

Réponse de l'UGA à l'appel d'offre national « Ateliers de la donnée »

F. Malbet

Jeudi 3 mars 2022

Stratégie Recherche Data Gouv (RDG)

Focus sur Recherche.Data.Gouv



Une plateforme à cinq modules

RECHERCHE DATA GOUV



- **Modules « entrepôt - catalogue »**
 - Projet confié par le MESRI à l'INRAE avec la participation d'établissements volontaires dont l'UGA
- **Module « ateliers de la donnée »**
 - AMI en cours. Objectif de labelliser des initiatives d'accompagnement des communautés recherche sur l'ensemble du cycle de vie des données

DATA TERRA : une infrastructure pour le système Terre

DATA TERRA est une infrastructure de recherche dont le but est de mettre à disposition des données atmosphériques, climatiques, océanographiques, sismiques et spatiales, pour comprendre et prévoir de manière intégrée l'histoire, le fonctionnement et l'évolution du système Terre.

Calendrier RDG



- Première version de l'entrepôt en **mars 2022**
- Première vague de labellisation des ateliers de la donnée en **avril - mai 2022**
- Le dispositif d'accompagnement va aussi s'appuyer sur :
 - Des **centres de ressources thématiques**, spécialisés sur des aspects disciplinaires
 - Des **centres de ressources rattachés à Recherche Data Gouv** (techniques, mutualisation de formations, ...)



UGA et Recherche.Data.Gouv ?

Rappels sur la Cellule Data Grenoble Alpes



- **Structure transverse** regroupant 16 membres issus de :
 - GRICAD
 - BAPSO
 - MSH
 - Laboratoires
 - Délégué à la Protection des Données
- Une **implication forte des membres GRICAD et BAPSO** et plus ponctuelle des autres participants
- Des personnels à la fois **UGA et CNRS**
- Des compétences complémentaires : **informatique, documentation, juridique, science ouverte**
- Propose un **ensemble de services autour de la donnée** en s'appuyant sur les infrastructures du site.
- S'appuie sur un **réseau de référents données dans les laboratoires**

Rappel de l'AMI « ateliers de la donnée »



- Objectif : labelliser les initiatives existantes ou nouvelles sur **l'accompagnement des communautés scientifiques tout au long du cycle de vie des données.**
- En **proximité**, regroupant des **compétences complémentaires**
- Différents niveaux de labellisation selon la maturité et le niveau de service apporté
- Calendrier :
 - Appel à intentions : clôture le **20 décembre 2021**
 - **6 janvier 2022** : présentation des propositions
 - Appel à projet du 10 janvier au **23 février 2022**
 - Premières labellisations fin avril 2022



#1 Guichet unique

- *Une adresse mail unique est disponible pour toute question sur les données (sos-data@univ-grenoble-alpes.fr).*
- *Cette adresse est adossée à un système de ticket et les demandes sont réparties entre les membres de la CDGA les plus pertinents. Le système permet d'assurer le suivi, mais aussi de fournir des indicateurs sur l'impact de la CDGA.*
- *Cette adresse est affichée sur le site science ouverte de l'UGA, ainsi que sur le site de GRICAD. Elle est également largement communiquée lors des formations, séminaires et workshop organisés, ainsi que lors des rencontres avec les membres des laboratoires.*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis moins d'un an



#2 Site web Science Ouverte

- *Le site science ouverte de l'université Grenoble Alpes et de ses partenaires (scienceouverte.univ-grenoble-alpes.fr) regroupe toutes les informations nécessaires pour la bonne orientation des équipes de recherche . Il permet un premier niveau d'information des communautés vers les ressources et le guichet unique.*
- *Le site couvre les problématiques des publications et des données de la recherche. Dans les prochaines semaines, des éléments concernant les codes de recherche seront également intégrés. Il recense également toutes les actualités et évènements pertinents sur la science ouverte en général, y compris les actions et formations proposées par la CDGA.*
- *Il est animé par la CDGA pour la partie données et codes et la Cellule HAL de l'UGA pour la partie publications.*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis moins d'un an

