

Réunion du comité opérationnel de l'Atelier de la donnée Centre-Val de Loire

23/05/2024

Ordre du jour

- 01 Informations générales concernant l'ADCVL**
- 02 Les groupes de travail de l'ADCVL**
- 03 L'évolutions de l'ADCVL**
- 04 L'écosystème Recherche Data Gouv**
- 05 Questions diverses**

Les informations générales de l'ADCVL

01

**Site et guichet unique :
utilisation et évolution**

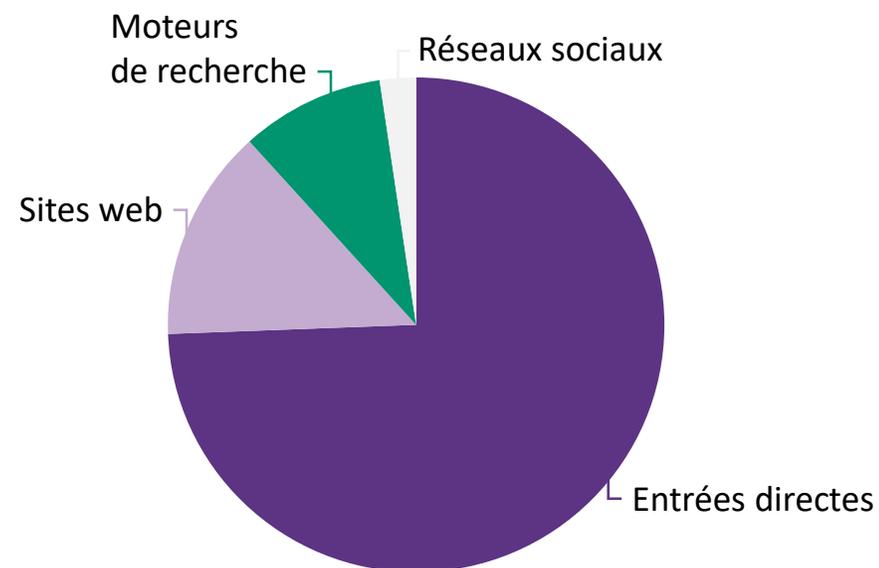


Les statistiques du guichet unique

- **14 demandes officielles** depuis l'ouverture du guichet unique
- Des demandes qui concernent principalement **l'accompagnement aux dépôts de jeux de données** et les **plans de gestion de données**
- Des demandes en provenance de chercheurs d'institutions diverses
- Deux vagues de demandes après l'ouverture du guichet et l'inauguration

