

DataNooS : bilan 2021

Nathalie Aussenac-Gilles et le comité de pilotage de DataNooS

6 avril 2022

Résumé

Ce document dresse un bilan de l'activité de la plateforme d'expertise DataNooS durant l'année 2021. Il a été rédigé par le comité de pilotage du groupe, composé des personnes suivantes (par ordre alphabétique) : Nathalie Aussenac-Gilles (IRIT et CNRS), Christophe Baehr (CNRM), Marianna Braza (IMFT et CNRS), Emmanuel Courcelle (CALMIP), Pascal Dayre (IRIT, INPT), Jessica Eynard (IDP et UT1C), Etienne Gondet (OMP et CNRS), Michelle Sibilla (IRIT, UTPS), Cassia Trojahn dos Santos (IRIT et UT2J).

1 Introduction

DataNooS a pour objectifs d'animer une réflexion académique transdisciplinaire sur les ressources numériques et les pratiques de la connaissance et de promouvoir leur partage et leur ouverture dans le cadre de la Science Ouverte. Nous entendons par ressources numériques, les données, les plateformes de capture/simulation de données, les plateformes de calcul et de stockage, les outils de traitement, d'analyse et de visualisation. DataNooS est devenu plateforme d'expertise de la MSHS-T courant 2021. Ses activités se déroulent d'une part au sein de groupes de travail qui ont pour objectif d'expérimenter le partage, la publication ou la réutilisation de données de la recherche dans des contextes de recherche interdisciplinaire, et du groupe DataCore, qui rassemble les retours d'expérience des premiers groupes pour en tirer des propositions en matière de catalogage et référencement des données à l'échelle de l'université de Toulouse ; d'autre part DataNooS est actif grâce à son comité de pilotage, qui anime la plateforme et la communauté scientifique concernée.

2 Activités des groupes de travail en 2021

2.1 UC1 : Modélisation aérodynamique enrichie et intelligente

Contact : Marianna Braza (IMFT)

<https://datanoos.univ-toulouse.fr/fr/use-case-modelisation-aerodynamique-enrichie-et-intelligente>

- **Présentation** Le UseCase UC1 développe une synergie entre les données scientifiques des trajectoires d'avion, modélisées par l'ONERA, et une partie des données du projet SMS «Smart morphing and Sensing for aeronautical configurations»¹ concernant l'optimisation de forme et du comportement vibratoire des ailes d'avion du futur pour en améliorer les performances de vol. Le financement Européen exige de porter un soin particulier à la caractérisation et la pérennité des données et des résultats produits (Plan de Gestion de Données, principes FAIR², science ouverte). Pour assurer le partage, l'interopérabilité et la réutilisation des données selon des principes FAIR, nous utilisons la plate-forme CALLISTO³ pour la science ouverte du méso-centre de calcul CALMIP, cette collaboration donnant lieu à des évolutions de CALLISTO. CALLISTO propose l'ontologie ARCAS⁴ qui permet de décrire sémantiquement les jeux de données et les workflows, et de faciliter ainsi la réutilisation des données en partie disponibles en Open Access.
- **Activités en 2021** • L'ontologie ARCAS a été spécialisée pour décrire des données scientifiques sur les nouveaux designs d'ailes d'avion bio-inspirés. • CALLISTO a été enrichie de services (mis en open access) d'accès à des algorithmes de traitement des données (workflows) et à leur re-traitement en ligne • Rédaction d'un preprint de datapaper⁵, accompagné de services facilitant la réutilisation et cross-fertilisation des données. Il souligne les bénéfices liés à l'utilisation d'une architecture FAIR

1. Européen H2020 N° 723402 www.smartwing.org/SMS/EU

2. Findable, Accessible, Interoperable and Reusable

3. <https://callisto.calmip.univ-toulouse.fr/>

4. ARCAS (ARming CALListo with Semantics) est disponible en ligne <https://callisto.calmip.univ-toulouse.fr/ARCAS.owl>

5. <https://doi.org/10.48531/JBRU.CALMIP/19XGRY>

« by design ». • Exemple de réutilisation possible grâce à DATANOOS : A partir de trajectoires d'avion de type A320 en phase d'approche à l'aéroport de Blagnac, l'ONERA-Toulouse a évalué le débit de consommation de carburant sans et avec morphing et élaboré en synergie avec le CALMIP le service permettant la cross-fertilisation entre des données du projet SMS et celles des trajectoires modélisées.