#3 Planification : aide à la rédaction de Plans de Gestion de Données

- *Ce service vise à répondre à toutes les demandes sur les plans de gestion des données, ainsi que sur les documents requis dans le cas de données sensibles (RGPD / AIPD).*
- *La cellule apporte des conseils, une aide à la rédaction et à la relecture. Elle rédige des documents à intégrer aux Plans de Gestion de Données : par ex dans le cas de l'utilisation des infrastructures du site.*
- *Elle apporte les compléments techniques, juridiques et documentaires nécessaires. Elle accompagne à l'utilisation de DMP Opidor.*
- *A court terme, la CDGA va intégrer des recommandations spécifiques au site dans DMP Opidor. Elle travaille également à une bonne appropriation des outils maDMP.*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis un an ou plus

#4 Planification : aide juridique sur les données et les codes

- *Ce service vise à accompagner les communautés scientifiques sur les questions les plus courantes concernant le RGPD et la propriété intellectuelle. Dans les cas plus complexes, le DPD, membre de la cellule, prend le relais.*
- *La cellule apporte des éclairages sur les licences données et logiciels. Elle fait l'interface avec les services de valorisation sur ces questions et sensibilise les communautés selon leurs spécificités. A court terme, l'objectif est de formaliser le lien avec les services valorisation et faire monter globalement en compétences sur le RGPD et les licences. L'organisation d'une base de connaissances, la construction d'une FAQ, la rédaction de fiches pratiques sont les prochaines étapes.*
- *Il faut également poursuivre le travail d'identification des interlocuteurs au niveau national, experts sur certains points très spécifiques, que la cellule peut solliciter si nécessaire.*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis un an ou plus

#5 Planification : audit de laboratoires, démarche qualité données

- *Ce service a pour objectif d'aider les laboratoires qui souhaitent déployer une politique de gestion de la donnée et des codes en interne. La cellule accompagne sur la réalisation d'un audit de l'existant et fait des recommandations d'organisation, d'utilisation de ressources, de bonnes pratiques comme l'application des principes FAIR.*
- *Le but est également d'inciter à l'intégration d'une démarche qualité reprenant ces principes à la fois autour des données mais aussi dans le cadre des plateformes déployées dans les laboratoires ou sur le site.*
- *A moyen terme, l'ambition est de généraliser la démarche sur le site en incitant les laboratoires à se poser la question de la gestion des données à leur échelle. Afin de renforcer la proximité et la compréhension des problématiques spécifiques, des expérimentations d'immersion de membres de la cellule dans les laboratoires pourront être menées.*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis moins d'un an

#6 Acquisition : conseil et accompagnement pour la recherche de données

- *L'objectif est de construire un service d'accompagnement des communautés scientifiques à la recherche de données en suivant plusieurs axes :*
 - *développer les connaissances des bases de données et des entrepôts disponibles en fonction des disciplines (développer une expertise),*
 - *faire l'interface avec différentes Infrastructures de Recherche au niveau national,*
 - *identifier des contacts dans les collectivités territoriales.*
- *Il s'agira également de s'approprier les outils comme data.gouv et éventuellement, créer un annuaire de sources en cohérence avec les centres de ressources thématiques de RDG et valorisables sur le site web SO et dans les formations.*
- *Le réseau des référents données du site, ainsi que les collègues des autres ateliers, seront des appuis importants pour identifier les sources les plus pertinentes.*

Niveau actuel d'accompagnement du service ou de l'action : En projet

#7 Acquisition : conseils et accompagnement sur le stockage

- *Ce service répond aux nombreuses demandes concernant la problématique stockage qui sont extrêmement variées :*
 - *orientation vers la plateforme la plus adéquate,*
 - *aide au partage des données entre partenaires d'un projet,*
 - *aide à l'organisation des données,*
 - *aide à la sécurisation des données (dont chiffrement, cloisonnement ...), - prise en compte des spécificités disciplinaires.*
- *La CDGA accueille et encadre quelques CDD data scientists des projets scientifiques. Elle rédige des documentations et des recommandations. Pour certains projets impliquant la cellule comme partenaire, elle met en œuvre les architectures techniques nécessaires.*
- *La CDGA s'appuie pour ce service sur les infrastructures du site qui seront hébergées à terme dans le DC labellisé GR3@, et inscrites au catalogue EOSC. Elle peut aussi orienter les utilisateurs vers d'autres plateformes selon les besoins (services de stockage des organismes de recherche, TGIR HumaNum, CINES, ...).*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis un an ou plus