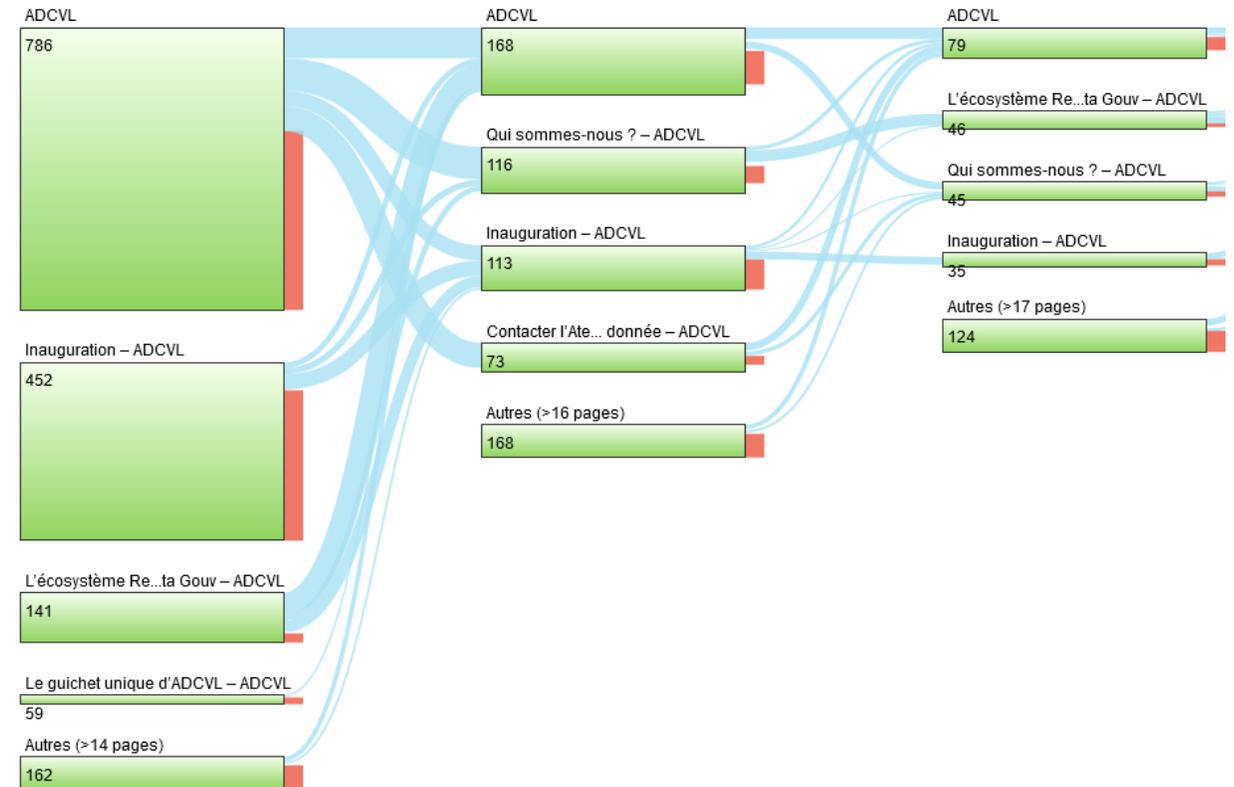
Les statistiques de visites du site

- **1614 visites** depuis le lancement du site et autour de 800 visiteurs uniques
- Une majorité de visiteurs (75%) se rendent sur le site en tapant directement l'adresse
- L'effet des réseaux sociaux paraît faible mais l'activité est très récente
- Les sites redirigeant le plus vers l'ADCVL sont très disparates, les boîtes mails (en cumulés) en majorités et les site de plusieurs institutions sont les principaux



Les statistiques de consultation du site

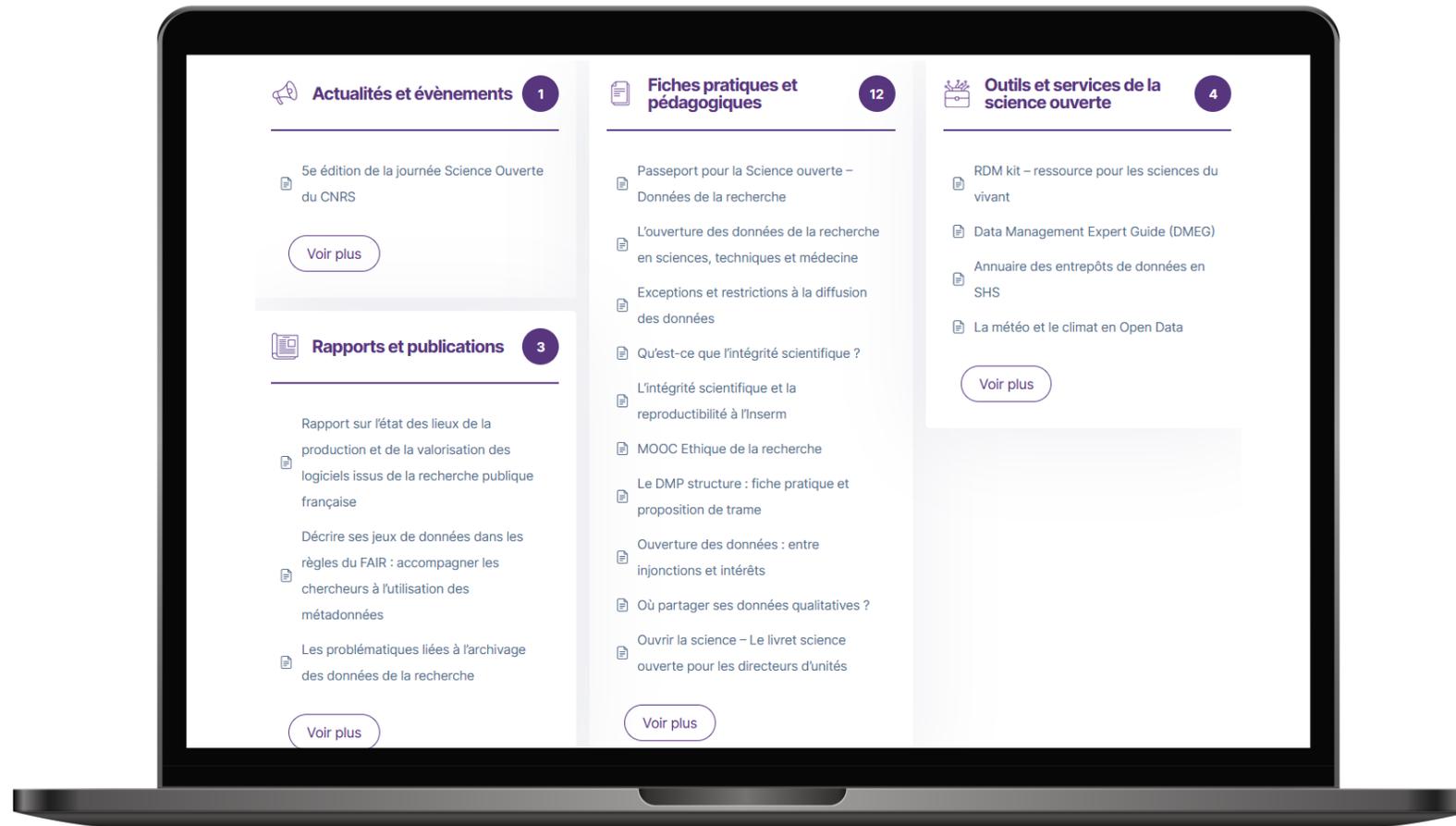
- Temps moyen de consultation du site : **3 minutes**
- Les pages les plus consultées :
 - La page d'accueil
 - La page dédiée à l'inauguration
 - La page sur le réseau Recherche Data Gouv
 - La page « Qui sommes-nous ? »
 - La page de contact du guichet unique
 - La page de la boîte à outil
- Parcours sur le site assez court :
 - En moyenne 2 actions par visite (clic, téléchargement, retour en arrière ...)
 - Les parcours se consistent bien souvent à un accès par la page d'accueil et la consultation d'une page informative avant la fin de session



