RÉSEAU NATIONAL
DES BIBLIOTHÈQUES
DE MATHÉMATIQUES

GT-DONNÉES

AG du RNBM

12-13 septembre 2022

RÉSEAU NATIONAL
DES BIBLIOTHÈQUES
DE MATHÉMATIQUES

GT-DONNÉES

- Membres, organisation, objectifs, projets et réalisations du GT
- Pourquoi parle-t-on des données ?
- Contextes international et national
- Stratégie portée par le MESR et focus sur Recherche Data Gouv
- Et les Maths ? Présentation des résultats de l'enquête
- Accompagnement technique et documentaire
- Annonce du Webinaire

MEMBRES ET ORGANISATION

- **Brigitte Bidegaray** - Laboratoire Jean Kuntzmann, Grenoble
- Catherine Araspin - Bibliothèque IMAG, Montpellier
- Céline Benoit - Bibliothèques MIR, Paris
- **Céline Smith** – Mathdoc, Grenoble
- Christophe Berthon - Laboratoire de Mathématiques Jean Leray, Nantes
- Dominique Barrère – Bibliothèque IMT, Toulouse
- Elisabeth Kneller - Bibliothèque Jacques Hadamard, Orsay
- Henri Massias* – Laboratoire XLIM, Limoges
- Laurent Facq* - Institut de Mathématiques de Bordeaux
- Lucie Albaret – Bibliothèques Universitaires, Grenoble
- Nadège Arnaud – Bibliothèque LMV, Versailles
- Nathalie Granottier – Bibliothèque CIRM, Marseille
- Olivier Labbe - Mathdoc, Grenoble
- Paolo Lai - INIST, Vandœuvre-lès-Nancy
- Romain Vanel - Institut Fourier, Grenoble
- Sandrine Layrisse* - Institut de Mathématiques de Bordeaux
- **Violaine Louvet*** - GRICAD, Grenoble

Contact : gt_donnees@listes.rnbm.org

Co-animatrices du GT
Membres du sous-groupe sur les Forges*

RÉSEAU NATIONAL
DES BIBLIOTHÈQUES
DE MATHÉMATIQUES

OBJECTIFS ET RÉALISATIONS

Objectifs

Proposer aux membres des laboratoires de mathématiques les **services et l'accompagnement** les plus pertinents autour des données de la recherche.

Projets et réalisations

- Enquête sur les données de la recherche : Résultats et perspectives

- **Formation** des membres des laboratoires de mathématiques, du RNBM et de Mathrice aux données de la recherche :

  → Webinaire sur la rédaction des PGD sous l'angle des besoins de la communauté mathématique prévu le 6 octobre

- Rédaction de **guides et fiches pratiques** à destination de la communauté scientifique des mathématiciens (à lancer)

POURQUOI PARLE-T-ON DES DONNÉES ?

Parce qu'on n'a pas le choix …

- De **nouvelles obligations** pour les acteurs de la recherche : par exemple plans de gestion de données, ouverture des données et des codes des projets financés sur fonds publics, exigences de certains éditeurs …

- Et donc des **sollicitations à venir** pour les mathriciens et les membres du RNBM !

Parce que la science est un **bien commun** et que le partage de toutes les productions scientifiques doit être une évidence

- Concrètement ce partage nécessite des **changements méthodologiques** dans le processus de recherche

- Mais aussi la mise à disposition de **services et d'outils adaptés**

- Et un **accompagnement technique et documentaire** : donc nous !

AU NIVEAU INTERNATIONAL ET EUROPEEN

Research Data Alliance (RDA)

- **Organisation internationale** basée sur les contributions de ses membres qui développe de l'infrastructure et des activités communautaires pour réduire les obstacles au partage et aux échanges de données, et pour accélérer l'innovation à l'échelle mondiale en misant délibérément sur les données.
- Plus de **12 000 membres** venant de 146 pays, de multiples disciplines, domaines et thématiques et différents types d'organisations à travers le monde.
- **Auto-organisation** en Groupes de Travail spécialisés (Working Groups) et en Groupes d'Intérêt prospectifs (Interest Groups) pour partager et échanger expertise, constats et connaissances nouvelles, discuter obstacles et solutions potentielles, explorer et définir des politiques, et mettre à l'épreuve ou harmoniser des standards, dans le but d'**améliorer et de faciliter le partage et la réutilisation** des données à l'échelle mondiale.

