



Bilan

20 novembre 2018

*Christine Collet, Vincent Claveau
Pierre Gançarski, Aurélien Garivier*

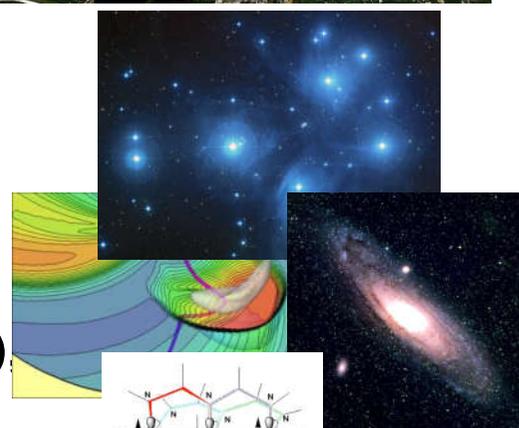
Masses de données (scientifiques)

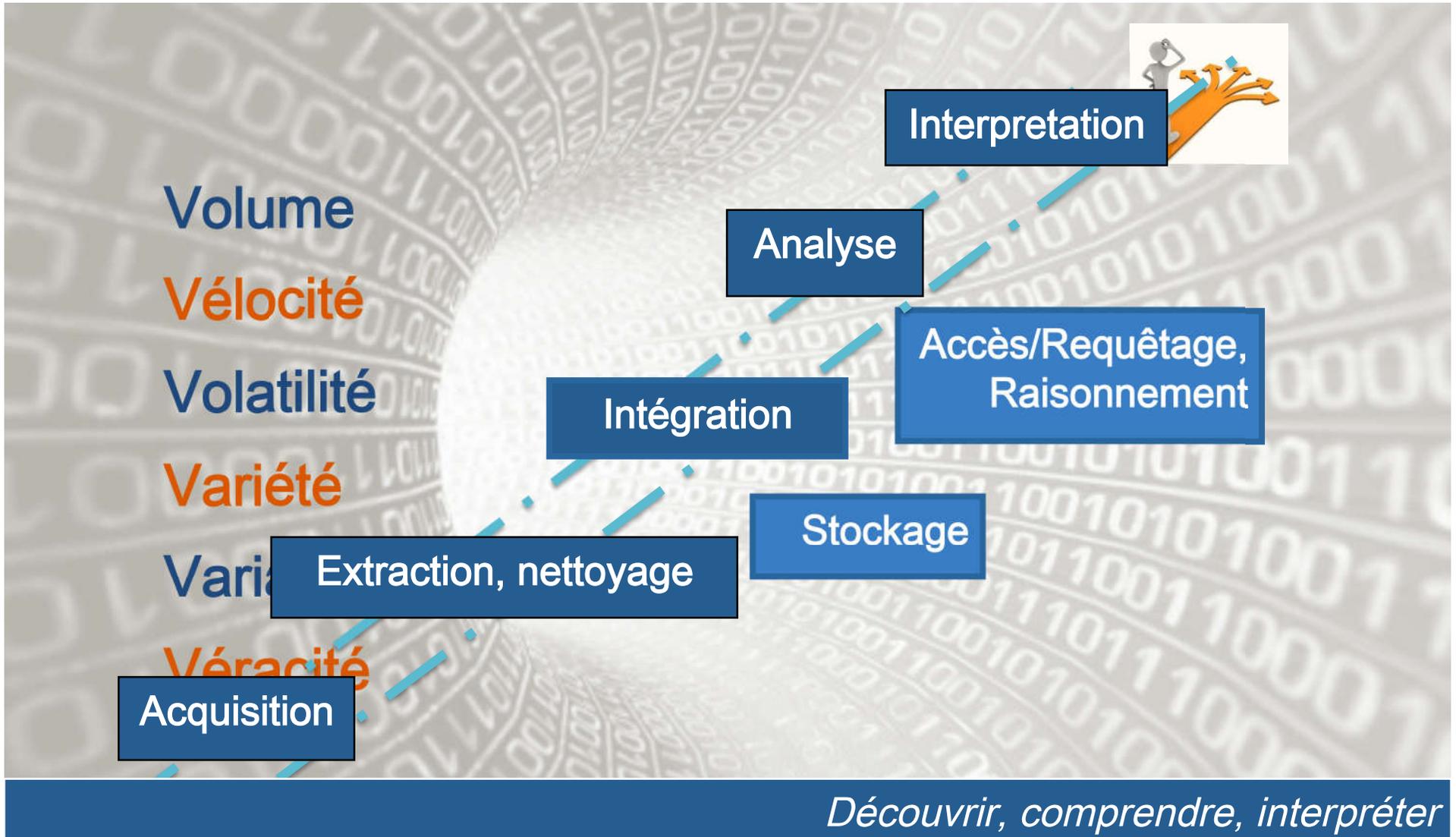
provenant d'instruments, de simulations numériques, de multiple dispositifs de collecte de données ... produites en grande quantité dont la taille difficilement appréhendable (**Volume**), à un rythme intense (**Vélocité**), ne pouvant plus être stockées (**Volatilité**), de nature très hétérogènes (**Variété**), avec des sémantiques variables au cours du temps (**Variabilité**), de qualité et précision peu contrôlables (**Véracité**).