Données issues du Matomo

La page documentation

- Espace pour partager des actualités et des ressources autour de la science ouverte
- Espace informel ouvert aux seuls membres de l'ADCVL
- Peut porter des informations et des ressources en provenance des institutions partenaires de l'atelier



Fiches pratiques



Des productions de ressources destinées aux chercheurs à l'échelle de l'ADCVL

Conseiller pour le choix de :

De l'entrepôt de données



Entrepôt de données NAKALA

Sciences humaines et sociales

Type de service : Entrepôts de données
URL : <https://www.nakala.fr/>
Contact : nakala@huma-num.fr
Institution porteuse : INRAE
Structure d'appartenance : Huma-Num
Tutelles : CNRS, Campus Condorcet, Aix-Marseille Université

Communauté d'utilisateurs

Communautés de chercheurs et d'institutions travaillant dans le domaine des sciences humaines et sociales.

Type de données

Données de recherche issues de projets scientifiques, comprenant des données de type textuelles, numériques, temporelles et spatiales...

Fonctionnalité

Les données sont stockées sur des serveurs gérés par Huma-Num. Ses services sont gratuits et offrent les fonctionnalités suivantes :

- Attribution d'identifiants pérennes citables dans vos publications (DOI).
- Utilisation du Dublin Core, un schéma généraliste et standard de description des données.
- Accès illimité dans le temps et en volume à vos données.
- Gestion fine des droits d'accès, de partage et de publication des données.
- Moteur de recherche qui exploite et indexe les métadonnées ainsi que le contenu des fichiers textuels (fichiers markdown, CSV, Images OCRisées, PDF, etc.) pour une recherche élargie.
- Mise à disposition des métadonnées de manière interopérable par des API et le protocole OAI-PMH qui permettent leur exploitation et/ou leur référencement par des agrégateurs externes (ex ISIDORE).

Conditions d'usage

Les jeux de données sont diffusés sous **Licence Ouverte**. Il n'y a pas de limitation pour la taille des données déposées. Cependant, si vous envisagez de déposer de gros volumes (par exemple, supérieurs à 10 GO par fichier), des précautions doivent être prises.

Cinq métadonnées sont obligatoires, mais il est fortement recommandé d'en utiliser davantage.

Standards et directives

- Le code ISO de la langue d'une métadonnée doit appartenir au référentiel des langues de NAKALA (Norme ISO-639-1 quand cela est possible, sinon Norme ISO-639-3).
- Les données déposées restent la propriété du déposant et demeurent sous sa responsabilité.
- Les données publiées ne peuvent pas être supprimées.

Liens utiles :

- <https://documentation.huma-num.fr/>
- <https://documentation.huma-num.fr/nakala-faq/>

à l'identifiant pérenne



L'identifiant pérenne DOI

Qu'est-ce qu'un DOI ? L'identifiant pérenne DOI (Digital Object Identifier, identifiant numérique d'objet), géré par le consortium DataCite est destiné à l'identification des ressources scientifiques et en particulier des données de la recherche. Ce système performant fournit un accès pérenne à la ressource par l'intermédiaire d'un lien cliquable permanent, et permet l'identification et le référencement des données les rendant ainsi découvrables, accessibles et citables (Opidor, 2023).

→ L'objectif est de fournir un identifiant unique et permanent aux objets numériques, tels que des articles de recherche, des publications, des données, des fichiers, etc.

Domaines d'applications

Les DOI sont largement utilisés dans le domaine de la recherche académique et de l'édition, que ce soit pour identifier les données ou pour les publications scientifiques dans toutes les disciplines confondues.

Qui en fait la demande ?

La demande ne peut être effectuée que de manière institutionnelle et non individuelle.

Coûts

Afin de pouvoir enregistrer des DOI, les organismes ou institutions de recherche doivent souscrire à un abonnement payant.

- Pour les établissements d'enseignement supérieur et les organismes publics, le coût en septembre 2023 est de 180 euros par an pour un nombre illimité de DOI.
- Pour les organismes privés, le coût s'élève à 180 euros HT par an pour 500 DOI.