European Open Science Cloud (EOSC)

- Ambition de l'Europe pour construire un **réseau de services et de données pour la recherche**
- En s'appuyant sur les **infrastructures existantes**
- Existence d'un **portail des services**
- Structuration de la gouvernance à travers l'**EOSC Association**
- De **nombreux appels à projets** à travers le pilier Infrastructures de Recherche de la commission européenne (financements conséquents)

RÉSEAU NATIONAL
DES BIBLIOTHÈQUES
DE MATHÉMATIQUES

AU NIVEAU NATIONAL

Plan National Science Ouverte v1 (2018) et v2 (2021)

- Structurer, partager et ouvrir les **données de la recherche**,
- Ouvrir et promouvoir les **codes sources** produits par la recherche,
- **Transformer les pratiques** pour faire de la science ouverte le principe par défaut
- S'inscrire dans une **dynamique durable**, européenne et internationale

Structuration au niveau du MESR : COmité pour la Science Ouverte (COSO)

- Composé de plusieurs instances qui proposent des **orientations**, instruisent les dossiers, effectuent des arbitrages, impulsent et accompagnent les actions associées pour la **mise en œuvre coordonnée du PNSO**
- **Copil / Secrétariat Permanent** (MESR, Universités, Ecoles, Organismes de recherches, Couperin, ANR, HCERES, …)
- **5 collèges** : publications, données de la recherche, europe et international, compétences et formation, logiciels
- **2 groupes d'expertises** : édition scientifique ouverte, juridique

Mise en place d'un **Administrateur Ministériel des Données, des Algorithmes et des Codes** (AMDAC) : Isabelle Blanc

- Chaque établissement et organisme doit désigner un **Administrateur des Données** (des Algorithmes et des Codes) de la Recherche

STRATÉGIE PORTÉE PAR LE MESR : LES PILIERS STRUCTURELS

RÉSEAU NATIONAL DES BIBLIOTHÈQUES DE MATHÉMATIQUES

Publications : HAL

science ouverte

Données : Recherche Data Gouv

Codes : Software Heritage

RÉSEAU NATIONAL
DES BIBLIOTHÈQUES
DE MATHÉMATIQUES

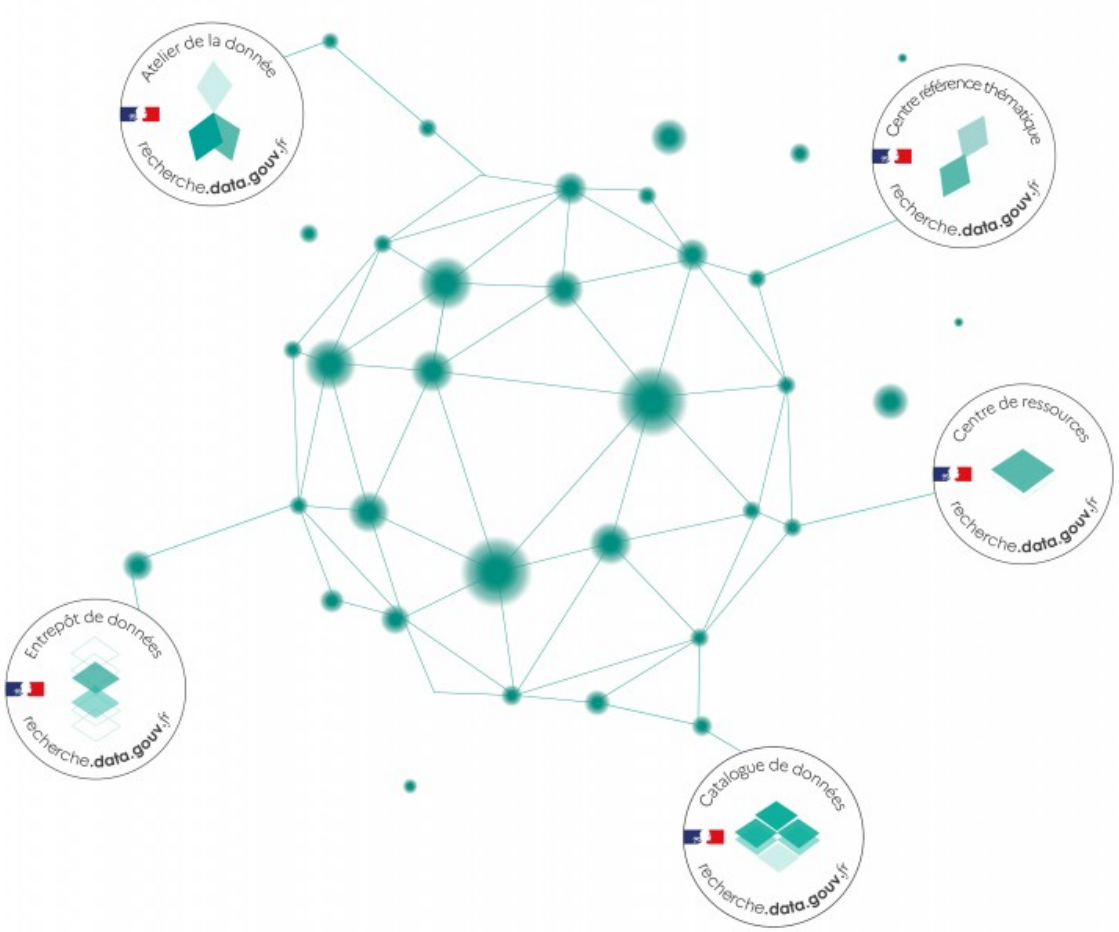
RECHERCHE DATA GOUV

Trois modules pour accompagner les équipes de recherche sur toute question relative aux données :