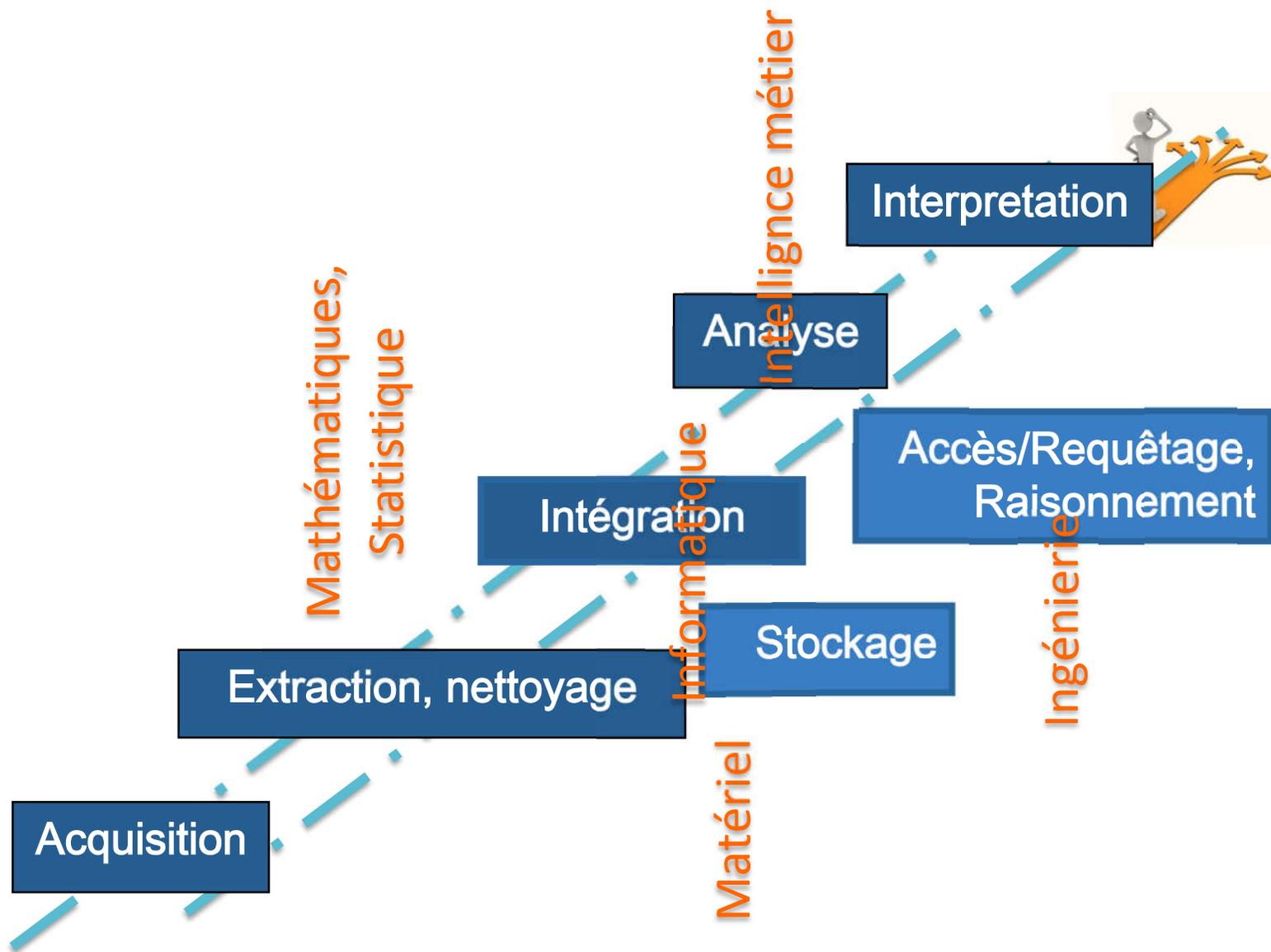
Nouvelle démarche scientifique

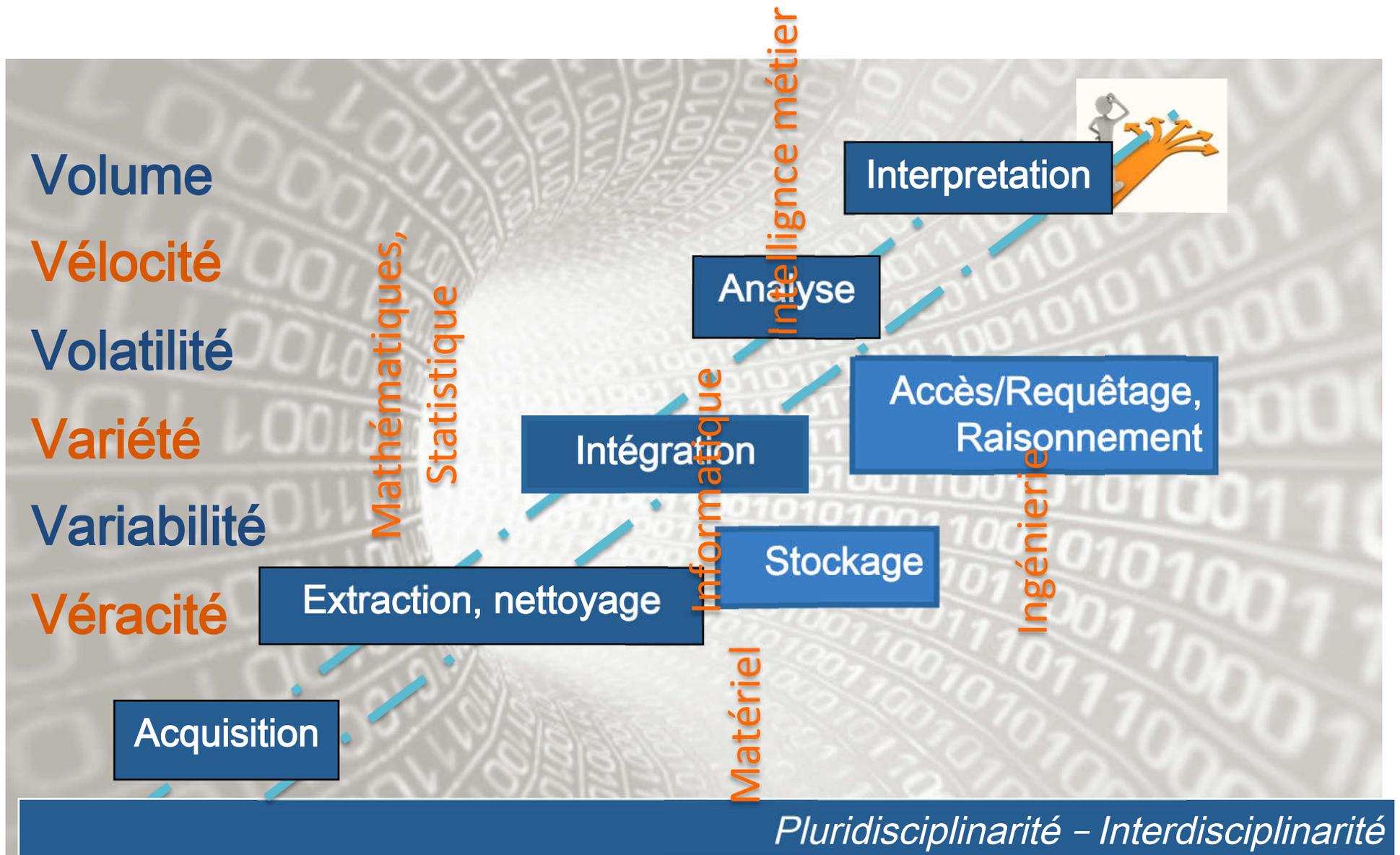
De : Croire -> Vérifier (avec des données)

