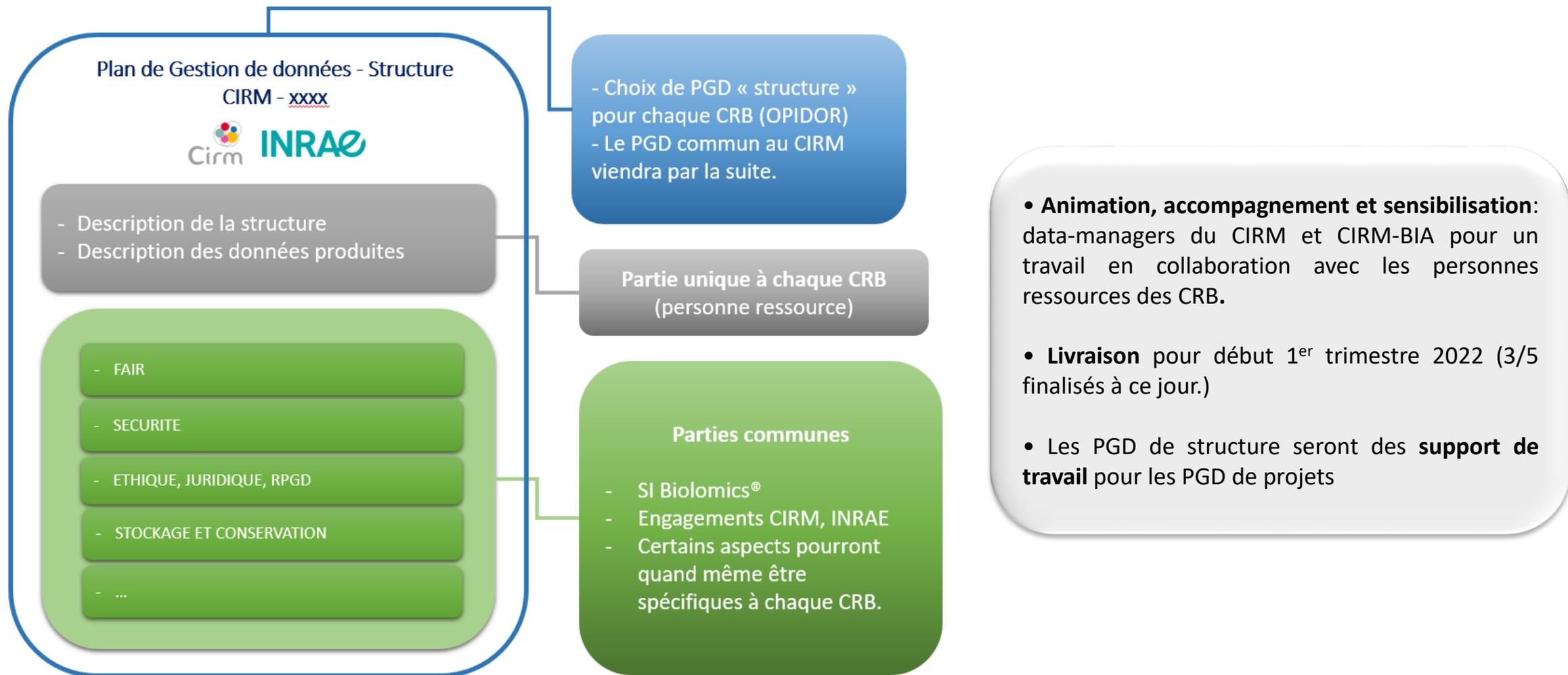
➤ DM : Systèmes d'Information du CIRM



> Data-management au CIRM

- **DM** : Plan de Gestion des données (PGD)