- Des ateliers de la donnée
- Des centres de référence thématiques
- Des centres de ressources

Deux modules pour déposer, publier et signaler des données :

- Un entrepôt pour déposer et utiliser des données
- Un catalogue pour rechercher les données publiées sur l'entrepôt ou sur des entrepôts externes



RÉSEAU NATIONAL
DES BIBLIOTHÈQUES
DE MATHÉMATIQUES

ET LES MATHS ?

Une enquête sur les usages et besoins de la communauté mathématique autour des données et des codes de recherche

- Identifier les pratiques autour des données (et des codes) pour les mathématiciens, sachant qu'elles sont la plupart du temps très disciplinaires

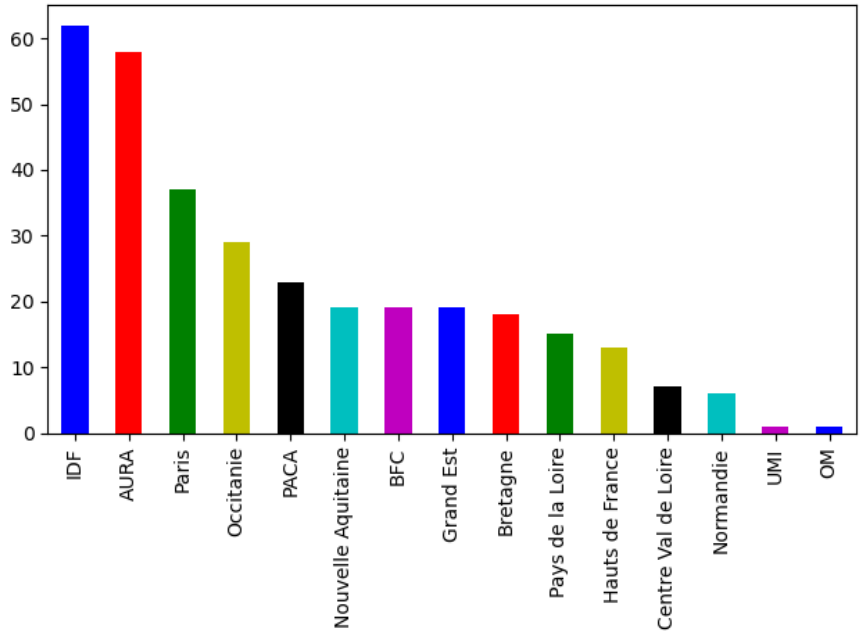
.

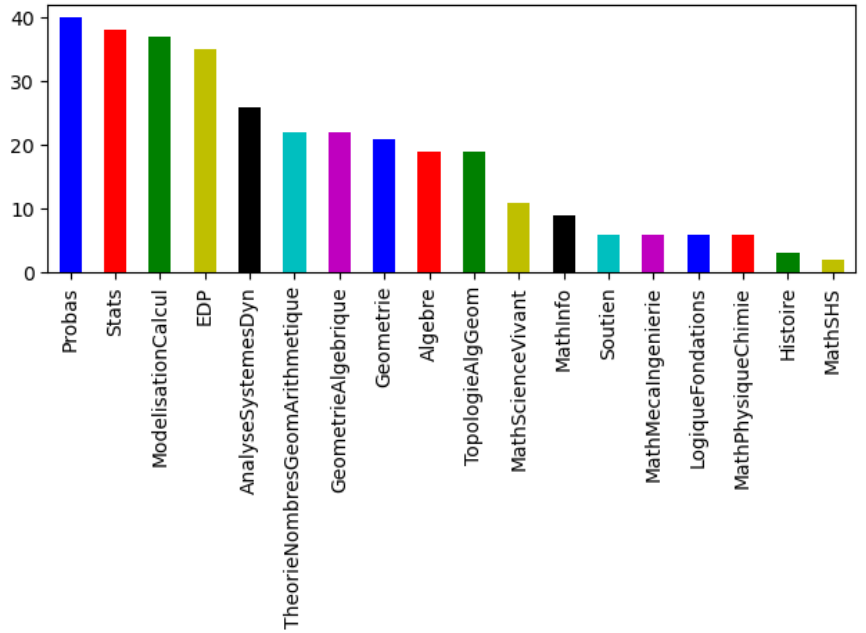
- Comprendre les freins, manques, succès pour proposer des services et un accompagnement adapté

Profils des répondants

- 352 réponses

- Bonne répartition à la fois au niveau géographique, dans les différentes sous-disciplines des mathématiques et dans les différentes fonctions (chercheurs, enseignants chercheurs, personnels d'appui, doctorants …)





QUI EST CONCERNÉ PAR LES DONNÉES ET LES CODES?