A : Découvrir (dans les données) => Comprendre









Volume

Vélocité

Volatilité

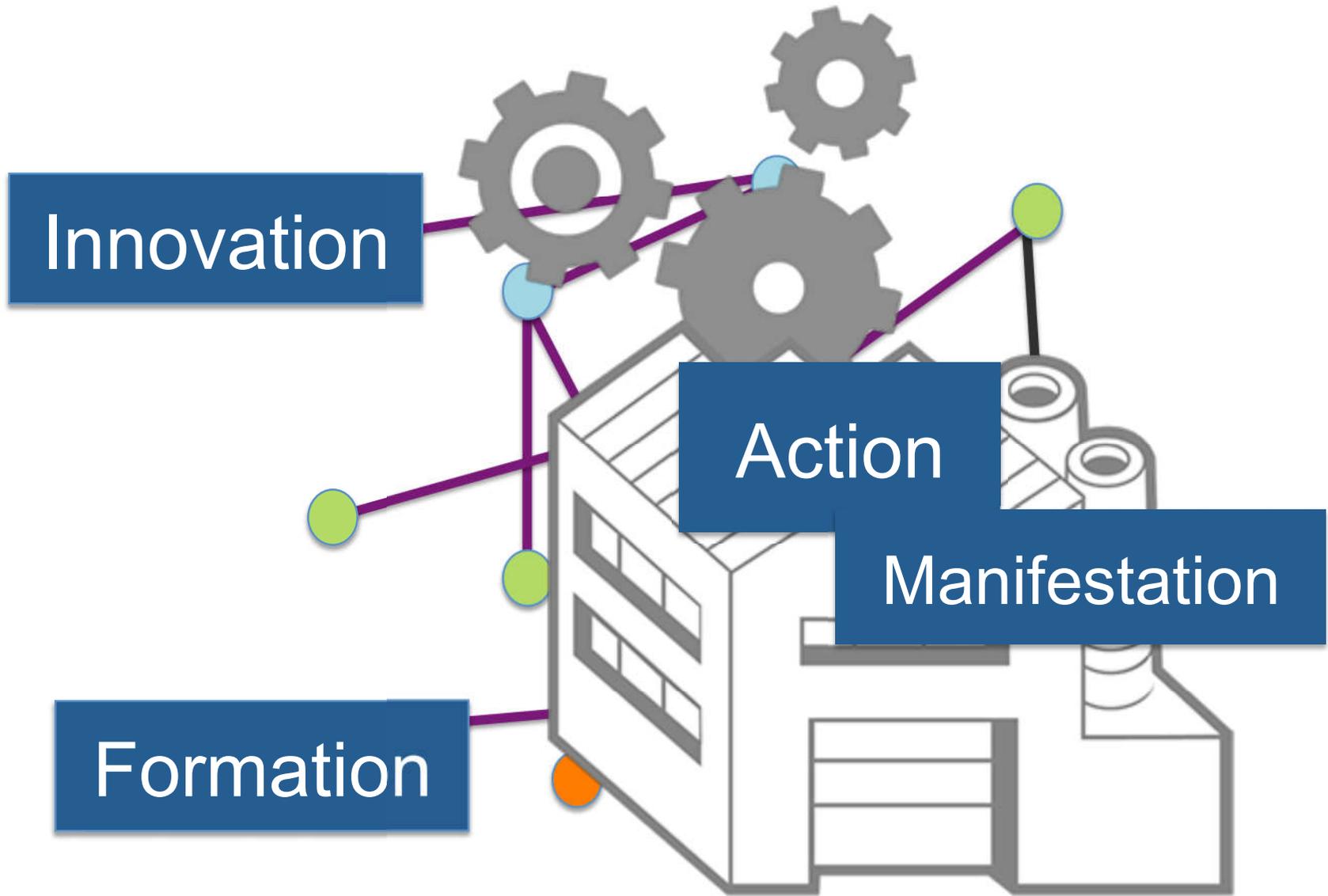
Variété

Variabilité

Véracité



MaDICS



- **Action** : coordonne diverses activités (Manifestations) autour d'un **projet d'animation interdisciplinaire**
- **Manifestation** : journée animation, groupe de travail, études, écoles, ateliers, hackathon ...
(problématiques, applications, données)
- **Réseau Formation en Sciences des Données** (soutien à des formations, **espace doctorants**)
- **Réseau innovation** : liens avec le monde socio-économique (**club des partenaires**)

Comité de direction : directeur e s. chargé e s. de

Comité (hors direction) : membres renouvelés au 1/3 tous les trois ans ; candidatures spontanées, suite à sollicitations, ex-responsables Actions ; décision bureau - avis consultatif du comité.

principales décisions et les choix de politique scientifique et de gestion.

Assemblées consultatives :

Assemblée générale / Rencontre annuelle : bilan des activités, lancement et discussions de perspectives de recherche

Comité d'Orientation Scientifique : apporte au GDR un **point de vue extérieur** sur la réalisation de son projet, sur son fonctionnement et organisation, et sur les orientations d'évolution tant scientifiques qu'organisationnelles.

Membres

- Christiane Weber, DR CNRS UMR TÉTIS, Montpellier
- Cédric Demeure, VP R&D France Thalès
- Vincent breton, DR CNRS IN2P3
- Pierre Senellart, Prof. ENS Paris
- Christel Vrain, Prof. Université d'Orléans
- Gersende Fort, DR CNRS, Toulouse

Pour le programme voir : <https://www.madics.fr/intranet/cos/>

Recommandations

Comité d'Orientation Scientifique du GDR
MaDICS

Originalité

- Interdisciplinarité
- structuration par actions

depuis sa création

Il note l'originalité de son positionnement interdisciplinaire autour des données massives de la science, qui est un domaine riche voire foisonnant. Une seconde originalité est sa structuration par actions proposées par la base. Elles sont toutes ancrées dans les concrets de communautés scientifiques variées, certaines abordant toute la chaîne des traitements, de la production à l'exploitation des connaissances. La reprise des actions initiées dans Mastodons semble tout à fait réussie ; de nouvelles actions ont aussi émergé.

Le conseil recommande au GDR de mieux mettre en avant la production scientifique dans ses rapports d'activités, même si ce n'est pas un des indicateurs prioritaires. Le GDR dont le rôle est avant tout de permettre à des communautés scientifiques de se rencontrer encourage la mise en valeur d'autres indicateurs de la production scientifique et de l'évaluation de l'impact des actions de type Mastodons.

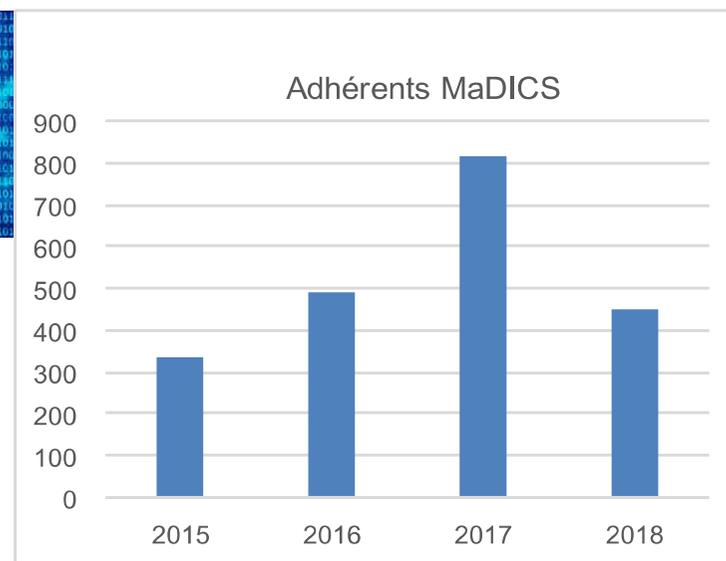
Le conseil recommande aussi d'insérer une description plus détaillée de la production scientifique dans les rapports d'activités. Il serait pertinent d'intégrer dans la gouvernance des personnes issues d'autres communautés scientifiques. Il serait aussi utile de structurer la communication du GDR vers ses membres pour mieux cibler les informations distribuées. Pour l'avenir, le COS encourage le GDR à continuer dans cet objectif d'animer une communauté interdisciplinaire. Il approuve l'objectif d'avoir 1/3 des membres issus de la recherche en informatique et 2/3 des membres du GDR issus des autres communautés scientifiques. Il semble important d'accroître la coordination notamment par rapport aux éléments de stratégie scientifique du GDR (formation, partenariat, plateformes etc.) et aux échanges entre les actions, qui constituent des écosystèmes vivants et actifs.