➤ DM : Plan de Gestion des Données

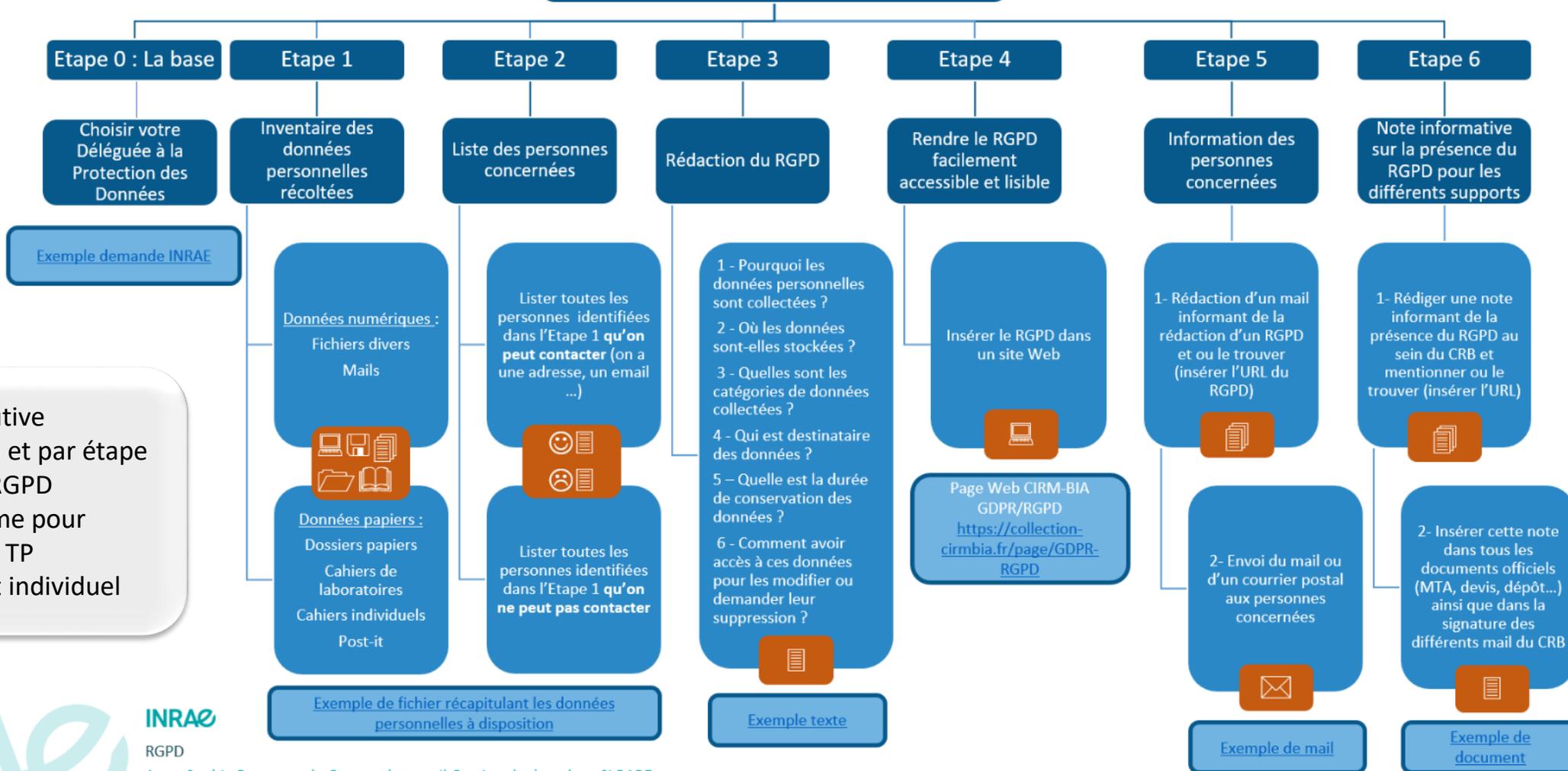


> Data-management au CIRM

- **DM** : Règlement Générale sur la protection des données (RGPD)

> DM : RGPD

Mise en place d'un RGPD pas à pas



- Approche évolutive
- Faire « simple » et par étape
- Démystifier le RGPD
- Travail en binôme pour l'animation et les TP
- Suivi collectif et individuel