RÉSEAU NATIONAL DES BIBLIOTHÈQUES DE MATHÉMATIQUES

Données : 48 % concernés, 45 % pas du tout et 7 % ne sait pas.

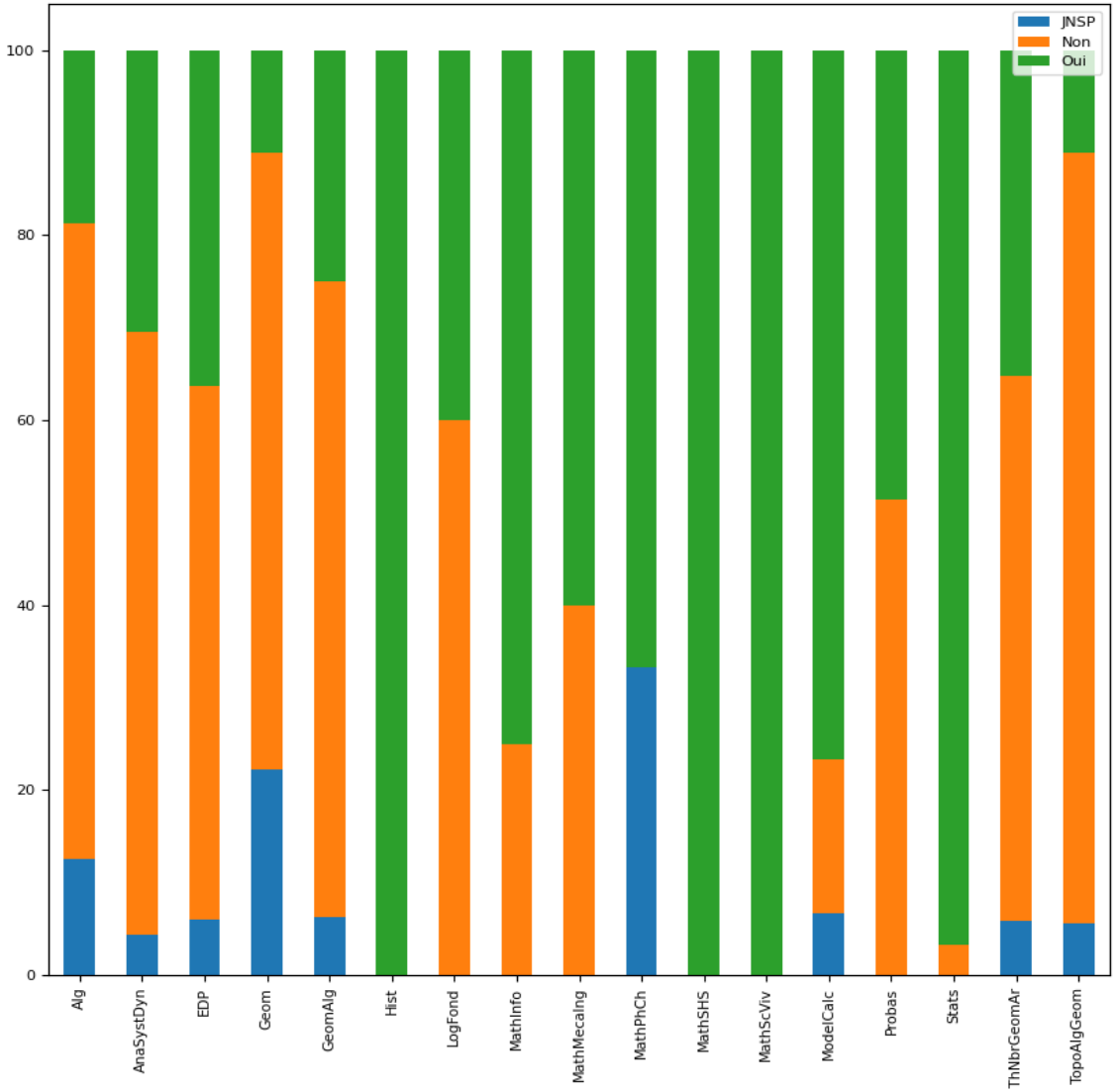
- La **compréhension** de ce que sont les données de recherche n'est pas toujours claire. Amalgame avec les publications.

Différences disciplinaires importantes

Codes : 75 % utilisent des logiciels pour leur recherche, 33 % développent du code

Volumes des données faibles à très faibles

Répondants concernés ou non par les données en fonction des disciplines (JNSP = Je ne sais pas)



RÉSEAU NATIONAL
DES BIBLIOTHÈQUES
DE MATHÉMATIQUES

QUELS TYPES DE STOCKAGE ?