Les efforts pour accroître la participation des industriels sont tout à fait pertinents. La décision prise de lancer un Club des Partenaires, de manière à construire un pont entre le monde académique et industriel a également été appréciée (on remarque d'ailleurs une augmentation du nombre de représentants du monde industriel inscrits sur le site web du GDR). Cette initiative permettra aussi d'augmenter les moyens du GDR et du coup de financer de nouvelles actions.

- mieux mettre en avant la production scientifique ; autres indicateurs (Hackathon)
- autres communautés scientifiques dans la gouvernance
- structurer la communication vers ses membres
- accroître : la coordination (formation, partenariat, plateformes etc.) ; les échanges entre les Actions, qui constituent des écosystèmes vivants et actifs.

Le RA de la période

- Introduction
 - Contexte
 - Outils
 - Fonctionnement
- **Bilan général**
 - **Le GDR a 5 ans**
 - **La communauté**
 - **Actions**
 - **Manifestations**
 - **Réseaux**
 - **Autres éléments ou outils**
 - **Conclusion**
- **Bilan des Actions**



452 Membres

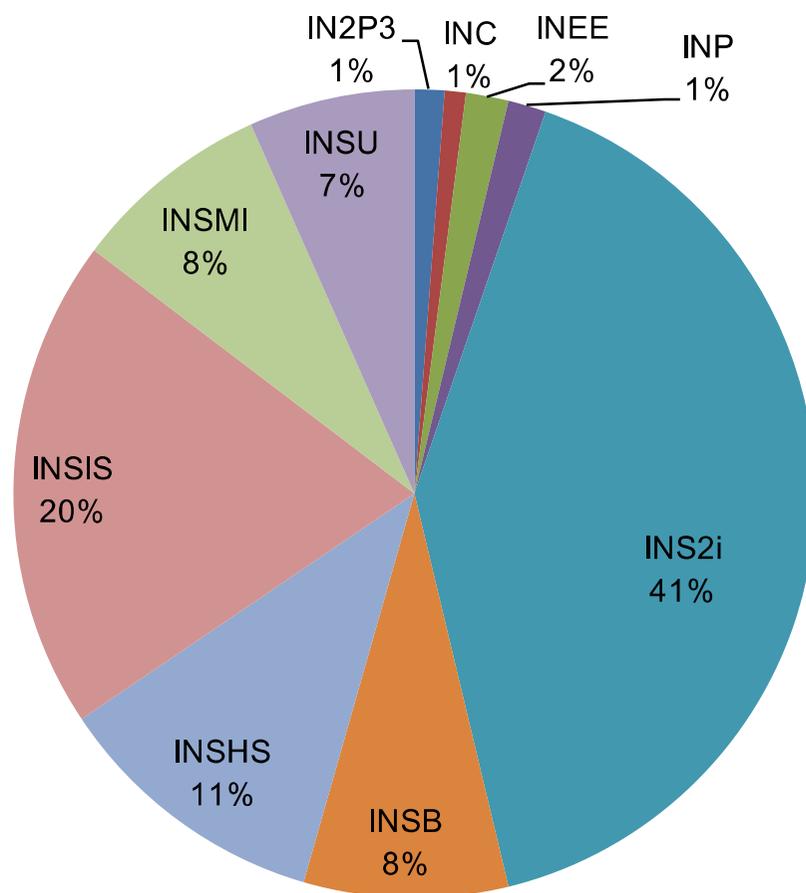
- 291 Chercheurs et Enseignant-chercheurs
- 63 Post-Doc + Doc
- 53 Industriels
- 45 Autres Instituts (INRA, INRIA, INSERM, INA, IRD, CEA, CHU, ...)

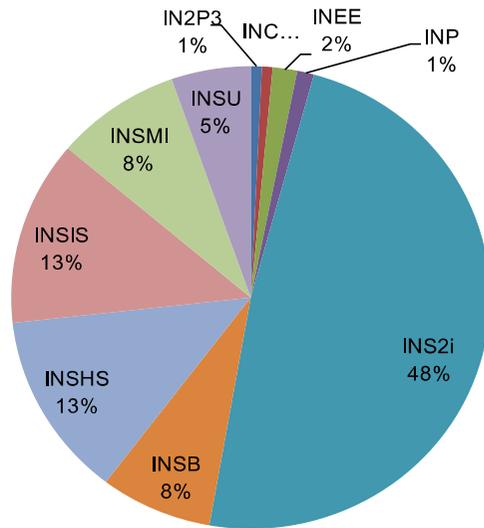
134 laboratoires - 292 équipes de recherche