INRAE

RGPD

Anne-Sophie Bage pour le Groupe de travail Gestion de données_ SI RARE

Source : Anne-Sophie BAGE, CIRM-BIA

> Data-management au CIRM

- **DM:** Un outils pour les CRB

➤ DM : Outils de suivis RGPD & PGD

Nom du référent données opérationnel	MINEAU Jonathan		David Navarro		Anne-Sophie Bage			
Mail du référent données opérationnel	jonathan.mineau@inrae.fr		david.navarro@inrae.fr		anne-sophie.bage@inrae.fr			
Institut tutelle du CRB	INRAE	INRAE	INRAE	INRAE	INRAE	INRAE		
RGPD (Règlement général sur la protection des données)	CIRM-L	CIRM-CFBP	CIRM-CF	CIRM-BP	CIRM-BIA	SAMBO	UBOC	MISP
Référent RGPD	Grondin Cécile	Perrine Portier	David Navarro	Emmanuelle Helloin	Anne-Sophie Bage			
Mail référent	cecile.grondin@inrae.fr	perrine.portier@inrae.fr	david.navarro@inrae.fr	emmanuelle.helloin@inrae.fr	anne-sophie.bage@inrae.fr			
DPO choisi et déclaré par la structure d'accueil					INRAE			
1. Inventaire des données personnelles récoltées. (informatiques, mail, cahier de laboratoire, postit...)	Oui	Non	Oui	Non	Oui	Non		
2. Liste des personnes à contacter pour les informer de l'existence du RGPD au sein du CRB.	Oui	Non	Oui	Non	Oui	Non		
3. Liste des personnes que l'on ne peut pas contacter (pas de mail, d'adresse postale, pas de numéro de téléphone) à conserver précieusement.	Oui	Non	Oui	Non	Oui	Non		
4. Rédaction du RGPD								
a. Pourquoi des données personnelles sont collectées ?	Oui	Difficulté	Oui	Non	Oui	Non		
b. Ou sont stockées Les données ?	Oui	Difficulté	Oui	Non	Oui	Non		
c. Quelles sont les catégories de données collectées ?	Oui	Difficulté	Oui	Non	Oui	Non		
d. Qui est destinataires des données ?	Oui	Difficulté	Oui	Non	Oui	Non		
e. Durée de conservation des données ?	Oui	Difficulté	Oui	Non	Oui	Non		
f. Comment peut-on avoir accès à nos données personnelles pour les modifier ou demander leur suppressions.	Oui	Difficulté	Oui	Non	Oui	Non		
5. Rédaction d'un mail d'information	Oui	Non	Oui	Non	Oui	Non		
6. Envoi du mail d'information	Non	Non	Oui	Non	Oui	Non		
7. Rédaction d'un texte synthétique à insérer dans les divers documents transmis aux différents partenaires et clients, ou autres..	Non	Non	Oui	Non	Oui	Non		
8. Insérer le texte dans les divers docs	Non	Non	Oui	Non	Oui	Non		
9. Insérer le RGPD dans une page Web du CRB pour que l'accès au registre soit facilité.	Non	Non	Oui	Non	Oui	Non		
10. Cookies et traçeurs Web droit et accès facilité et élicite à l'utilisateur pour modification.	Oui	Non	Oui	Non	Oui	Non		
11. Analyse des autres formalités à engager https://intranet.inrae.fr/donnees-personnelles/Les-demarches-de-mise-en-conformite/Outil-d-aide-aux-formalites	Non	Non	Difficulté	Non	Non	Non		
12. soumission du RGPD à la DPO choisie	Non	Non	Non	Non	Non	Non		
PGD (plan de gestion de données)	CIRM-L	CIRM-CFBP	CIRM-CF	CIRM-BP	CIRM-BIA	SAMBO	UBOC	MISP
Référent PGD	Grondin Cécile	Perrine Portier	David Navarro	Emmanuelle Helloin	Anne-Sophie Bage			
Mail référent	cecile.grondin@inrae.fr	perrine.portier@inrae.fr	david.navarro@inrae.fr	emmanuelle.helloin@inrae.fr	anne-sophie.bage@inrae.fr			
1. Informations sur la structure								
a. Nom de la structure	Oui	Non	Oui	Oui	Oui	Non		
b. Type de structure	Oui	Non	Oui	Oui	Oui	Non		
c. Identification de la structure	Oui	Non	Oui	Oui	Oui	Non		
d. Responsabilité dans la structure	Oui	Non	Oui	Oui	Oui	Non		
e. Etablissement(s) tutelle(s)	Oui	Non	Oui	Oui	Oui	Non		
f. Département de rattachement INRAE	Oui	Non	Oui	Oui	Oui	Non		
g. Financeur(s) (permettant l'acquisition des jeux de données – hors projet)	Oui	Non	Oui	Oui	Oui	Non		
2. Informations sur le plan de gestion								
a. DOI	En cours	Non	En cours	En cours	En cours	Non		
b. Historique des versions	En cours	Non	En cours	En cours	En cours	Non		
c. Rédigé par	En cours	Non	En cours	En cours	En cours	Non		
3. Présentation générale des données.								
a. types et nature des données	Oui	Non	Oui	Oui	Oui	Non		
b. recueillies et/ou produites	Oui	Non	Oui	Oui	Oui	Non		
c. La production des données, la saisie des données, le suivi de la qualité des données ainsi que leur stockage et leur partage (qui réalise ces différentes étapes)	Oui	Non	Oui	Oui	Oui	Non		
4. Origine, Mode d'obtention et Droits de propriété intellectuelle des données								
a. Origine	Oui	Non	Oui	Oui	Oui	Non		
b. Mode d'obtention	Oui	Non	Oui	Oui	Oui	Non		