- **Contrats et collaborations** • Présentation d'un poster [BLEa21] à la 18^e plénière mondiale de la RDA • Communication orale aux JCAD 2021,[BJBT21]
- **Projet pour l'avenir** Les activités du UC1 ont contribué au montage et à la sélection d'un nouveau projet européen TEAMAERO, " Towards Effective Flow Control and Mitigation of Shock Effects in Aeronautical Applications ".

2.2 UC2 : Personnalisation de données environnementales

Contact : Nathalie Aussenac-Gilles (IRIT)

<https://datanoos.univ-toulouse.fr/fr/use-case-2-personnalisation-de-donnees-environnementales>

- **Présentation** Ce cas d'usage vise à montrer la valeur ajoutée de la description des données à l'aide de méta-données sémantiques en se plaçant dans la perspective de l'utilisateur, pour répondre à sa recherche de jeux de données dans un domaine scientifique dont il n'est pas expert, et au besoin de transfert de connaissances dans les termes utilisés par les usagers finaux. Il s'appuie sur le projet ANR Semantics4FAIR⁶, dans lequel le cas étudié est celui d'un chercheur en palynologie de l'OMP qui recherche des données météorologiques pour anticiper la propagation de l'Ambrosie, espèce invasive allergène. La problématique a évolué vers la définition d'un schéma de métadonnées qui réutilise des ontologies et vocabulaires standards, et qui optimise les critères FAIR des jeux de données.
- **Activités** L'année 2021 a permis de finaliser et mettre en ligne DMO⁷, une ontologie dédiée à la représentation de métadonnées de jeux de données du site de MétéoFrance, à une communication lors d'un atelier [AAGC+21b], à la publication de 2 articles présentés lors d'une conférence nationale [AKAG+21] et internationale [AKT+22], à un rapport interne IRIT [AKT+21] et un poster présenté lors de la 18^e plénière de la RDA [AAGC+21a].
- **Contrats et collaborations** Le projet a obtenu le soutien du chantier ENVIA du RTRA STAE afin de financer un nouveau stage courant 2022, consacré à la réalisation d'un prototype de catalogue des jeux de données du site de MétéoFrance à l'aide des métadonnées de DMO.
- **Projet pour l'avenir** Nous lançons un nouveau projet, MOCK-UP, où l'on se place du point de vue d'un utilisateur non scientifique, devant déterminer et utiliser des données environnementales pour décider du pilotage de la transition écologique d'un territoire. Une demande de co-financement de thèse est en cours auprès de UT2J (accord obtenu) et de la région Occitanie (en attente).

2.3 UC3 : plate forme de science ouverte pour une communauté disciplinaire

Contact : Pascal Dayre (INPT, IRIT)

<https://datanoos.univ-toulouse.fr/fr/use-case-plateforme-science-ouverte-pour-une-communauté-int>

- **Présentation** Ce use case émane du projet ANR "SO-DRIIHM" qui vise la spécification, la conception et le développement de la plateforme pour la science ouverte de la communauté du labEx DRIIHM. Il fédère plus de 1200 acteurs de la recherche issus de . 13 Observatoires Hommes-Milieus (OHM) L'objectif de cette plateforme est de faciliter le dépôt et le partage des données en offrant de nouvelles solutions à la communauté de recherche et à la communauté en soutien à la gestion des données. Ce projet développe la couche d'intermédiation des utilisateurs avec l'éco-système de la science ouverte. Les besoins et usages attendus pour la recherche et le dépôt de données sont recueillis en étroite collaboration avec la communauté intéressée.
- **Activités** Présentation des objectifs et activités de So-DRIIHM lors de la réunion plénière de DataNooS en juin 2021 [LD22], lors d'une journée du GDR MAGIS [ALJC+21], lors des journées MATE-SHS [LDM+21], lors de journées du labex DRIIHM [LAA+21].