Liens utiles :

- <https://coop-ist.cirad.fr/trouver-l-information/utiliser-un-doi/1-les-principes-du-doi>
- https://doranum.fr/identifiants-perennes-pid/zoom-doi_10_13143_jsxt-6j4/

L'agence habilitée à attribuer un DOI

L'Inist-CNRS est l'agence qui attribue des DOI pour les données de recherche en France via le service DataCite. Vous pouvez obtenir un DOI en utilisant le portail OPIDOR : <https://opidor.fr/identifier/>

⚠ Cependant, des DOI sont systématiquement attribués aux jeux de données lorsqu'ils sont déposés dans des entrepôts tels que Recherche Data Gov ou ceux des centres de référence.

Peut-on modifier ou supprimer un DOI ?

Non, le DOI ne peut être supprimé et ne changera jamais, même si la ressource est renommée ou déplacée (changement de serveur, d'entrepôt de données, etc.). Le DOI facilite ainsi l'identification, le référencement et la citation des données de la recherche.

Fonctionnement

Un préfixe propre à l'organisme est attribué et sera composé d'un nombre qui commence toujours par "10". Le suffixe peut être généré automatiquement ou choisi. Le DOI ainsi créé est associé à un fichier de métadonnées et à une URL stable qui peut renvoyer vers la ressource, ou vers une « landing page » (page descriptive de la ressource).

Exemple de DOI : <https://doi.org/préfixe/suffixe>

Téléchargements

URL DE TÉLÉCHARGEMENT	TÉLÉCHARGEMENTS UNIQUES	TÉLÉCHARGEMENTS
<input type="checkbox"/> adcvl.pu-cvl.fr	20	24
<input type="checkbox"/> /wp-content/uploads/2023/11/Fiche_synthetique_NAKALA.pdf	10	14
<input type="checkbox"/> /wp-content/uploads/2023/11/Fiche_synthetique_DOI.pdf	9	9

But : Cibler et synthétiser les informations essentielles

Des productions de ressources destinées aux Ateliers

Distinction entre URI, URN et URL			
	URI (Uniform Resource Identifier)	URN (Uniform Resource Name)	URL (Uniform Resource Locator)
	<i>L'URI est le sur-ensemble d'un URN et d'une URL.</i>	<i>L'URN est un sous-type spécifique d'URI.</i>	<i>L'URL est le sous-ensemble de l'URI.</i>
Définitions	identifie de manière unique une ressource, qu'elle soit localisable (URL) ou non localisable (URN) . C'est-à-dire que les identifiants pérennes peuvent être utilisés comme des URIs, et lorsque vous les incorporez dans une URL, vous créez une adresse web qui peut être utilisée pour accéder de manière persistante à la ressource associée. Les systèmes de résolution liés à ces identifiants garantissent que l'URL reste pérenne même si l'emplacement physique de la ressource change.	Un URN est un sous-type d'URI. Contrairement à une URL qui indique l'emplacement spécifique d'une ressource, un URN est conçu pour servir d'identifiant unique et persistant pour une ressource, indépendamment de son emplacement.	C'est un sous-type d'URI qui indique l'emplacement spécifique d'une ressource sur le web . Une URL est composée de plusieurs éléments, y compris le protocole (comme "http" ou "https", "ftp"), le nom de domaine, le chemin d'accès et éventuellement des paramètres.
Caractéristiques	L'identifiant pérenne, tel qu'un Handle, DOI, ou ARK, peut être utilisé comme une URI qui identifie de manière unique une ressource numérique.	>Les URNs sont conçus pour rester stables dans le temps, même si la ressource qu'ils identifient est déplacée ou si sa méthode d'accès change. >Contrairement aux URLs qui décrivent où trouver la ressource, un URN se concentre uniquement sur l'identification unique de la ressource elle-même.	Contrairement à un URN, cet identifiant de ressource Internet n'est pas toujours persistant et indépendant de l'emplacement puisqu'il spécifie le chemin d'accès à la ressource, qui peut changer au fil du temps.
Exemples	Pour l'URL : https://example.com/ressource (URL) Pour l'URN : urn:example:ressource (URN)	Pour l'URN : urn:doi:10.1234/example	En intégrant l'identifiant pérenne dans une URL, on peut créer une forme d'URL pérenne. Pour l'URL d'un DOI : https://doi.org/10.1234/example
Synthèses	URI est le terme générique qui englobe à la fois les URLs et les URNs.	URN est un type spécifique d'URI qui sert d'identifiant unique et persistant pour une ressource, indépendamment de son emplacement.	URL est un type spécifique d'URI qui indique comment accéder à une ressource, en spécifiant son emplacement.

Distinction entre les identifiants pérennes

C'est une chaîne de caractères alphanumériques qui identifie de manière unique :

LES IDENTIFIANTS OBJETS

doi

Une ressource

Un identifiant international qui donne accès aux ressources numériques (publications, données, revues, rapports, ect.)