- Majoritairement l'**ordinateur professionnel** (environ 38%)
- Puis les **serveurs de laboratoire** (environ 20%)

- Outils cités :
  - Ceux de la PLM : PLMBox, gitlab
  - Les outils établissements à base de nextcloud ou owncloud
  - Les outils commerciaux (Dropbox, Google …), en particulier dans le cadre de collaborations à l'international (mais seulement 5%)

- **Sauvegarde**
  - Ordinateur professionnel (26%)
  - Serveurs de labo (24%)
  - DD externe (20%)

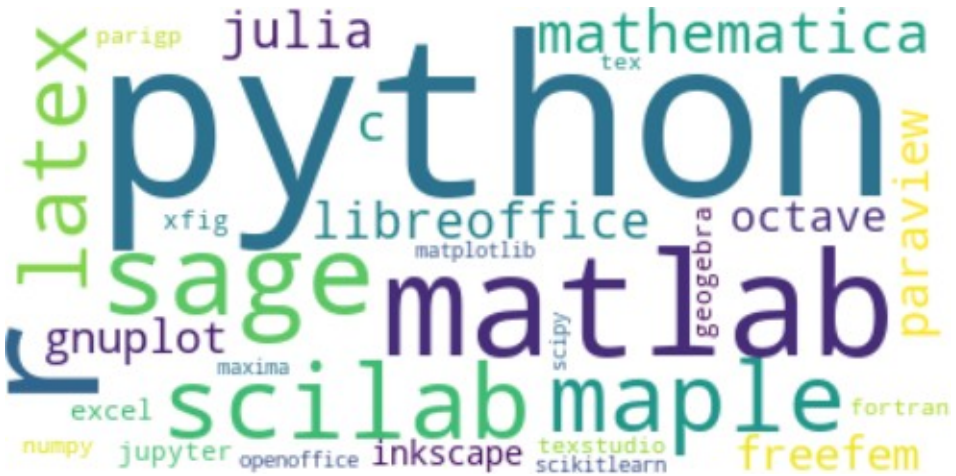
- Beaucoup de **pratiques individuelles**, des difficultés exprimées pour partager les données, le manque de simplicité des solutions proposées ...

QUELS TRAITEMENTS ? UTILISATION ET DÉVELOPPEMENT DE CODES

RÉSEAU NATIONAL DES BIBLIOTHÈQUES DE MATHÉMATIQUES

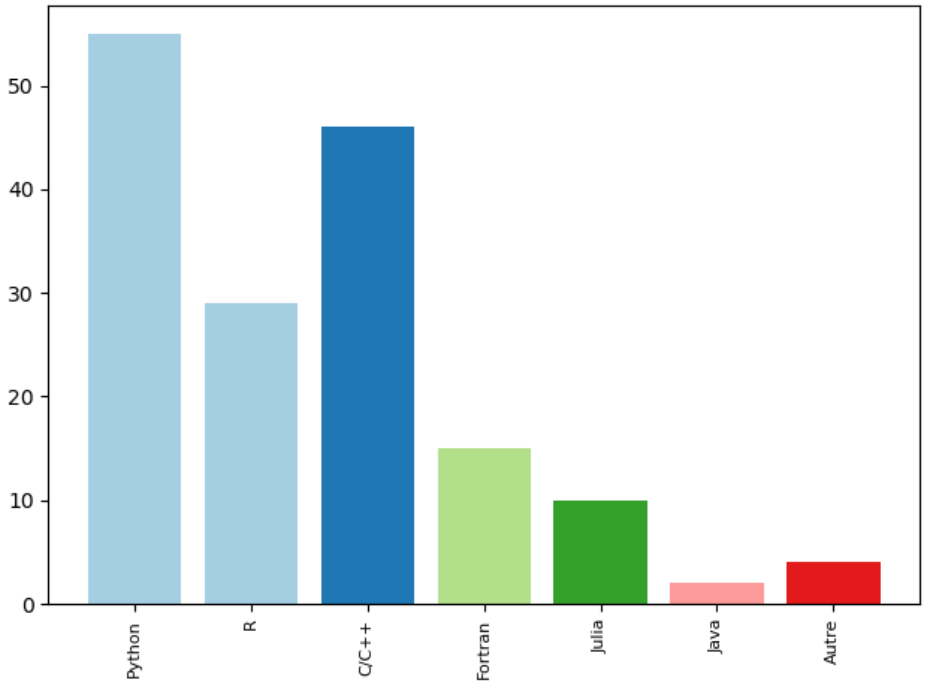
- Traitements réalisés : simulations, statistiques, représentation graphique, vérification de conjectures, post-traitement, ...