75 % Chercheurs, 14% Post-doc + Doc

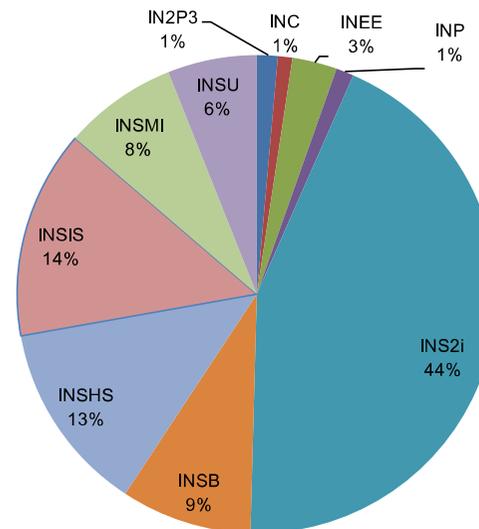
89% académiques, 11% industriels

Pluridisciplinarité

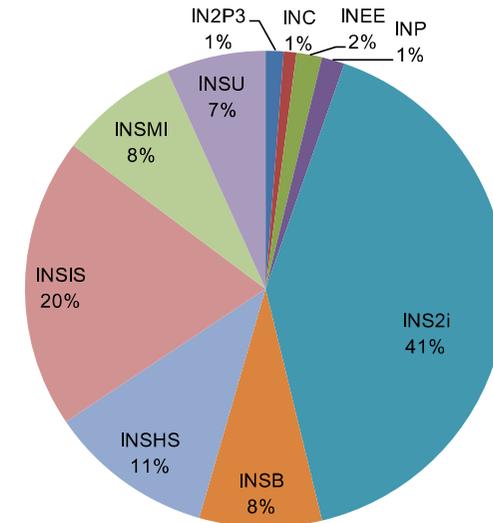




2016



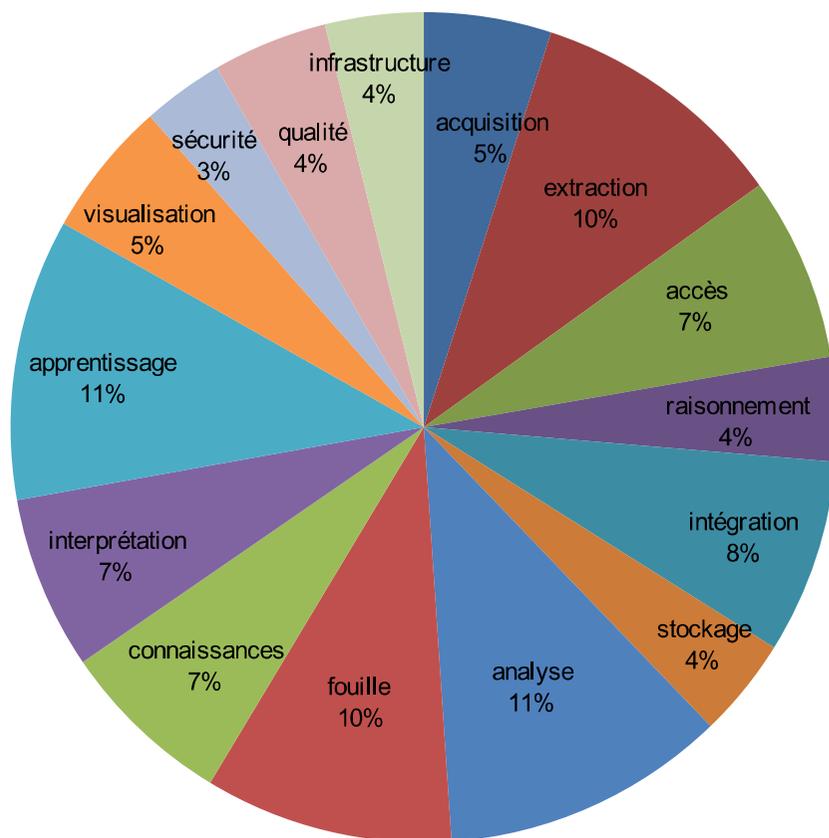
2017



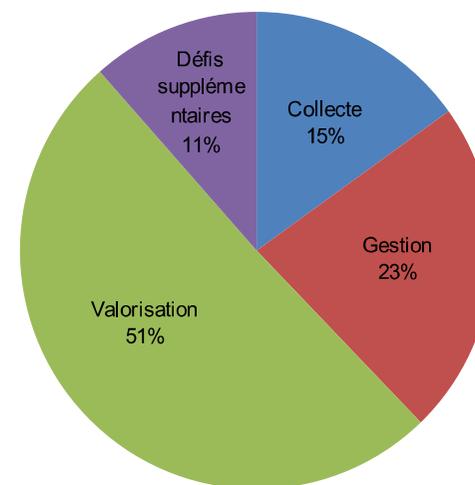
2018

INS2i : 48%, 44%, 41 %

Compétences



Domaines



Thématiques

Une Action propose un projet d'animation interdisciplinaire coordonnant diverses activités.

Sélection (un appel par an)

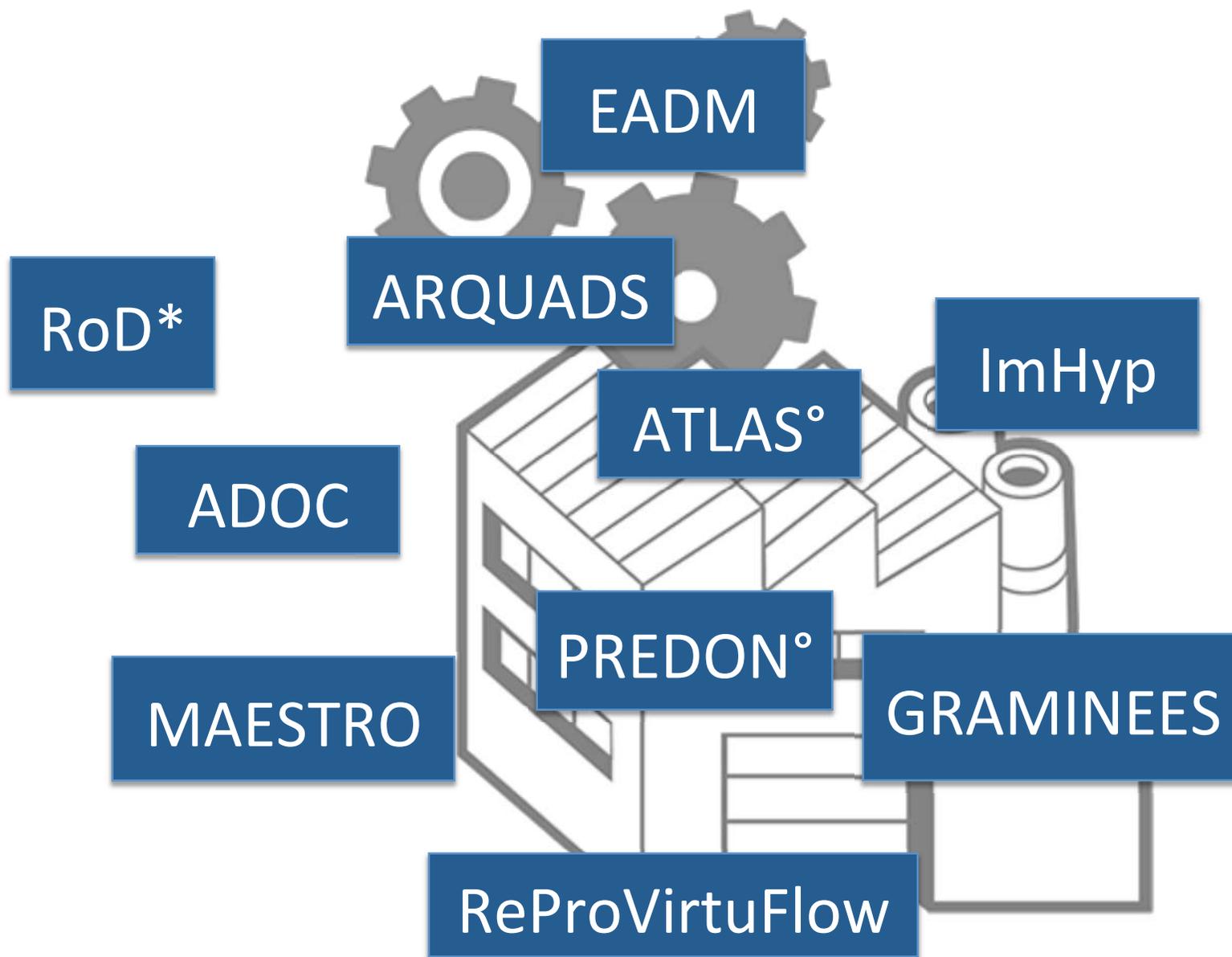
- Verrous scientifiques et techniques identifiés
- Nature des données (produites et consommées) clairement identifiée ;
- Au moins deux responsables de disciplines différentes (pas industriels)

Contraintes :

- Durée de vie limitée (2 ans, renouvelable une fois)
- Pas de format de fonctionnement type ; prend la forme qui paraît la plus adéquate aux responsables ; autonomie ;
- Manifestations affichées sur le site web MaDICS de l'Action, et communiquées largement (appel à participation)
- Un rapport annuel scientifique et financier en fin d'année (financement révisable chaque année)

Un correspondant attribué au sein du comité de direction

Actions en cours



* Action partagée GDR IA

ADOC : ENTrepôt et Analyse de DOcuments

Jérôme Darmont, ERIC, Olivier Teste, IRIT, INS2i, Natalia Grabar, STL, INSHS/ISCC

ATLAS : Apprentissage, opTimisation Large-échelle et cAlculs distribuÉS

Marianne Clausel, LJK, INSMI, Massih Amini, LIG, INS2i

ARQUADS : Action de Recherche sur la Qualité des Données Scientifiques

Allel Hadjali, LIAS/ENSMA, INS2i Angela Binifati, LIRIS, INS2i, Laure Berti-Equille, LIF UMR CNRS 7279, INS2i et Bastien RANCE, INSERM UMRS 1138, HEGP, AP-HP