DOI : 10.1234/5678

LES IDENTIFIANTS CONTRIBUTEURS

ORCID **ROR**

Un chercheur

Research Organization Registry

Pour identifier de manière distincte un chercheur, il est utilisé afin de simplifier la gestion des publications et des activités de recherche, ainsi que de faciliter la collaboration entre différentes bases de données.

Ces identifiants ROR facilitent la gestion, la recherche et l'interconnexion des informations sur les institutions de recherche.

ORCID ID : 0000-0003-0634-3277
ROR ID : <https://ror.org/03v4gs38>

Objectif → Accroître la visibilité et l'accessibilité des travaux de recherche



Relation entre URN, URL, et URI dans le contexte des identifiants pérennes (DOI, ARK, Handle)

	DOI (Digital Object Identifier)	ARK (Archival Resource Key)	Handle
URN (Uniform Resource Name)	Le DOI peut être considéré comme un URN car il fournit un identifiant unique et persistant pour une ressource, indépendamment de son emplacement.	L'ARK peut être vu comme un URN, car il sert également à identifier de manière unique et stable des ressources numériques.	Un Handle peut être considéré comme un URN, car il attribue un identifiant unique et stable à une ressource, indépendamment de son emplacement physique.
URL	Un DOI peut également être utilisé dans une URL. Exemple https://doi.org/10.1234 . L'URL facilite l'accès à la ressource identifiée par le DOI.	Un ARK peut être incorporé dans une URL. Exemple https://n2t.net/ark:/12345/x78901 . Pour permettre l'accès à la ressource associée.	Un Handle peut être utilisé dans une URL. Exemple https://hdl.handle.net/123/456 . Pour fournir un moyen d'accéder à la ressource identifiée.
URI (Uniform Resource Identifier)	Tous ces identifiants (DOI, ARK, Handle) peuvent être regroupés sous le terme générique d'URI, car ils sont tous des chaînes de caractères utilisées pour identifier de manière unique des ressources, qu'elles soient localisables (URL) ou non localisables (URN).		
Synthèse	Dans le contexte des identifiants pérennes, on peut les considérer à la fois comme des URNs pour leur rôle d'identification unique et stable, et comme des URLs lorsque incorporés dans une adresse web pour faciliter l'accès à la ressource. Le terme global pour les englober est celui d'URI.		

But : Cibler et synthétiser les informations essentielles

Des productions de ressources destinées aux chercheurs à l'échelle de RDG

À la découverte des centres de référence thématiques... DATA TERRA

Thématiques concernées	Sciences de la terre et de l'environnement Les cinq pôles de données et de services sont centrés sur des objets disciplinaires spécifiques au sein des sciences de la terre et de l'environnement : ➢ THEIA : surfaces continentales https://www.theia-land.fr/ ➢ AERIS : atmosphère https://www.aeris-data.fr/ ➢ ODATIS : océan https://www.odatis-ocean.fr/ ➢ ForM@Ter : terre solide https://www.poleterresolide.fr/ ➢ PNDP : biodiversité https://www.pndp.fr/
Données concernées	Tous types de données du système terre et de l'environnement
Typologie de données traitées	Données de longue traîne Données d'observation prises en charge par les pôles de données et de services, avec spécificités propres à chaque pôle. Ces dernières sont acquises à partir de moyens basés au sol, d'instruments montés à bord de satellites, de ballons, d'avions, de bateaux ou de drones, de mesures de laboratoire, de modèles complétant les jeux d'observations, etc...
Statut administratif	Infrastructure de recherche virtuelle, Data Terra s'appuie sur l'UAR* CPST 2013 (Coordination des Pôles de données système Terre).
Missions	<ul style="list-style-type: none"> Faciliter l'accès aux données multi-sources Développer des services qui couvrent tout le cycle des données Développer des produits et services répondant aux critères FAIR pour chaque composante du système Terre et leurs interactions Coordonner, fédérer et optimiser toutes les institutions, structures et ressources existantes dans ce domaine Permettre la mutualisation et l'interopérabilité des services Mettre en œuvre des approches multidisciplinaires intégrées pour l'utilisation des données de recherche en observation de la Terre Contribuer aux initiatives internationales et européennes en soutien aux politiques publiques de développement durable
Échelles d'intervention	<ul style="list-style-type: none"> Européenne DataTerra est partie prenante de plusieurs projets européens autour de la gestion de données FAIR (EOSC-PILLAR, FAIR-EASE, ENVRI FAIR...) et des infrastructures de calcul et de stockage (PHIDIAS). Les pôles de données et de services participent par ailleurs également à de nombreux projets, programmes, consortiums de recherche européens. <ul style="list-style-type: none"> Nationale

