

# Cartographie des actions des réseaux métiers autour de la gestion des données

## Synthèse

Version 1.0 (22 mai 2017)

Initiative soutenue et financée par la MI Plateforme des réseaux et suivi par la DIST-CNRS .  
Animation : Caroline Martin (MEDICI) et Emmanuelle Morlock (RENATIS).

### Réseaux impliqués

- RENATIS (Colette Orange/ Phillippe Eyraud/ Emmanuelle Morlock)
- MEDICI (Stéphane Renault / Caroline Martin)
- QeR (Alain Rivet)
- RESINFO (Olivier Band-Foissac, Maurice Libes)
- RBDD (Marie-Claude Quido/Geneviève Romier)
- CALCUL (Anne Cadiou)
- DEVLOG (Lyriane Bonnet)

## **I- Bref historique :**

Ce groupe est le fruit de plusieurs actions convergentes. En 2015 : RENATIS prend initie de premiers échanges avec d'autres réseaux ainsi que l'identification de plusieurs thématiques communes (annuaire, formation, droit, données). Une réunion a été organisé le 15 avril 2015. En janvier 2016, lors des ateliers des journées réseaux au CNRS (12-13 janvier), l'atelier « données » a été l'occasion de construire la base d'une réflexion sur la gestion des données porteuse d'une vision « métiers » et « réseaux ».

Cette action se poursuit désormais dans le cadre d'un groupe transversal à plusieurs réseaux.

## **II-Activités du groupe depuis sa création**

- **5 réunions en visioconférences en 2016 (18/02, 01/06, 12/07, 22/09, 12/12)**

Production: rapport des comptes rendus accompagné d'une cartographie schématisée des activités des réseaux autour du cycle de vie des données.

- **Une journée d'étude 19/10/2016**

Objectifs : Echanger et intégrer des regards différents sur le sujet, approfondir notre réflexion et nos travaux avec une vision extérieure ; établir une feuille de route pour 2017 sur ce sujet pour le groupe.

La matinée : 10H-13H :

- Échanges avec des extérieurs : Francis André (DIST) - Alain Bénard (PEPI données- INRA)

L'après-midi : 14H-17H

- Travail du groupe pour fixer feuille de route et travail approfondi sur la cartographie des activités des réseaux autour des données. Livrable : une cartographie approfondie des activités des réseaux autour de la gestion des données.

- **2 réunions en présentiel / visioconférence au 1<sup>er</sup> trimestre 2017 (13/01, 15/03)**

### **III-Objectifs du groupe :**

- Etablir et diffuser une vision transversale de la gestion des données afin d'enrichir la pratique de chaque réseau dans le domaine des données et permettre le développement de la complémentarité entre réseaux ;
- Valoriser l'apport des expériences et expertises « métier » dans une vision transversale de gestion de données dans les réseaux technologiques et scientifiques de la MI ;
- Sensibiliser les communautés professionnelles de l'appui à la gestion des données (organisation de journées thématiques par exemple);
- Identifier les problématiques concernant les « data » sur chaque réseau (livrables à définir) ;
- Mise en commun et partage de nouvelles pratiques en réseau et au sein de chaque réseau.

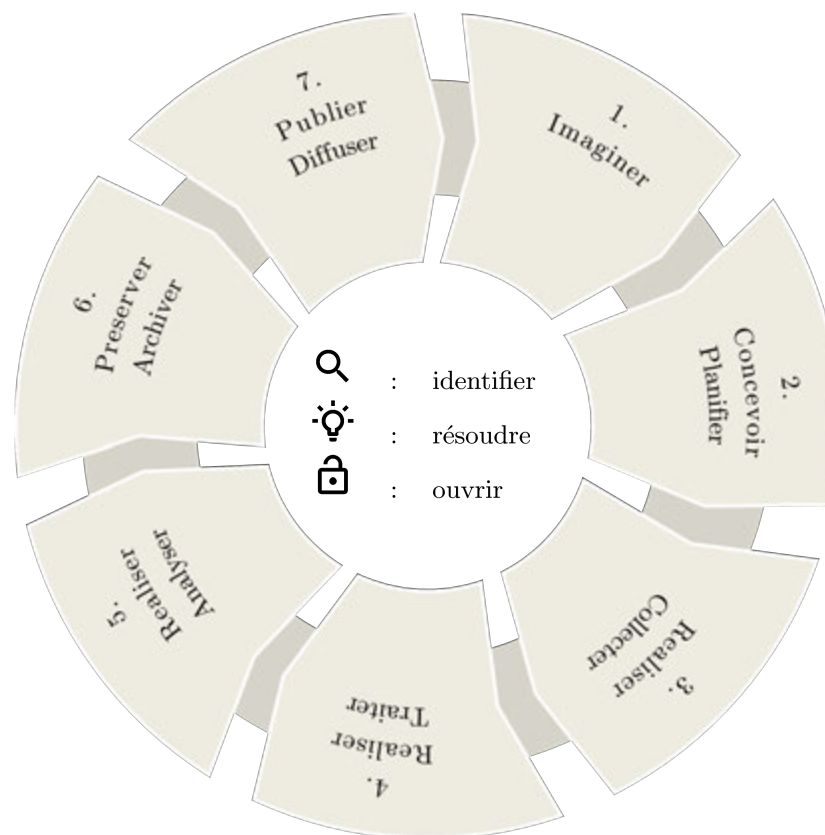
Ces objectifs pourront se concrétiser au travers d'actions d'information et sensibilisation au sein de chaque réseau et également plus largement auprès des écoles doctorales et autres secteurs.