EADM : Environmental Acoustic Data mining

Hervé Glotin, LSIS, INS2i, Frédéric Precioso, I3S, INS2i

MAESTRO : MAsses de données En aSTRONomie et astrophysique

Sofian Maabout, laBRI, INS2i, Christian Surace, LAM, INSU

GRAMINEES : GRaph data Mining in Natural, Ecological and Environnemental Sciences

Romain David, IMBE, Dino IENCO, TETIS, IRSTEA

ImHyp : Imagerie Hyperspectrale

Jocelyn Chanusot, GIPSA-Lab, INS2i, Mauro Dalla Mura, GIPSA-Lab, INS2i
Marie-Françoise Devaux, INRIA Nantes

PREDON : PREservation des DONnées scientifiques

Cristinel Diaconu, CPPM, IN2P3, Salima Benbernou, LIPADE Paris Descartes, INS2i

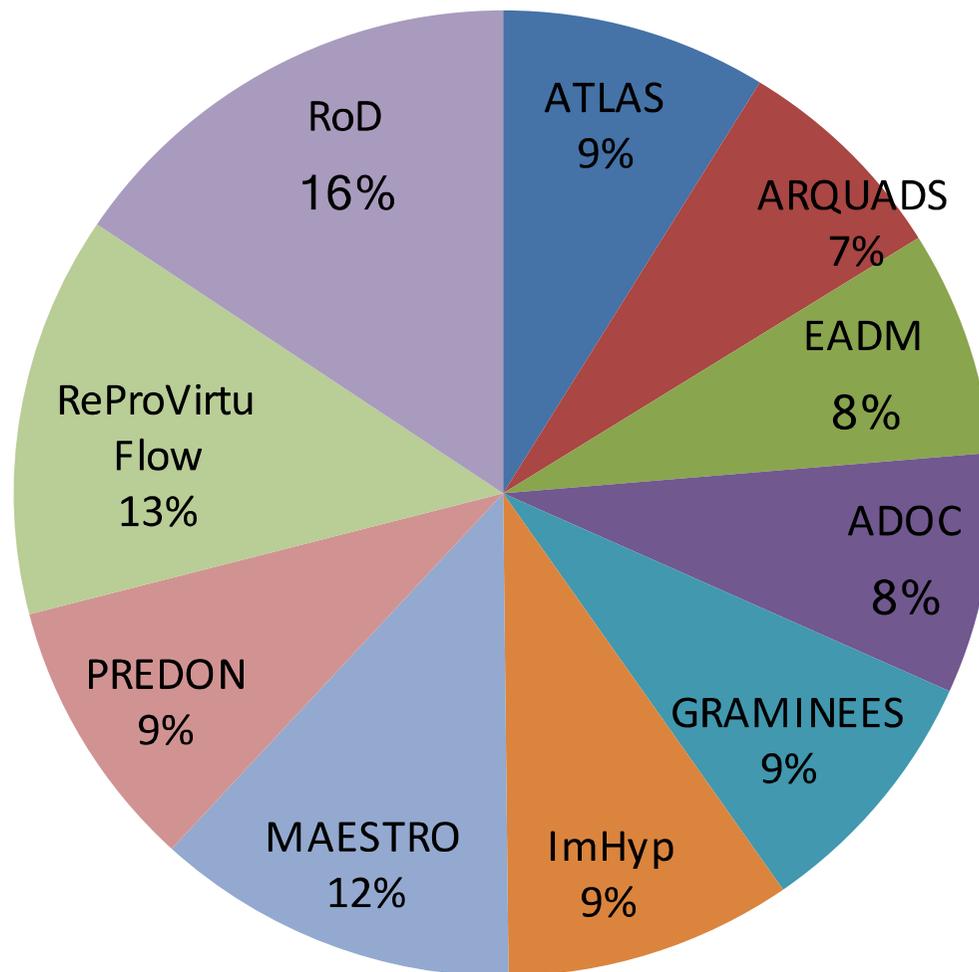
ReProVirtuFlow : Reproductibilité des expériences d'analyse de données scientifiques

Sarah Cohen-Boulakia, LRI, INS2i, Christophe Blanchet, IFB, INSB

RoD : Reasonner sur les données (Reasoning on Data)

Marie-Laure Mugnier, LIRMM, INS2i, Marie-Christine Rousset, LIG, INS2i,
Véronique Bellon, ITAP (UMR IRSTEA / SupAgro) Olivier Palombi, LADAF-CHU Grenoble-Alpes

MaDICSiens par Action



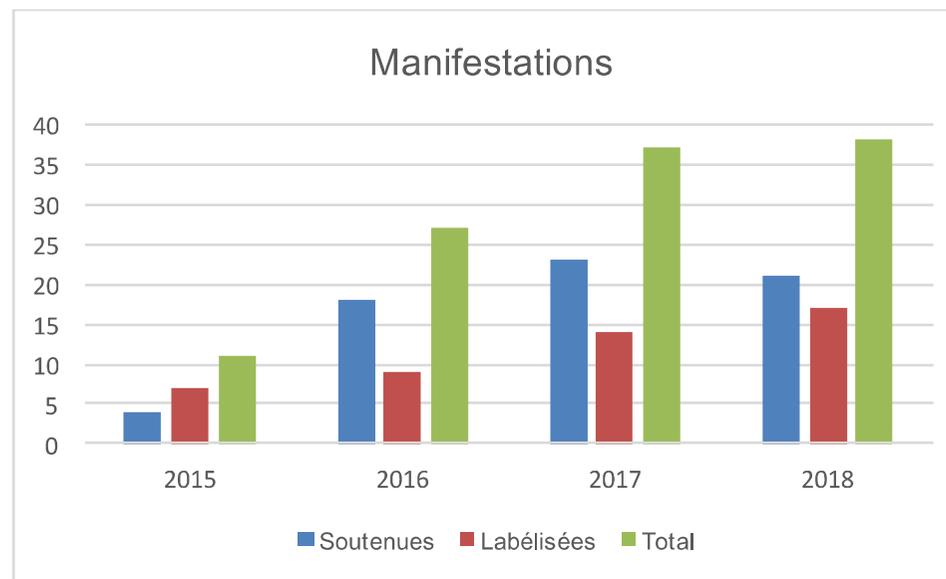
65% membres sont rattachés à des Actions

- **Vecteur d'animation scientifique**
 - montage de projets interdisciplinaires : lancement/ organisation de workshops, conférences, écoles, challenges, réponses à appels, sondages pour rapprochement de communautés, numéro spéciaux
 - émergence de nombreux *Hackathons*
 - fournisseurs de problématiques et solutions
- **Résultats** : publications, mise à disposition de jeux de données, de problèmes, logiciels et workflows de reproductibilité
- **Visibilité en croissance** : rayonnement et communication ; liens entre Actions, avec d'autres structures