#8 Acquisition : aide sur les cahiers de laboratoire électroniques

- *Une plateforme de cahiers de laboratoire électroniques a récemment été mise en œuvre sur le site grâce à une démarche mutualisée de certains laboratoires et de GRICAD.*
- *En cohérence avec l'expertise menée au sein du CoSO, la CDGA va proposer des formations et un accompagnement à son utilisation.*
- *Un des objectifs sera également de communiquer largement sur l'existence de ce nouveau service, et de s'assurer qu'il répond aux besoins d'une majorité de communautés scientifiques.*
- *A plus long terme, l'ambition serait de créer des liens avec les autres plateformes : calcul, stockage, gitlab en particulier, et d'accompagner les scientifiques à une utilisation multimodale de ces infrastructures.*

Niveau actuel d'accompagnement du service ou de l'action : En cours de création

#9 Transformation : aide pour l'analyse de données, le calcul et l'IA

- *Ce service apporte un support aux utilisateurs des plateformes de calcul, de traitements de données et d'IA présentes sur le site.*
- *La cellule peut aider à l'installation et à l'utilisation des logiciels. Elle propose également des conseils et des formations.*
- *Elle peut réaliser des développements spécifiques dans le cadre de partenariat sur des projets scientifiques.*
- *La CDGA s'appuie pour ce service sur les infrastructures du site qui seront hébergées à terme dans le DC labellisé GR3@, et inscrites au catalogue EOSC. Elle peut aussi orienter les utilisateurs vers d'autres plateformes selon les besoins en particulier GENCI.*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis un an ou plus

#10 Transformation : conseil et aide sur le développement de codes

- *Ce service vise à accompagner les communautés scientifiques à l'utilisation d'outils pour le développement collaboratif : gitlab, intégration continue sur le cloud, ...*
- *La cellule met également en place des formations autour du génie logiciel. Elle aide et conseille sur les problématiques d'optimisation des codes.*
- *Elle soutient les démarches de reproductibilité à travers un travail important sur le packaging.*
- *L'objectif à moyen terme est d'inscrire les actions autour de la reproductibilité au niveau européen et augmenter la communauté autour du packaging Nix et Guix (utilisés sur le site) au niveau national.*
- *Une des actions pourra être de travailler à la mise en place d'un espace (channel) au niveau national pour le partage des paquets.*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis un an ou plus

#10 Transformation : conseil et aide sur le développement de codes

- *Ce service vise à accompagner les communautés scientifiques à l'utilisation d'outils pour le développement collaboratif : gitlab, intégration continue sur le cloud, ...*
- *La cellule met également en place des formations autour du génie logiciel. Elle aide et conseille sur les problématiques d'optimisation des codes.*
- *Elle soutient les démarches de reproductibilité à travers un travail important sur le packaging.*
- *L'objectif à moyen terme est d'inscrire les actions autour de la reproductibilité au niveau européen et augmenter la communauté autour du packaging Nix et Guix (utilisés sur le site) au niveau national.*
- *Une des actions pourra être de travailler à la mise en place d'un espace (channel) au niveau national pour le partage des paquets.*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis un an ou plus