recherche.data.gov.fr

Services : Outils mis à disposition	DataTerra intervient auprès des équipes de recherche concernées via les pôles de données et de services (parfois représentés par des centres de données et de services en local). <ul style="list-style-type: none"> FaSy Data : entrepôt de diffusion des données de longue traîne Dinamis : plate-forme d'acquisition et de diffusion de données spatiales d'observation de la Terre Earth System Analytics Labs (ESAL) et Environnements Virtuels de Recherche (VRE) : outils d'analyse et de traitement Centres de données HPC-Tier2 : ressources de calcul et stockage dédié à l'hébergement et l'exploitation de données volumineuses. D'autres outils sont développés et mis à disposition par chaque pôle.
Services : Accompagnement des chercheurs	<ul style="list-style-type: none"> Formations et séminaires Accès aux données, description des données (modèles de métadonnées, vocabulaires partagés, ontologies alignées) Services de FAIRisation Attribution d'identifiants pérennes.
URL	Site web : https://www.data-terra.org/ CATOPIDoR : https://cat.opidor.fr/index.php/DATA_TERRA
Contact	contact@data-terra.org

Fiche produite par le réseau des ateliers de la donnée
Dernière mise à jour : 16/05/2024

Fiches pratiques sur les centres de référence thématiques

À la découverte des centres de référence thématiques... PROGEDO

Thématiques concernées	Sciences humaines et sociales
Données concernées	Données quantitatives
Typologie de données traitées	Données d'enquêtes de la statistique publique issues des ministères (justice, éducation, agriculture), données administratives (données fiscales notamment), données d'entreprises.
Statut administratif	PROGEDO (PROduction et GÉstion des DONnées) est une infrastructure de recherche étoile (IR*), qui s'appuie sur l'unité d'appui et de recherche PROGEDO (UAR2506). Soutenir la création (notamment par du financement) et la diffusion de grandes enquêtes internationales et pluridisciplinaires. Exemples : <ul style="list-style-type: none"> ESS (<i>European Social Survey</i>), SHARE (<i>Survey on Health, Ageing and Retirement in Europe</i>), GGP (<i>Generations and Gender Programme</i>), EVS (<i>European Value Study</i>), ISSP (<i>International Social Survey Programme</i>), etc.
Missions	<ul style="list-style-type: none"> Développer en SHS l'utilisation de données quantitatives (principalement issues d'enquêtes de la statistique publique) par les chercheurs.
Échelles d'intervention	<ul style="list-style-type: none"> Européenne Investissement dans les consortiums européens (ERIC CESSDA <i>Consortium of European Social Science Data Archives</i>) et diffusion des grandes enquêtes. <ul style="list-style-type: none"> Nationale Avec les départements de la documentation et la diffusion des données par le département Quetelet-PROGEDO Diffusion <ul style="list-style-type: none"> Locale Via les plateformes universitaires de données implantées sur de nombreux sites universitaires.
Services :	Quetelet-Progedo-Diffusion : entrepôt et catalogue de données qui propose plus de 1500 jeux de données, issus : <ul style="list-style-type: none"> des <i>Archives de Données Issues de la Statistique Publique</i> :

recherche.data.gov.fr

Outils mis à disposition	enquêtes et bases de données produites par l'INSEE, plusieurs services statistiques ministériels (DARES, INSEE) et des institutions de recherche publiques. <ul style="list-style-type: none"> du service d'enquêtes de l'Ined. Centre d'Accès Sécurisé aux Données : équipement hautement sécurisé permettant de travailler à distance sur des bases de données individuelles détaillées, dans le respect des exigences imposées par le comité du secret statistique.
Services : Accompagnement des chercheurs	Les 17 Plateformes Universitaires de Données (PUD) visent à stimuler l'utilisation des données quantitatives en SHS, plus particulièrement celles mises à disposition au travers du dispositif Quetelet-PROGEDO Diffusion. Elles apportent un soutien direct aux utilisateurs de données d'enquêtes par : <ul style="list-style-type: none"> l'aide à la recherche, à l'acquisition et à la mise en œuvre des fichiers de données, la formation à la compréhension, à l'utilisation et au traitement de ces données, la formation à l'utilisation d'outils informatiques et statistiques.
URL	Site web : https://www.progedo.fr Carnet hypothèse : https://progedo.hypotheses.org Entrepôt thématique : https://data.progedo.fr
Contact	info@progedo.fr

Fiche produite par le réseau des ateliers de la donnée
Dernière mise à jour : 19/03/2024

Diffusion dans Recherche Data Gov



But : Cibler et synthétiser les informations essentielles

Communication



Inauguration - 09 avril

- Journée d'étude : interventions des différents acteurs de Recherche Data Gouv
- Cérémonie d'inauguration : présence de tous les partenaires