2.4 DataCore

<https://datanoos.univ-toulouse.fr/fr/datacore-groupe-dexpertises-transdisciplinaires>

- **Présentation** DataCore est un groupe de travail qui capitalise les travaux interdisciplinaires des différents cas d'utilisation de DataNoos pour faciliter la coopération et les échanges de savoir-faire et d'outils. Durant l'année 2021, il s'est focalisé sur l'étude d'une méta-infrastructure pour

6. <https://www.irit.fr/semantics4fair/contratANR-19-DATA-0014-01>.

7. <https://www.irit.fr/recherches/MELODI/ontologies/DMO/core/index-en.html>

accompagner l'ouverture et le partage des données selon des principes FAIR ; faciliter le croisement de données issues de plusieurs sources, Préparer les données pour l'analyse et l'I.A.

- **Activités en 2021** • Proposition d'une méta-architecture pour coordonner les contributions pour publier, cataloguer ou réutiliser des jeux de données • Développement de la plateforme Callisto pour partager, découvrir et analyser facilement des données • Modules d'une plateforme DataNooS pour créer un référentiel ouvert de données avec gestion des schémas de méta-données, génération de formulaires pour la publication et la recherche de jeux de données de la science ouverte, guidage de l'utilisateur pour le dépôt de jeux de données. • Recrutement de 3 stagiaires pour implémenter une partie de cette architecture, et coordination avec 2 stages recrutés dans le cadre de l'UC2 et l'UC3. • Recrutement Post-doctoral d'AMina Annane fin 2020 et début 2021, suite à son travail post-doctoral au sein de l'UC2
- **Projet pour l'avenir** Ce travail est poursuivi à titre expérimental, et la réflexion associée est développée dans le cadre du CésO (Comité de réflexion sur la science ouverte de l'UFT-MIP), piloté par Michelle Sibilla, également membre du comité de pilotage de DataNooS. La mise en place de l'entrepôt recherche.data.gouv ouvre également de nouvelles pistes d'expérimentation et de collaboration.

3 Activités du comité de pilotage

Outre la coordination des échanges avec les groupes de travail, le comité de Pilotage a mené en 2021 deux autres types d'activités : (i) se structurer et réfléchir à son statut une fois terminé le chantier du RTRA STAE ; (ii) assurer une animation scientifique et des échanges avec la communauté scientifique toulousaine et au delà.

3.1 Structuration et visibilité de DataNooS

Pour mener cette réflexion, le comité de Pilotage a bénéficié de la présence en son sein d'une spécialiste du droit des données personnelles, Jessica Eynard, et du recrutement d'un stagiaire de M2 en droit de l'informatique, Mike Souza Barbeiro. Le Copil a réalisé quatre tâches :

- **Choix d'un statut** : Suite à sa prolongation (due à la pandémie de COVID-19), DataNooS est resté un chantier du RTRA STAE jusqu'à début 2021. Afin d'afficher son interdisciplinarité, son lien avec l'université de Toulouse et le CNRS, DataNooS a mené une réflexion pour déterminer le statut le mieux adapté à ses objectifs et à son fonctionnement. Le comité de pilotage a choisi de devenir un groupe d'expertise de la MSHS-T, statut que lui a accordé le comité de pilotage de la MSHS-T.
- **Définition d'un règlement intérieur** : Afin de clarifier les missions des groupes de travail, les relations entre le copil et les cas d'usage, le statut des jeux de données manipulés dans ces projets et ou encore ce que signifie de participer à DataNooS, nous avons rédigé un règlement intérieur sur la base de la proposition de nos collègues juristes. Il a ensuite inspiré le règlement intérieur proposé par la MSHS-T pour toutes ses plateformes d'expertise.
- **Réalisation du site web** <http://datanoos.univ-toulouse.fr> : Fin 2020, nous avons eu recours aux services d'un prestataire extérieur pour la réalisation du site web de Datanoos mis en service début 2021. Ce site joue le rôle de "vitrine" pour Datanoos et ses "use cases", il annonce des événements en lien avec le partage FAIR des données, dont ceux organisés par DataNooS pour lesquels il comporte une fonctionnalité d'inscription. Ce site utilise le CMS Drupal (version 8 puis 9 début 2022) et il est hébergé sur une machine virtuelle de l'Université de Toulouse, administrée par Emmanuel Courcelle.
- **Relations avec le CésO, ENVIA, l'IRD et Occitanie Data** Les membres du COPil de DataNooS sont très impliqués dans l'animation du CésO (M. Sibilla en est la responsable) et de ses groupes de travail (GT état des lieux, feuille de route, contribution à l'élaboration d'une enquête auprès de la communauté scientifique ...). Des membres de DataNooS participent à la mission Science Ouverte de l'IRD. D'autres sont actifs au sein du groupe "données" du chantier ENVIA. Des échanges réguliers ont eu lieu avec Occitanie Data puis avec Hervé Luga au sujet du futur centre de données régional DR'Occ.