Manifestation

Élément de base de l'animation dans le cadre d'une Action / hors Action



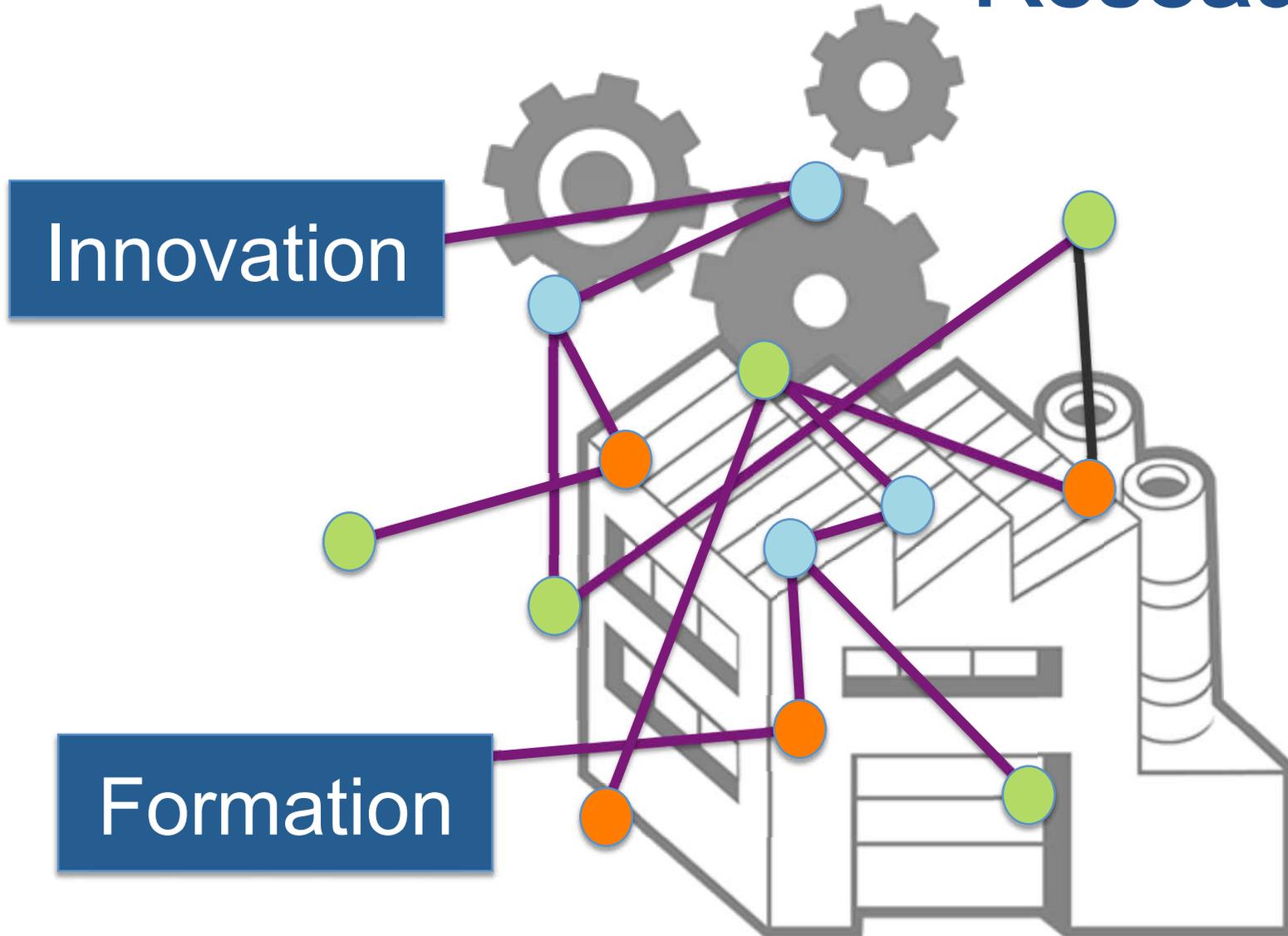
- Flux significatif de manifestations : 20/35 Manifestations soutenues/ labélisées en moyenne par an
- Modèle de soutien de manifestations visant à intégrer au maximum les communautés et à maximiser la visibilité du GDR
- Exigences en cohérence avec ce modèle (talks scientifiques, appels à interventions, pas de manifestation récurrente)

Manifestations MaDICS majeures

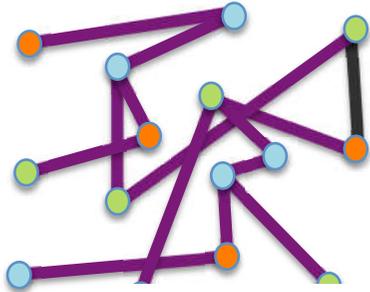
- Assemblée Constitutive, Lyon, les 24 et 25 juin 2015
- Rencontre annuelle, Paris, le 14 octobre 2016
- Journées Scientifiques MaDICS, Marseille, les 22 et 23 juin 2017 (et ateliers)
- Rencontre annuelle, Strasbourg, 20 et 21 novembre 2018 (Santé, échanges industriels, retour ateliers)

Booster les rencontres, Ouverture & Visibilité MaDICS

=> 2019 : 1^{er} SYMPOSIUM Data Science MaDICS



Réseau Formation

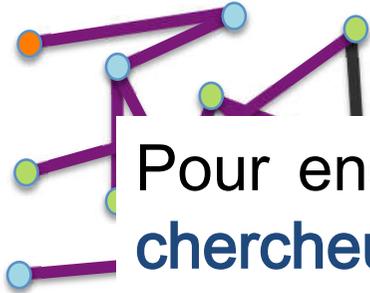


Pour soutenir les **activités de formation** autour des sciences des données et des données massives

- ressources : liste des Masters, MOOCs, etc.
- écoles : formations labellisées MaDICS reconnues par les ED, écoles thématiques CNRS, autres.

25 /10 écoles labélisées /soutenues financièrement

Espace doctorants



Pour encourager les **activités des doctorants et jeunes chercheurs** en sciences des données

-**aides à la mobilité** : bourses pour favoriser les échanges inter-disciplinaires,

-**aides pour participation à des Manifestations MaDICS** : bourses prenant en charge – partiellement – certains frais

- **posters et démonstrations** (sur appel / rencontre annuelle, manifestation spécifique)

Contrainte : être membre du GDR, appartenir à un laboratoire français

Sélection : adéquation des projets et/ou travaux aux thématiques MaDICS, le nombre de demandes par laboratoire et le budget disponible.