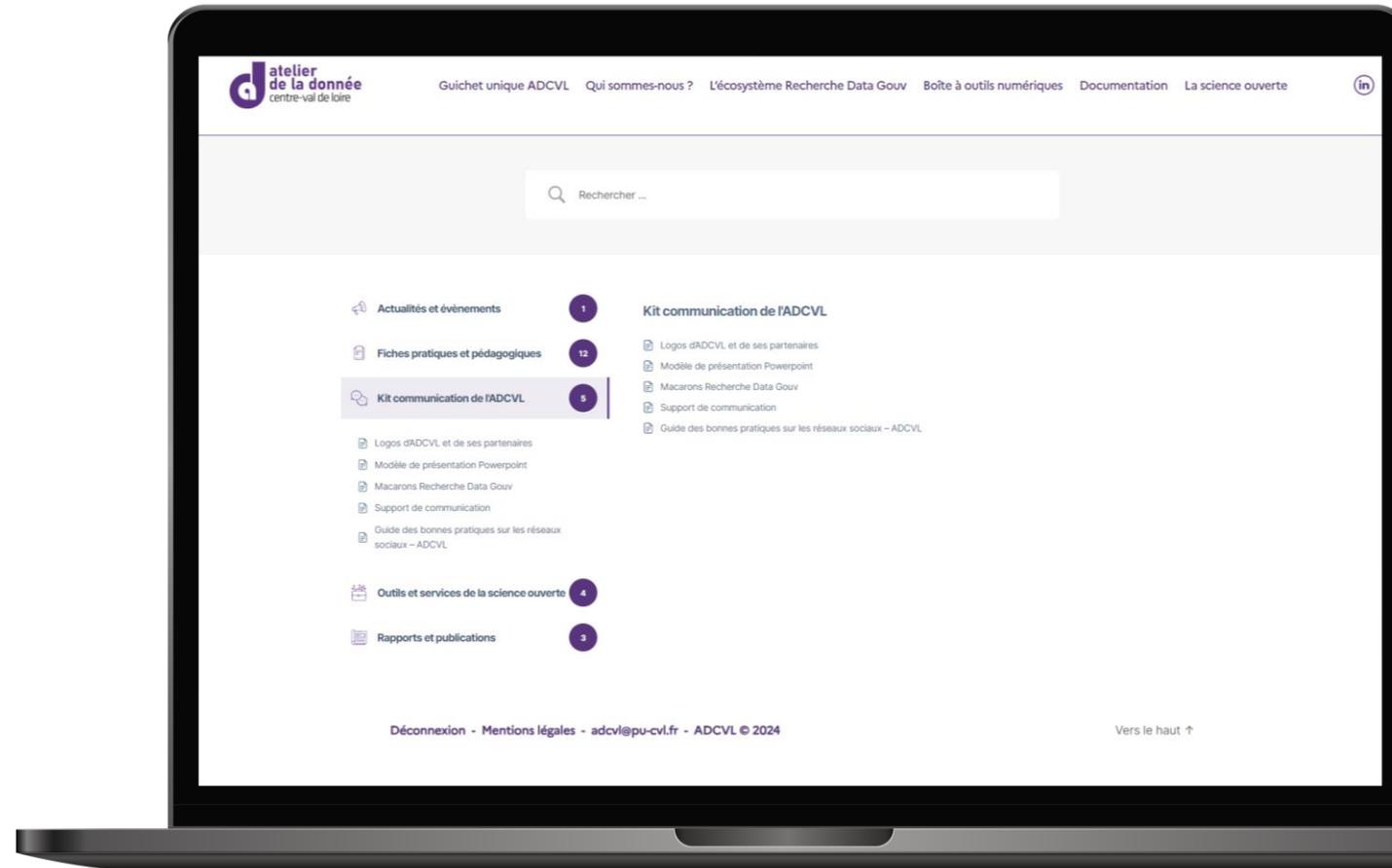
Communication interne

- Newsletter interne pour les membres du comité opérationnel
- Envoi mensuel à chaque fin de mois
- Composées de 3 grandes parties :
 - Des actualités sur l'ADCVL
 - Un agenda
 - Quoi de neuf en science ouverte



Kit communication

→ Sur le site de l'ADCVL
dans la rubrique
« documentation »