- **Besoin en ressources très limité** : l'ordinateur professionnel suffit dans la grande majorité des cas



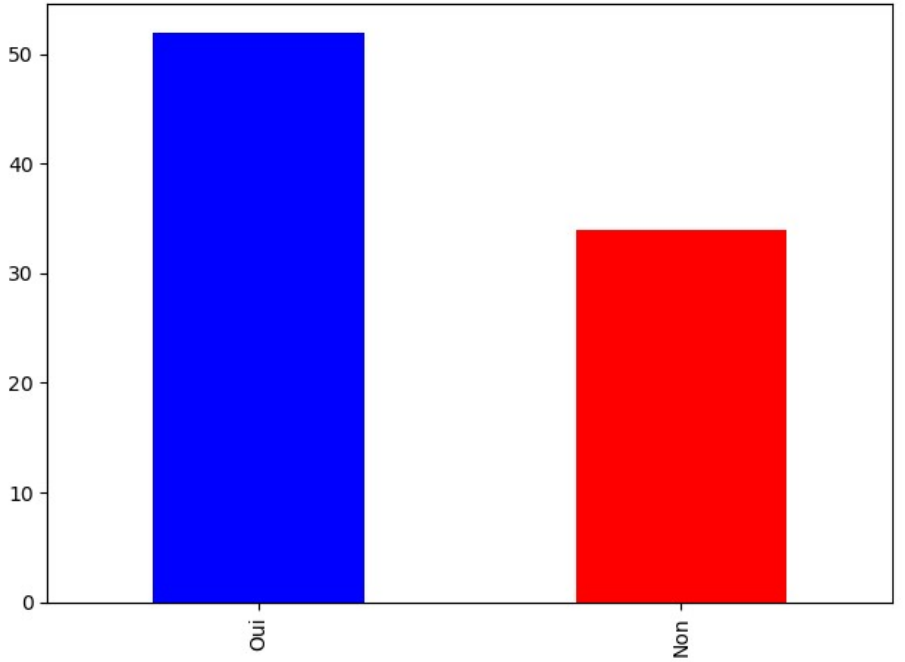
Logiciels / outils utilisés

PRATIQUES DE DÉVELOPPEMENT



Langages de programmation

Utilisation de forges



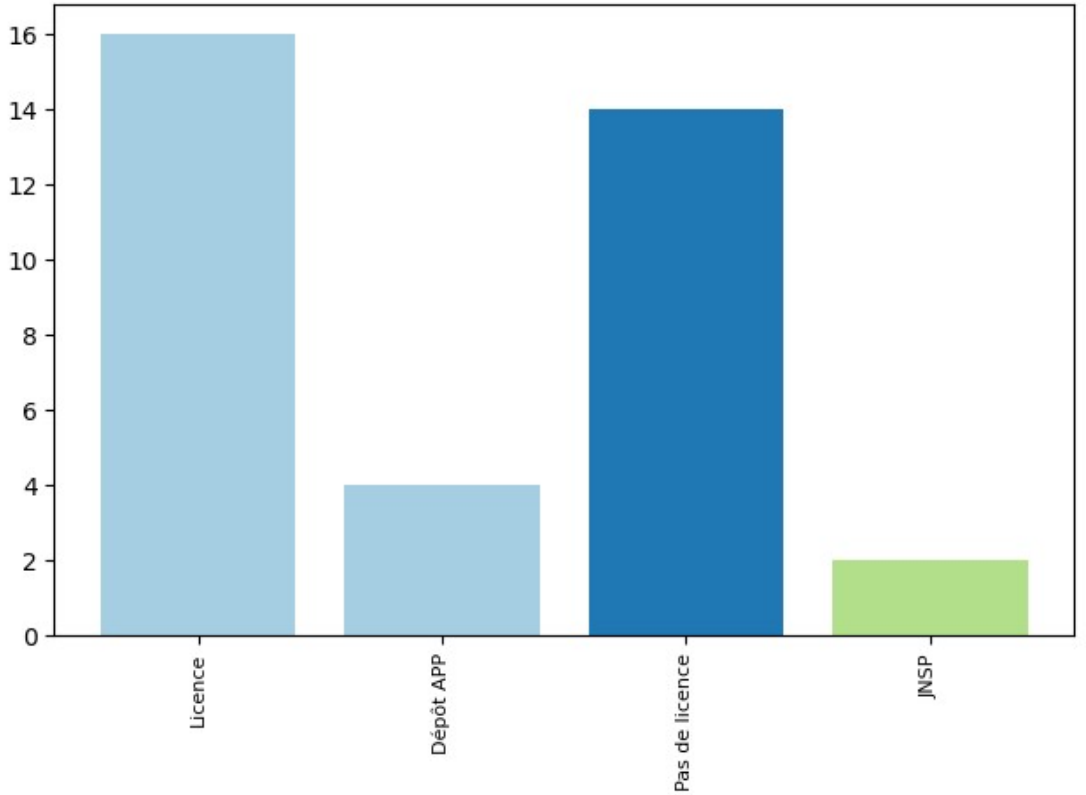
DIFFUSION ET PARTAGE DE CODES

- Développements **majoritairement partagés** : avec sa communauté (37%), avec les partenaires des projets (35%), avec son équipe (32%)

- Nombreux codes **partagés sans licence**. Licences très majoritairement **open source** sinon.