#### IV-La cartographie :

**Objectif : à partir d'une réflexion de type « cycle de vie de la donnée », cartographier les usages dans chaque réseau autour de la gestion de la donnée.**

Cf. pages suivantes. Voici pour rappel la représentation visuelle des différentes étapes du cycle tel qu'il a été défini par le groupe de travail. Les différents réseaux ont structuré leurs réponses autour de trois questions :

- « Identifier » : repérage des manques ou des interrogations principales concernant les données ;
- Résoudre : actions et solutions mises en oeuvre ;
- Ouvrir : manques ou perspectives à explorer.



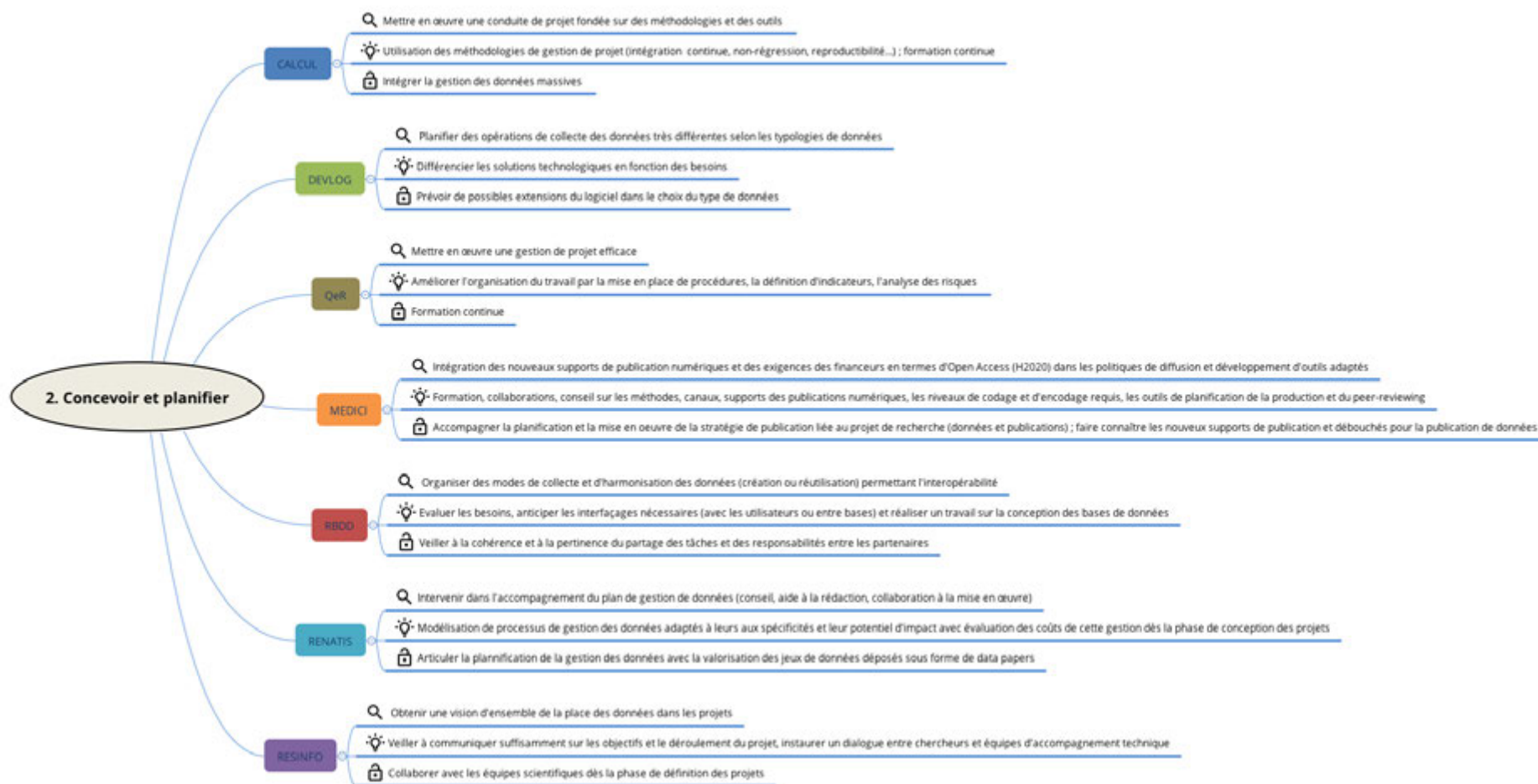
## Etape 1 : Imaginer (idée de recherche : état de l'art, veille, fouille de texte et de données (TDM))

Les réseaux participent à cette étape du cycle de vie en identifiant les problématiques techniques, en proposant des outils technologiques les plus pertinents pour accompagner la construction, la mise en œuvre du projet et la valorisation des résultats. Pour réaliser cet accompagnement, les réseaux organisent une veille sur les nouvelles technologies en émergence permettant le traitement et l'analyse des données, sur les nouveaux supports de valorisation de la recherche, les contraintes et opportunités qu'elles soient réglementaires, juridiques (politiques d'Open Data, H2020 par exemple) ou bien techniques, technologiques (TDM, ...) ou encore organisationnelles (plans de gestion de données). L'apport des réseaux dans cette étape est également important en termes de croisement des disciplines et des métiers permettant un accompagnement de haute valeur.