3.2 Animation scientifique

DataNooS a organisé une réunion plénière pour inviter la communauté toulousaine, et présenté ses activités lors de manifestations scientifiques.

- Contributions à une sensibilisation aux DMP et aux data (02/02/2021), organisée par Etienne Gondet pour l'OMP⁸.
- Présentation du CésO et de DataNooS lors de l'OPen Access week (mars 2021)
- Organisation d'une réunion plénière (en visio) le 15 juin 2021⁹, envisagée comme la première journée d'un cycle d'ateliers "Passer de la donnée brute aux croisements de données entre disciplines". Présentation des use case et intervention de M. Hoarau sur le projet de référentiel de site de Strasbourg
- présentation de nos travaux aux JCAD 2021 (dec 2021)
- présentation de 3 posters à la 18e plénière de la RDA¹⁰ (Nov. 2021) [AAGC⁺21a, BLea21, MS21]

4 Projet pour 2022-2023

Pour l'année 2022, le comité de pilotage cible les objectifs suivants :

- accueillir de nouveaux cas d'usage : plusieurs contacts sont en cours, avec le projet NéoCampus, la maison intelligente de Blagnac ou encore des collègues portant des projets faisant appel à des jeux de données
- poursuivre le cycle des ateliers : le prochain, prévu pour le 12 avril 2022, sera consacré aux pratiques de dépôt et réutilisation de jeux de données
- poursuivre les développements réalisés au sein du dataCore et des cas d'usage pour parvenir à un environnement de catalogage sémantique.
- poursuivre les échanges avec le CésO pour apporter une contribution adaptée et pertinente.

