

Le projet **R4multidata** a été officiellement ouvert le 28/01/2025.

Objectif : La communauté R4multidata vise à créer un environnement standardisé pour tester et comparer des fonctions de packages R avec des données réelles et simulées. Au départ, les packages RGCCA et mixOmics seront utilisés, avant de s'intéresser à d'autres packages. Quatre méthodes sont envisagées prioritairement : l'analyse canonique des correspondances régularisée généralisée sparse ou classique ((sparse)RGCCA) et la régression PLS sparse ou classique ((sparse)PLS).

Livrables :

- Un compte rendu des analyses et comparaison (méthodologie appliquée, résultats) avec des recommandations
- soit un webinaire ouvert à tous les membres de CATIs, soit des interventions dans des Assemblées Générales ou animations scientifiques de CATIs, équipes ou unités des agents ayant participé au projet.
- une communication via le forum de la DipSO, et lors des séminaires de cohésion CATI/PEPI/Pépinière.
- les jeux de données, scripts et dockers d'exécution qui auront servi au travail de comparaison, mis à disposition via un entrepôt de données (data.gouv.fr) sous licence, et la forge institutionnelle.

Phases du projet :

- préparation du travail de comparaison de fonctions de packages R (5 mois ⇒ janv – mai 25) : Théorie des méthodes; établissement des critères de comparaison, et du plan de comparaison pour prendre en compte différentes structures de données qui peuvent impacter les résultats; mise en place des outils collaboratifs de travail; préparation des jeux de données réelles et simulées
- travail de comparaison lors d'un hackathon (2 jours ⇒ mai 2025): programmation sous R, application aux données, synthèse des résultats obtenus.
- l'analyse et mise en forme des résultats (3 mois ⇒ juin - aout 25)
- la diffusion des résultats (3 mois ⇒ sept-nov 25)

Etat d'avancement :

préparation du travail de comparaison de fonctions de packages R

- outils collaboratifs mis en place: projet sur la forge institutionnelle, espaces collaboratifs de travail, tutoriels pour l'utilisation des outils
- jeux de données: les données de 3 projets ont été packagées et déposées sur la forge institutionnelle

travail de comparaison lors d'un hackathon

le hackathon s'est déroulé du 20 au 22 mai 2025 à Balma

l'analyse et mise en forme des résultats

en cours

la diffusion des résultats

une présentation du projet a eu lieu:

- lors de l'AG du CATI codex (mars 2025)
- lors du séminaire interCATI/PEPI 2G (juin 2025)
- lors des journées du PEPI IBIS (octobre 2025)

un site internet est en cours de finalisation pour la diffusion des résultats.

reporting financier au 17/06/2025 :

recettes : 6500

- SaPI : 6000
- CATI empreinte : 500

dépenses : 4598.2

- hackathon (réservation du lieu - somme engagée): 4230
- frais de déplacement pour le hackathon (somme engagée) : 368.2

From:
<https://pepi2g.wiki.inrae.fr/> - pepi2g

Permanent link:
<https://pepi2g.wiki.inrae.fr/doku.php?id=communaute:sapi:sapi:sapis2025:r4multidata&rev=1765794672>

Last update: **2025/12/15 11:31**