Réseau Innovation



Pour faciliter et intensifier les échanges entre le monde de la recherche académique et les entreprises

- offrir aux entreprises une vision de compétences académiques
- offrir aux laboratoires un accès à des problématiques industrielles
- mise à disposition de jeux de données, de problèmes
- permettre un accès à l'information, aux résultats

au moins un des responsables du réseau est un industriel

Participation *Innovatives Big Data* (octobre 2016)

Club des partenaires

- première réflexion fin 2016, printemps 2017
- mars 2017 : GDR GREEN, GOLD
- mars 2018 : Convention -> [Bulletin d'adhésion](#) 
- mars - juillet 2018 : validation des [Conditions de participation](#) par INS2i
- mise en place du Club de partenaires (en cours) : courriers de juillet et septembre

Conditions de participation au « club des partenaires » du GDR MaDICS

La SOCIETE aura librement accès :

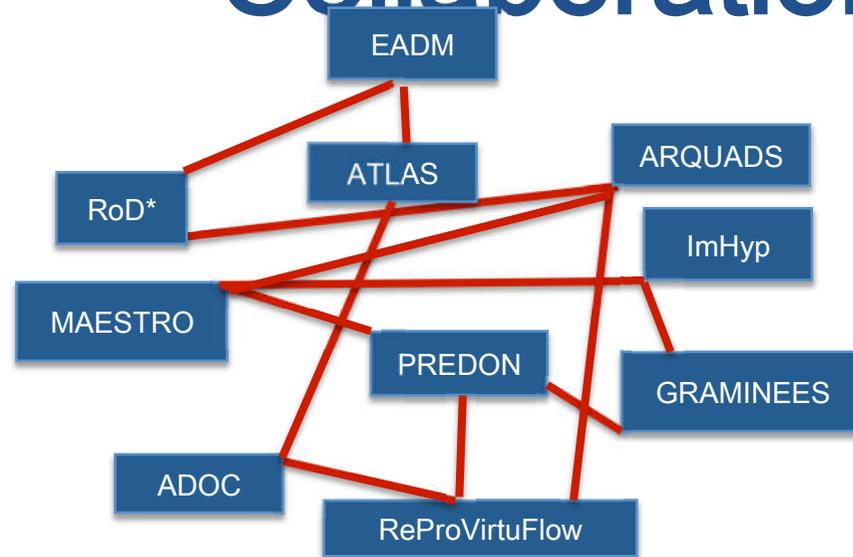
- aux lettres du GDR, couvrant les évènements passés et à venir du GDR ainsi qu'aux actualités scientifiques dans le domaine des Science des données ;
- au site internet dédié au club des partenaires sur lequel sont mis en ligne des informations et rapports des membres du GDR ;
- à un outil de recherche sur ce site permettant d'identifier des compétences ;
- aux réunions annuelles du GDR ; espace privilégié (présentations, démonstrations, flyers, etc.) ;
- aux outils du GDR pour communiquer des offres d'emploi (stage, postes CDD, CDI, des bourses de thèses, etc.)
- à des appels à **Actions Innovantes de Rupture (AIR)** dédiés à renforcer les collaborations entre jeunes chercheurs académiques et industriels

Club des partenaires

=> Restructuration site web / Plateforme intermédiation

- Pages Club
 - notes prospectives / livre blancs spécifiques
 - outil d'interrogation de compétences
 - formulaire de dépôt de problèmes
 - appel à Actions Innovantes de Rupture (AIR)
- Restructuration base de données sous-jacente
 - Adhérents (académiques / industriels)
 - Mise à jour compétences
- Espace privilégié au moment des rencontres / Symposium

*** dépôt de données ouvertes, logiciels / algo libres(liens vers plateformes)*



- Internes : entre ACTIONS

- Externes :

- GDR : manifestations communes, liens via les Actions



- Centres de recherche / observatoires / grands projets Inserm

Modalités de collaboration à ajuster, définir



- Grands instruments (TIGIR) / Platerformes (PerSCIDO)
- Associations (AFIA, SDFS,...)

- Une dizaine de communications annuelles sur MaDICS ou en rapport avec ses thématiques (sans compter celles faites par les Actions)
- Participation à l'ouvrage collectif et pluridisciplinaire « *Big Data à découvert* »,
- Numéros spéciaux, Ouvrages collectifs/revues
- Résultats Hackathon, bases de données
- NewsLetter bi-mensuelle mise en place

Manque des documents de synthèse de vulgarisation et/ou spécialisés
(utiles à la communauté et nécessaire pour le partenariat)



Site Web

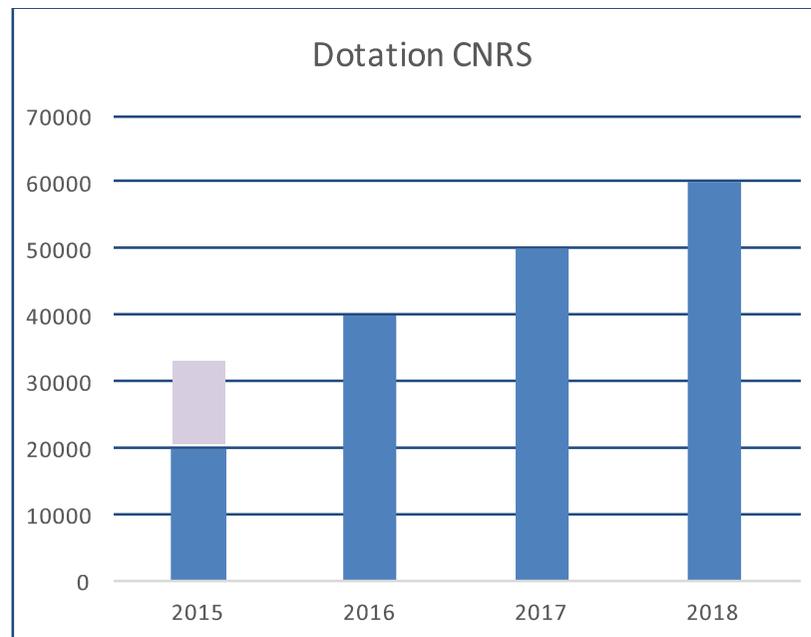
- une gestion des profils des adhérents ; un annuaire
- des formulaires de demandes de soutien, d'aides, de diffusion, d'offres d'emploi
- un outil d'aide à la gestion des manifestations pour les organisateurs
- un site web pour chaque Action et son calendrier associé
- des pages de présentation des Manifestations
- des pages de présentation du Réseau Formation et son espace dédié aux doctorants
- des pages Club des partenaires (outil d'interrogation, documents, dépôts de problématiques)

Associé à

- Listes de diffusion
- Un compte twitter



Environ +20%



Forces / Faiblesses

Masses de données,
pluridisciplinarité et
interdisciplinarité

Outils de structuration scientifique
souples (Action, Manifestation,
Formation, Innovation)

Action interdisciplinaire dynamique
(résultats), agile et bien perçue.

Visibilité en croissance (résultats,
réseau d'Actions...), label MaDICS
en construction

Fonctionnement efficace et réactif
(comdir, ara, COS, site web)

Absence de certaines disciplines ; Pluralité
des interactions avec d'autres structures ;
Peu d'activités intra discipline/domaine.

Coordination, suivi des interfaces ; Peu
implication dans la formation, les expertises
(pool d'experts), sondages ..

Risque de repli sur l'Action ; ressources
limitées

Peu de doctorants ; faible participation aux
rencontres annuelles

Communication, ouverture ; Fragilité
administrative et technique

Opportunité / Menaces

Visibilité nationale et internationale de la recherche en Sciences des données (nombreux appels)

Reconnaissance de la thématique par les entreprises (industrie 4.0)

Diffusion dans le grand public des notions fortes liées au Big Data : IA, Deep Learning,...

Données au cœur des défis sociétaux (sécurité, protection vie privée, santé, révolution industrielle, ...)

Trop de sollicitations externes pour répondre aux appels à Actions MaDICS

Nombreux outils existants (hôpitaux à projets, plateformes, chaires, associations, clubs de partenaires, ...) facilitant l'accès aux recherches académiques

“Réduction” du “Big Data” à l'IA et au Deep Learning



Merci pour votre attention