Références

- [AAGC⁺21a] A. Annane, N. Aussenac-Gilles, C. Comparot, M. Kamel, C. Trojahn, C. Baehr, P. Gaillard, and Y. Auda. Semantics4fair, ontology-based approach for fair datasets https://drive.google.com/file/d/11S4xJBhE1m0Wed8ftcE82UNMT_JwF7m1/preview. In *RDA 18th virtual - Poster Exhibition - Nov., 3rd-21st, 2021*, 2021.
- [AAGC⁺21b] A. Annane, N. Aussenac-Gilles, C.e Comparot, M. Kamel, C. Trojahn dos Santos, and C. Baehr. Un modèle sémantique en vue d'améliorer la FAIRisation des données météorologiques. In *Séminaire In-OVIVE 2021 - INRAE*, Paris (en distanciel), France, September 2021.
- [AKAG⁺21] A. Annane, M. Kamel, N. Aussenac-Gilles, C. Trojahn, C. Comparot, and C. Baehr. Un modèle sémantique en vue d'améliorer la FAIRisation des données météorologiques. In M. Lefrançois, editor, *Journées Francophones d'Ingénierie des Connaissances (IC 2021) @ PFIA 2021*, pages 20–29, Bordeaux (en distanciel), France, June 2021.
- [AKT⁺21] Amina Annane, Mouna Kamel, Cassia Trojahn, Nathalie Aussenac-Gilles, Catherine Comparot, and Christophe Baehr. SYNOP Data Evaluation Using FAIR Maturity Model. Research Report IRIIT/RR-2021-03-FR, IRIIT - Institut de Recherche en Informatique de Toulouse, April 2021.
- [AKT⁺22] A. Annane, M. Kamel, C. Trojahn, N. Aussenac-Gilles, C. Comparot, and C. Baehr. Towards the fairification of meteorological data : a meteorological semantic model. In E. Garoufallou and M. A. Ovalle-Perandones, editors, *Metadata and Semantic Research - 14th International Conference MTSR 2021, Madrid, Spain, Revised Selected Papers*, volume 1537 of *CCIS*, pages 1–12. Springer, 2022.
- [ALJC⁺21] F. Arnaud, E. Lerigoleur, Arnaud J.-C., I. I. Le Berre, C. Pardo, J.-C. Raynal, J. Fozzani, K.ristell Michel, M.-L. Trémélo, and D. Roux-Michollet. Gérer et partager des données multidisciplinaires dans des Observatoires Hommes-Milieus : retours d'expérience du LabEx DRIIHM. In *Atelier du GdR MAGIS " Réflexions et actions collaboratives autour des observatoires scientifiques Milieux-Sociétés "*, Montpellier, France, December 2021.
- [BJBT21] M. Braza and Th. Louge et al. J.-B. Tô. Etude de performances aérodynamiques d'ailes d'a320 par déformation électro-active; bénéfices d'une approche centrée sur les données. In *Journées Calcul et Données (JCAD 2021), Toulouse (F), 14/12/21, 2021*.

8. https://sourcesup.renater.fr/wiki/atelieromp/dmp_dataper

9. <https://datanoos.univ-toulouse.fr/datanoos2021>

10. <https://datanoos.univ-toulouse.fr/fr/3-posters-datanoos-la-18e-pleni%C3%A9re-virtuelle-de-la-rda>

- [BLEa21] M. Braza, Th. Louge, and J.-B. Tô et al. Fairified aerodynamic data cross-fertilization <https://youtu.be/ONqOqKcihr0>. In *RDA 18th virtual - Poster Exhibition - Nov., 3rd-21st, 2021*, 2021.
- [LAA⁺21] E. Lerigoleur, G. Amal, F. Arnaud, P. Dayre, A. Decaulne, P. Duboz, S. Dunesme, J.-C. Arnaud, I. Le Berre, M. Massaviol, K. Michel, B. Morandi, C. Pardo, J.-C. Raynal, G. Skupinski, and M.-L. Trémélo. Vision and first actions of the SO-DRIIHM open science project. In *International Symposium of LabEx DRIIHM (2021)*, Toulouse, France, September 2021.
- [LD22] E. Lerigoleur and P. Dayre. Le projet ANR SO-DRIIHM : Aiguiller et aider une communauté interdisciplinaire à déposer ses données dans une jungle de référentiels disponibles. In *Plénière de l'alliance académique transdisciplinaire dataNooS (DATANOOS 2021)*, IRIT, March 2022.
- [LDM⁺21] E. Lerigoleur, P. Dayre, M. Massaviol, K. Michel, C. Pardo, J.-C. Raynal, and M.-L. Trémélo. Co-design d'un portail "aiguilleur de la science ouverte" DATA-DRIIHM pour le partage des données interdisciplinaires des Observatoires Hommes-Milieus. In *Journées Annuelles de Méthodes, Analyses, Terrains, Enquêtes en Sciences Humaines et Sociales (MATE-SHS 2021)*, Bordeaux (Webinaire), France, September 2021. MATE-SHS : Méthodes, Analyses, Terrains, Enquêtes en Sciences Humaines et Sociales.
- [MS21] P. Dayre M. Sibilla, N. Aussenac-Gilles. datanoos : Toward a generic and integrated fair data management platform <https://youtu.be/ONqOqKcihr0>. In *RDA 18th virtual - Poster Exhibition - Nov., 3rd-21st, 2021*, 2021.