## **Etape 2 : Concevoir et planifier (Définition des tâches, élaboration de plannings, recherche de partenaires, de financement, spécifications, métadonnées, plan de diffusion ...)**

Les réseaux apportent un appui sur la gestion de projet des méthodologies de conduite de projet : définition des indicateurs, interopérabilité des systèmes, aide à la rédaction des plans de gestion de données, planification du type de données à collecter, identification des nouveaux supports de publications...). Il est nécessaire de prévoir le mode de collecte et de stockage afin d'organiser la traçabilité en amont susceptible de garantir la réutilisation des données.



### Etape 3 : Réaliser et collecter (constitution des jeux de données, métadonnées)

Les réseaux participent également à la constitution des jeux de données, des métadonnées et du choix des référentiels associés, et à la mise en place d'un environnement leur permettant d'assurer le stockage, le « versionning », et l'accessibilité des données et des jeux de données via des bases ou des supports fiables et bien documentés, dans le respect des règles de traitement spécifiques des données personnelles.



## Etape 4 : Réaliser et traiter (filtrer, regrouper, choisir les données pertinentes, reformater, gérer les métadonnées)

Les réseaux participent également au traitement des jeux de données via la mise en place de procédures ou logiciels assurant la qualité des données et la qualité des traitements (modélisation et reformatage des données ; mise en base de données ; mise à disposition de moyens informatiques ; aide aux choix de la solution la plus adaptée, détection des erreurs de saisie ou des incohérences relatives aux modèles de données, mise en place d'un contrôle qualité).





## Etape 5 : Réaliser et analyser (exploitation des données et des métadonnées)

La participation à l'analyse des données se réalise au travers d'un accompagnement sur la mise en place de procédures qualité, par l'optimisation des calculs ou des temps d'accès, ainsi que dans le choix et la mise en œuvre des techniques et outils d'analyse et éventuellement dans leur conception, en fonction des besoins (formats, interoperabilité, performance, ergonomie, visualisation, travail collaboratif...)



## Etape 6 : Préserver et archiver (pérenniser la disponibilité, la traçabilité, la confidentialité, l'intégrité (sécurité) de l'ensemble nécessaire à la reproductibilité)

L'essentiel de l'accompagnement des réseaux sur cette étape se formalise par l'intégration des notions de pérennisation des données, de traçabilité et d'intégrité des données ainsi que des outils permettant de les exploiter. Les réseaux sont également amenés à participer à la définition des politiques de sauvegarde et d'archivage, et à la mise en place des procédures associées dans les collectifs de recherche.



## **Etape 7 : Publier et diffuser (choix du support de publication/diffusion, forme de diffusion, structuration de contenu, marketing de contenu, mise en visibilité (formatage, ...), brevet, dissémination)**

A cette étape finale d'un projet de recherche impliquant une gestion des données, l'accompagnement des réseaux s'exerce sur la documentation des données et leurs formats d'exploitation, pour en permettre l'accès et en assurer la reproductibilité ou la répliquabilité. Ainsi les réseaux travaillent sur l'ensemble des informations (données, modes opératoires, échantillons, publications, visualisation et interfaces graphiques) nécessaires à la mise en œuvre des supports de diffusion et de valorisation les plus pertinents en rapport avec l'objectif du projet initial.



## Synthèse générale :

Suite à, ce travail, on peut mettre en évidence quelques points saillants, parmi l'ensemble des points de recoupements dans les activités des réseaux participant à ce travail :

- Importance de la veille « métier »
  - sur les tendances, innovations en matière de technologies pour l'outillage tout au long du cycle de vie des données
  - dans la veille et l'analyse sur les nouveaux supports, les nouveaux usages pour diffuser et valoriser les jeux de données (utilisation d'identifiants pérennes, dépôts dans des entrepôts de données certifiés)
  
- Complémentarité de l'activité technique et scientifique au travers de l'accompagnement sur la mise en œuvre de moyens et conditions permettant la reproductibilité (replicabilité) et l'accès aux données dans les projets de recherche, dans le respect des restrictions nécessaires (données sensibles, propriété intellectuelle, embargos...).
  
- Utilisation systématique du mode projet dans l'accompagnement (aide à la rédaction de Data Management Plan, démarche qualité, analyse de risques, etc.)
  
- Volonté de dégager des thèmes communs dans lesquels des actions transversales inter-réseaux pourront s'inscrire.