#11 Diffusion : conseil et accompagnement pour la diffusion des données

- *L'objectif est de répondre aux demandes pour la diffusion en open data ou restreinte des données.*
- *La CDGA oriente les scientifiques en privilégiant les entrepôts disciplinaires certifiés, et prochainement vers les centres de ressources thématiques et vers la collection UGA de Recherche Data Gouv (RDG).*
- *La cellule conseille aussi sur les formats des données, leur organisation, leur documentation, ainsi que les licences les plus appropriées. Il s'agit là encore de développer l'application des principes FAIR pour favoriser la réutilisation des données diffusées.*
- *Elle sensibilise aussi à la publication de data papers. Ce service va fortement monter en puissance avec l'ouverture de RDG. L'organisation de l'accompagnement à l'échelle du site est en cours en impliquant au maximum les personnels de proximité dans les laboratoires et les référents données.*
- *Il faut développer collectivement les compétences pour assurer une curation et une modération de qualité des jeux de données déposés.*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis un an ou plus

#12 Diffusion : conseil et accompagnement pour la diffusion des codes

- *Actuellement, la seule diffusion se fait via l'instance gitlab présente sur le site.*
- *Il s'agit cette année de démarrer un travail d'accompagnement sur la diffusion et l'archivage des codes de recherche en s'appuyant sur HAL et Software Heritage. Au printemps va être commencé un premier référencement des développements logiciels des laboratoires.*
- *De nouvelles plateformes (en particulier un Jupyter/binder) va permettre d'intégrer les problématiques de notebooks et faire le lien avec gitlab.*
- *Il s'agit en parallèle de proposer des formations et de rédiger de la documentation pour faciliter la diffusion des codes de recherche, leur citation. Il sera fondamental d'assurer par ce service le lien entre les publications, les données et les codes.*

Niveau actuel d'accompagnement du service ou de l'action : En cours de création

#13 Diffusion : accompagnement sur l'archivage numérique

- *La question de l'archivage numérique des données revient très régulièrement dans les demandes des communautés. Il n'existe pas sur le site de plateforme d'archivage au sens strict du terme (cad préservation sur le long terme et dans toute leur intégrité des documents numériques).*
- *L'objectif est de pouvoir conseiller les communautés scientifiques sur leurs besoins en archivage : identifier précisément les données à archiver, les durées, les contraintes réglementaires, ...*
- *Il s'agit donc d'abord de développer les compétences de la CDGA sur l'archivage numérique en associant des archivistes professionnels.*
- *Ce service pourra prendre en compte l'ensemble du processus de recherche et s'inscrit en cohérence avec le service transverse de sensibilisation aux impacts environnementaux. Il sera également important d'identifier les expertises et créer un réseau de contacts sur le sujet au niveau national grâce en particulier au réseau des ateliers.*

Niveau actuel d'accompagnement du service ou de l'action : En projet

#14 Diffusion : liens publications, données et codes. Ed. universitaires

- *L'objectif est de faciliter le lien entre les publications, les données et les codes, en particulier à travers un travail commun avec les responsables de revues UGA et de UGA Édition, le centre Mersenne, ainsi que le réseau Medici et le collège publications.*
- *Il s'agit également d'encourager la collaboration entre les équipes d'accompagnement autour de HAL et autour des données, à travers la mise en place d'espaces de travail partagés, la co-organisation de formations et d'ateliers, la généralisation de la coopération entamée pour les porteurs de projet.*
- *Enfin, il faut aller vers l'usage de référentiels partagés (ORCID, ROR, idRef, etc.) en cohérence avec la politique nationale et conformément aux principes FAIR.*

Niveau actuel d'accompagnement du service ou de l'action : En projet

#15 Transverse : formations et animations

- *Les formations proposées par la CDGA sont organisées pour différents niveaux (collège doctoral, équipes de recherche, personnels d'appui) sur de nombreux sujets (génie logiciels, calcul intensif, IA, stockage, diffusion des données, PGD ...).*
- *La CDGA intervient également dans des formations disciplinaires et au niveau national (par exemple avec l'INIST). Elle organise des séminaires et ateliers et anime une liste de diffusion. Elle assure la création de différents types de ressources (fiches, vidéos ...) et vise à diversifier ses modalités pédagogiques en s'associant avec la MACI (Maison de la création et de l'innovation). Elle s'appuie fortement sur son réseau de référents laboratoires qu'il faut renforcer et former.*
- *Le rôle du réseau des ateliers est également primordial sur ces sujets puisqu'il va permettre de mutualiser les supports, d'organiser des formations communes, d'échanger autour d'usages et de bonnes pratiques.*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis un an ou plus