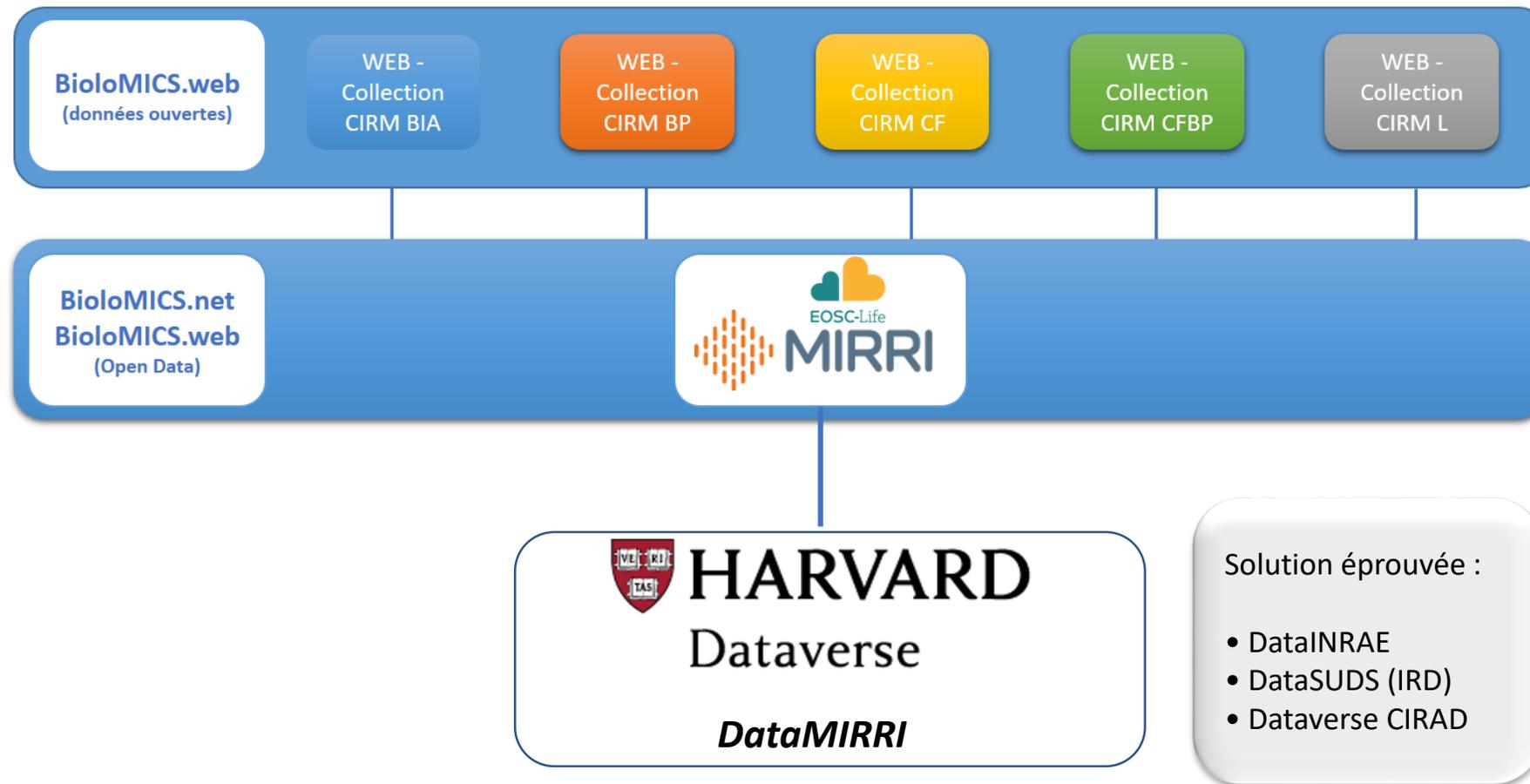
Outil déployé pour les CRB du CIRM et l'IR RARE