- Les **freins** pour la mise sous licence :

  - Le partage avec des collègues proches n'est pas considéré comme nécessitant la mise en place d'une licence
  - Manque de connaissance du sujet
  - Absence d'intérêt



RÉSEAU NATIONAL
DES BIBLIOTHÈQUES
DE MATHÉMATIQUES

DIFFUSION DES DONNÉES

- **Plan de Gestion de Données**
  - 80 % ne connaissent pas
  - Parmi ceux qui connaissent, seuls un peu moins de 30 % ont participé à la rédaction

- Plus de 67 % ne connaissent pas d'**entrepôt de données**
  - Amalgame avec HAL, gitlab ...

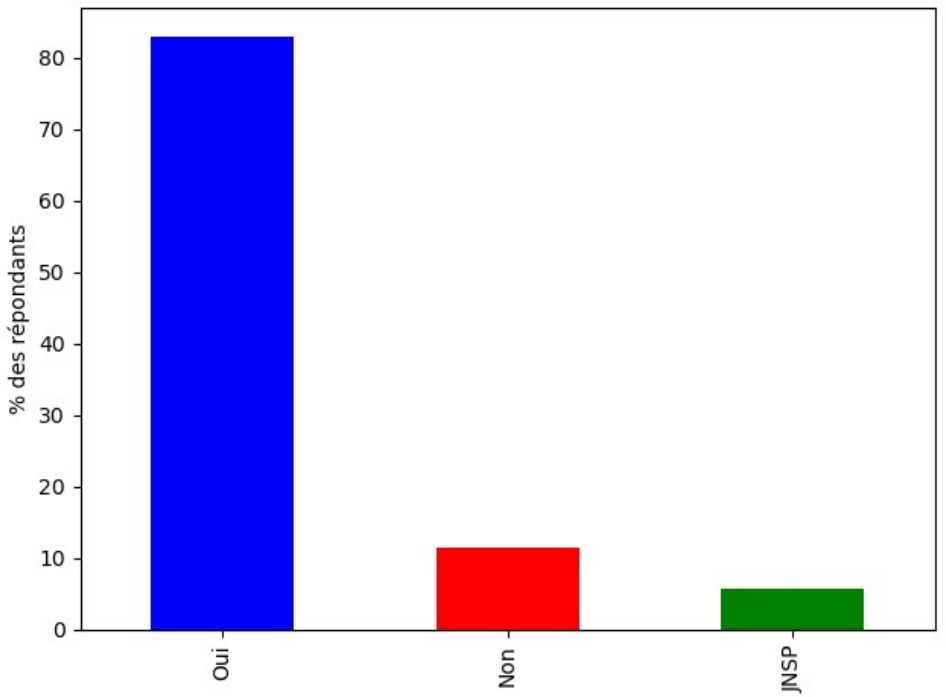
- Par contre, l'**intérêt de déposer ses données** est bien perçu :
  - Gain en visibilité
  - Accès libre aux recherches pour tous
  - Possibilité de vérifier et reproduire les résultats

- Les **freins** au dépôt :
  - Manque d'informations sur les entrepôts et le processus de dépôt
  - Problématique réglementaire, propriété intellectuelle
  - Données estimées non pertinentes
  - Manque de temps pour curer
  - Manque de valorisation pour le chercheur

REPRODUCTIBILITÉ, RÉUTILISATION

- **Partage de données et de codes** : pratique assez courante
  - Par mail, transfert de fichiers, clé USB … sur demande de collègues
  - Ou organisé via des clouds ou des forges

- **Réutilisation fréquente** des données et/ou codes de collègues



Reproductibilité des résultats de leurs recherches d'après les répondants

RÉSEAU NATIONAL
DES BIBLIOTHÈQUES
DE MATHÉMATIQUES

SYNTHÈSE DE L'ANALYSE

Compréhension très diverse de la notion de données de recherche, et une grande hétérogénéité entre les disciplines des mathématiques sur le fait ou non de manipuler des données.

Certaines thématiques **clairement pas concernées** et les processus associés comme les demandes de Plans de Gestion de Données, ou l'ouverture des données pas adaptés : suscitent de l'incompréhension voire du rejet.

Problématique des **codes de recherche beaucoup plus partagée.**

Stockage des données : usages extrêmement divers, avec beaucoup de pratiques individuelles et de système D.

Problème du partage des données avec des collaborateurs extérieurs souvent mentionné.

Volumétries considérées généralement assez faibles. Idem pour les **ressources de traitement / calcul**.

Peu de personnes concernées par des **données sensibles** (au sens du RGPD)

RÉSEAU NATIONAL
DES BIBLIOTHÈQUES
DE MATHÉMATIQUES

SYNTHÈSE DE L'ANALYSE

Activité autour des codes de recherche importante, avec des pratiques certes à consolider mais globalement plutôt justes (en particulier sur l'utilisation des forges).