#16 Transverse : sensibilisation aux impacts environnementaux

- *Le coût environnemental du numérique, et des données en particulier, est important. La CDGA travaille sur la mise en place d'indicateurs pertinents pour quantifier l'impact de l'usage des plateformes de stockage et de calcul.*
- *Elle sensibilise et informe les utilisateurs, et intervient à différents niveaux (local, national, international).*
- *Certains des indicateurs ont été bien identifiés et vont être intégrés dans les plateformes afin d'impliquer les utilisateurs.*
- *Outre la quantification des impacts, il s'agit maintenant d'aborder également les évolutions dans les usages par des réflexions qui doivent se tenir au niveau local, national (par exemple dans le cadre du réseau MésoNet) et international (via la RDA).*

Niveau actuel d'accompagnement du service ou de l'action : Effectif depuis un an ou plus

#17 Transverse : construire un réseau métiers des ateliers

- *L'objectif est d'initier, à travers des discussions déjà existantes, un réseau métier d'échanges de bonnes pratiques, afin de créer des synergies entre dispositifs d'accompagnement et mutualiser des réalisations.*
- *Plusieurs actions pourront être mise en œuvre : organisation d'événements, espaces d'échanges communs, rencontres annuelles, échanges de type Erasmus, journées d'étude,...*
- *Un des objectifs est de favoriser la montée en compétence des acteurs de l'accompagnement (membres des ateliers de la donnée UGA, URCA, UL, Alsace et UBFC, ayant vocation à s'ouvrir largement) au travers de séminaires communs ou de dispositifs intégrés à la formation continue.*
- *Il s'agit également de proposer des outils techniques communs aux ateliers, comme un assistant numérique virtuel (chatbot) sur les données reposant sur une base de connaissances alimentée par le réseau, un annuaire de personnes ressources,...*
- *Ces actions sont transverses et se retrouvent dans plusieurs des services de la CDGA.*

Niveau actuel d'accompagnement du service ou de l'action : En cours de création

Périmètre retenu et moyens demandés

Partenariats avec les établissements présents sur le site



- Périmètre : l'**ensemble des laboratoires du site**, quelque soit leur discipline et leurs tutelles
- Etablissements présents sur le site et sollicités pour être **partenaire** de la CDGA :
 - CNRS
 - INRIA
 - INRAE
 - CEA
 - CHUGA
 - INSERM
 - IRD
- **Différents niveaux d'implication** possibles
 - Implication de personnels dans la CDGA
 - Apport de ressources financières récurrentes ou spécifiques sur sollicitation
 - Co-organisation d'évènements, de formations ...
- **Coordination des actions locales et nationales**

Demande financière proposée



- **Accompagner la diffusion des données et des codes.**
La présence d'une personne de plus au côté des 16 membres de la cellule permettra d'impulser le mouvement dans un calendrier cohérent avec l'ouverture de l'entrepôt de la plateforme fédérée des données Recherche Data Gouv et avec le développement des liens entre Software Heritage et HAL. Un premier travail de référencement des codes développés sur le site devra être réalisé.
 - > Demande : financement de 2 ans de CDD niveau IE de formation plutôt technique avec une connaissance de l'open data.
 - > Total : 120 k€
- **Développer l'offre de formation et l'animation** : proposer des séminaires, ateliers, conférences.
 - > Demande : financement de missions (environ 4 par an pendant 2 ans, environ 500 euros par mission), financement de pauses et buffets (environ 4 par an pendant 2 ans, 50 personnes par évènement, 30 € / personne)
 - > Total : 16 k€
- **Permettre aux membres de la CDGA de se former** et de participer à des initiatives nationales ou européennes (environ 6 missions par an pendant 2 ans, environ 500 euros par personne)
 - > Total : 6 k€



Quelle place pour les activités OSUG-DC ?