Pour toute demande d'information et/ou avoir accès aux documents :
celia.michotey@inrae.fr
anne-sophie.bage@inrae.fr
jonathan.mineau@inrae.fr

> Data-management au CIRM

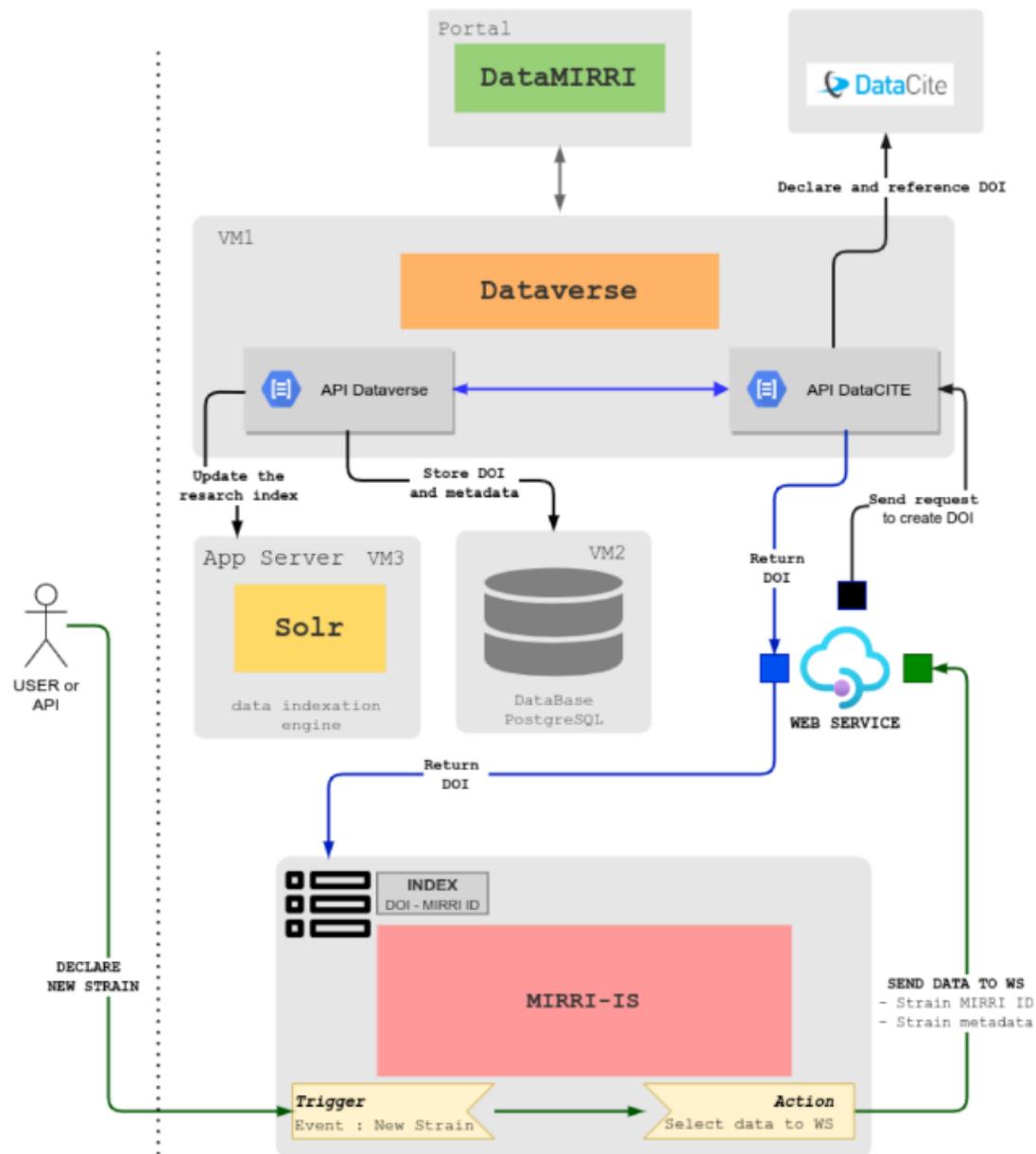
- **DM:** FAIRisation des données

> DM : FAIR – Dataverse MIRRI



> DM : FAIR – dataverse MIRRI

- Livraison fin 2022
- 1 IE informaticien à plein temps (EOSC-LIFE)
- Piloté par le data-manager du CIRM
- Dataverse déployé en Espagne (CECT)
- Permet **l'attribution de DOI** à toutes les ressources microbiennes, et le dépôt des datasets associés et de leurs métadonnées (elles-mêmes déposées dans l'entrepôt européens **DataCITE**)
- Pour l'instant en phase de développement après 6 mois de test en POC.
- Interactions fortes avec les équipes de BioAWARE et le WP1 d'IS-MIRRI-21
- Projet d'élaboration d'un **vocabulaire contrôlé** pour les métadonnées en cours.



> Data-management au CIRM

- **DM: Communication WEB**

> DM : Communication WEB



Amélioration du **site historique** sous eZpublish



Construction d'un **nouveau site** sous eZpublish



Construction d'un **nouveau site** sous eZplatform (solution ISC des CN-UE/OC)

- **Réunions mensuelles**
- En cours de finalisation après **1 an de conception et concertations**
- Livraison 1^{er} trimestre 2022
- Choix d'un **site de vulgarisation et de routeur** vers les sites des CRB (accès à leur collection sous Biolomics.WEB)
- Piloté et animé par le responsable national du CIRM et le data-manager

> Data-management au CIRM

- **DM: Bilan**

> Bilan

Certains freins limitants

- **Approches multidisciplinaires**, demande ressources et compétences diverses.
- **Décalage** entre les exigences et les moyens à disposition.
- **Changement de paradigmes**, parfois fortes réticences des producteurs de données.
- **Évolutions rapides** et conjointes des projets (nationaux et européens)

Des bonnes pratiques à respecter

- **Travail collaboratif** et non réservé à un domaine (qualité, informatique, ..)
- **Approches itératives** et évolutives dans le temps.
- **Savoir pourquoi** on le fait.
- Nécessité de s'appuyer sur des **réseaux** institutionnels et associatifs.
- **Formations** nécessaires.
- **Les outils et structures** d'aides sont là, cherchons-les et utilisons-les.
- **Communication, vulgarisation, dialogue.**
- Gérer **l'équilibre** dans l'investissement que demandent ces exigences.

➤ Merci pour votre écoute !