Ancrage fort dans le logiciel libre tant sur l'utilisation des logiciels que sur la production de codes de recherche.

Partage très intégré dans les pratiques, mais pas forcément formalisé.

Problématique des licences : pratiques de partage sans licence.

Peu de **citation des logiciels** utilisés dans les publications.

Sensibilité forte sur la reproductibilité des résultats de recherche.

Concept d'**entrepôt de données** peu connu, de même que les **Plans de Gestion de Données**.

Besoins d'accompagnement et de formations très généraux, tant sur les données que sur les codes.

β ACCOMPAGNEMENT TECHNIQUE ET DOCUMENTAIRE

RÉSEAU NATIONAL DES BIBLIOTHÈQUES DE MATHÉMATIQUES

- **Évolutions des métiers et des compétences** pour accompagner la gestion et le partage des données, des algorithmes et des codes
  ➔ Voir par exemple cet article

- Importance de la **complémentarité des compétences techniques** (informaticiens) **et documentaires** (bibliothécaires)

- Importance du **travail en réseau nationaux** (CoSO, Mathrice, RNBM) **et internationaux** (RDA, EOSC)

Comment s'organise-t-on collectivement pour répondre aux demandes des chercheurs concernant les données de la recherche (rédaction de PGD, etc…) ?

SOUS-GROUPES DE TRAVAIL : FORGES

Suite à la présentation de la synthèse de l'enquête faite par Violaine lors des JournéesMathrice2022, le GT-Données a validé la mise en place d'un sous-groupe dédié aux forges logicielles coordonné par Henri (réunion de lancement : 1er septembre).

Objectif :

- Référencer les supports de formation existants à l'utilisation des forges logicielles pour les mutualiser ;
- Monter des actions de sensibilisation à l'utilisation des forges logicielles (GitLab) : imaginer différents niveaux de formation ;
- Référencer les forges logicielles existantes et utilisées au niveau des laboratoires et des universités et identifier les freins à leur utilisation (problèmes d'identifications : nécessité de création d'un compte chez l'établissement hébergeur?).

Membres : **Henri**, Laurent, Sandrine, Violaine ; échanges sur Mattermost

Contact « Hotline » : codes-donnees@math.cnrs.fr

SOUS-GROUPES DE TRAVAIL : MÉTADONNÉES

Appel aux bibliothécaires du RNBM pour **rejoindre le GT-Données** et monter un groupe de travail spécifique autour des **standards de métadonnées pour les mathématiques**.

Objectif : proposer un standard de métadonnées pour l'**indexation thématique des données et des codes** dans les entrepôts de données et ainsi en faciliter la recherche. Il y a déjà des réflexions autour des métadonnées de description des codes au niveau de Software Heritage. L'idée serait de proposer des métadonnées d'indexation spécifiques aux maths et adaptées aux données et aux codes (à partir de la MSC).

Membres recherchés : collègues chercheurs et bibliothécaires du RNBM en coopération avec des chercheurs et bibliothécaires à l'étranger.

Rejoignez-nous : gt_donnees@listes.rnbm.org !

WEBINAIRE SUR LA RÉDACTION DE PGD

RÉSEAU NATIONAL DES BIBLIOTHÈQUES DE MATHÉMATIQUES

Jeudi 6 octobre 2022 de 13h à 14h

La plupart des porteurs de projets financés sont désormais confrontés à la **rédaction d'un PGD**. Or ce document n'est pas toujours bien compris et sa finalité bien appréhendée. L'objectif de ce webinaire est de présenter les PGD **sous l'angle des besoins de la communauté mathématique** et d'apporter les réponses à toutes vos questions, telles que :

- En tant que chercheur / chercheuse en Maths, pourquoi ai-je intérêt à rédiger un PGD ?
- Parmi les outils existants lequel utiliser ? Et quel type de modèle ?
- À qui m'adresser localement pour un accompagnement ciblé ?

Lien zoom

RÉSEAU NATIONAL
DES BIBLIOTHÈQUES
DE MATHÉMATIQUES

OPEN SCIENCE DAYS @ UGA

Du 13 au 15 décembre 2022

Les Open Science Days@UGA ont vocation à être un lieu d'échange sur la science ouverte, à l'échelle du site de Grenoble et de façon beaucoup plus large au niveau national.
La session 2022 sera consacrée aux codes et logiciels de recherche.

- **Keynotes :**
  - Histoire du logiciel libre et recherche académique
  - Reproductibilité et réutilisabilité
- **Sessions :**
  - Questions juridiques et valorisation
  - Production logicielle et évaluation des carrières
  - Communautés
- **Tutoriaux :**
  - Gitlab
  - Guix
  - Jupyter notebook
  - Software Heritage et Hal pour référencer les codes

Présentiel (Grenoble) et distanciel (sauf tutoriaux)
https://osd-uga-2022.sciencesconf.org/