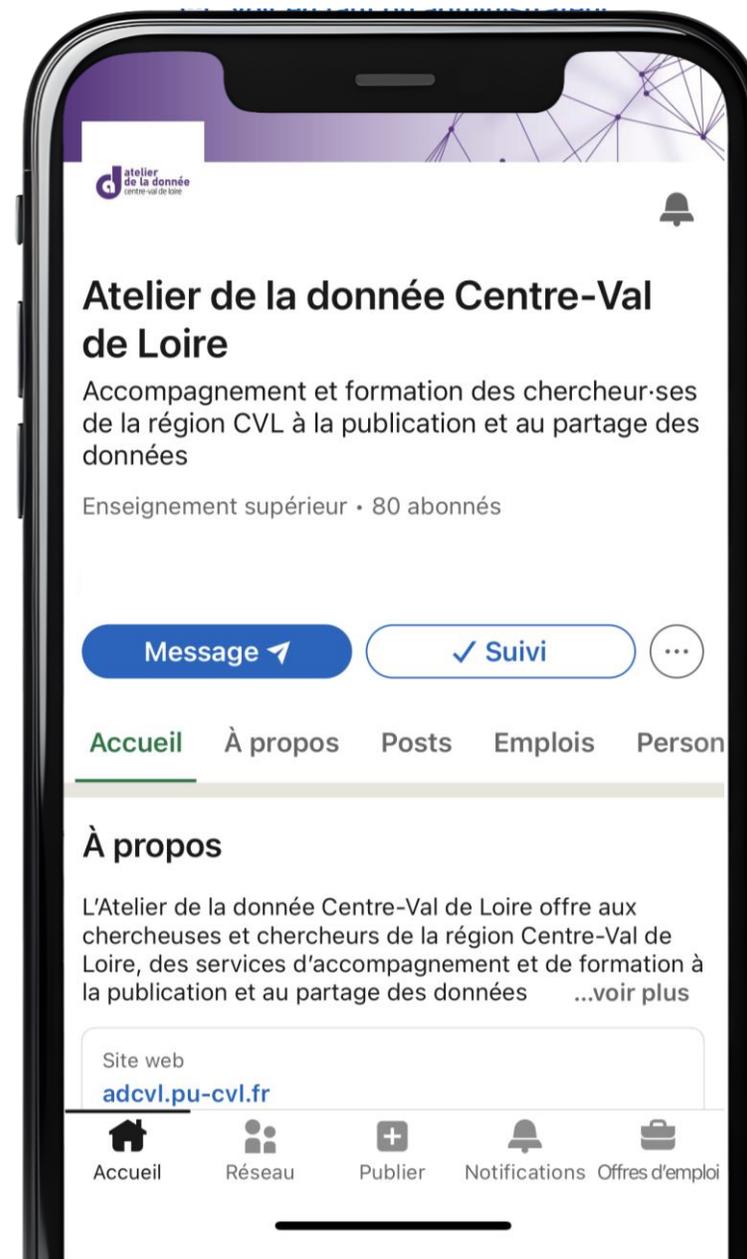


Communication externe : LinkedIn

- + de visibilité
- entretenir des relations professionnelles avec nos partenaires et les laboratoires et chercheur.ses de la région

Types de contenus :

- Mise en avant d'événements
- Mise en avant de formations ou de webinaires
- Informations liées à l'actualité de l'Atelier et à la science ouverte
- Offres d'emploi



Communication externe : Actions futures

- Réalisation d'une vidéo présentation de l'ADCVL
- Interventions dans chaque laboratoire = nous faire connaître et sensibiliser à la science ouverte

Pôle Université d'Innovation



Recrutements

**Ingénieur.e d'études chargé.e
du suivi des codes sources et
des logiciels**

→ Rémi Ossant

**Ingénieur.e d'études chargé.e
d'appui à la gestion des
données de la recherche**

→ Clara Monmarché

Formations



Formation URFIST

- 28 mars 2024
- Programme :

Matin : rappel autour de quelques notions essentielles

- Les données et leur définition
- Le cadre juridique et réglementaire actuel
- Les principes FAIR

Après-midi : état des lieux et perspectives pour l'ADCVL

Quels sont **vos besoins** en termes de formation ?

- Recherche de données de confiance
- ...

Les groupes de travail de l'ADCVL

02

GT Accompagnement au dépôt de jeux de données dans l'entrepôt approprié

- Service le plus sollicité
- Liste des entrepôts de confiance :
2 webinaires pour présenter la méthodologie et de la liste des entrepôts

Liste des critères d'exclusion

- Absence de modération des dépôts
- Absence d'identifiant pérenne
- Absence de garantie sur la pérennité de l'infrastructure
- Entrepôts pratiquant la cession de droits
- Politique tarifaire excessive
- Localisation des données hors Union européenne pour certains types de données
- Dépôt restreint par l'affiliation institutionnelle

Liste des critères de description

- Champ disciplinaire
- Données acceptées
- Fourniture d'un identifiant pérenne
- Pérennité des données
- Type de modération
- Possibilité d'embargo
- Limite de volume

GT Plan de gestion de données

- 2^{ème} service le plus sollicité
- Articulation entre le groupe de travail et le nouveau poste qui a été obtenu dans le cadre du PUI

GT Sensibilisation et formation

- Mise en place de formations par l'ADCVL
- Prochainement une formation sur le PGD
- Sinon sur quelles thématiques ? Quels intervenants ?

**Est-ce que vous proposez des formations
qui rentrent dans les thématiques de l'Atelier ?**

Formation Nakala – 25 mars 2024

Une journée en immersion avec les chercheurs en trois points :

- Entrepôt de données : concept, intérêt et exemple
- L'infrastructure de recherche Huma-Num
- Le dépôt de données en pratique dans Nakala



25 mars 2024

Formation Nakala :

Accompagnement au dépôt de données

GT Conseils éthiques et juridiques

- Evolution du groupe ? Possibilité de nouveaux personnels ?
Conditions d'ouverture ?
- Des demandes de part des chercheurs semblent émerger
concernant des aspects autour du RGPD et des licences.

GT Préservation et sécurité des données

→ 3 sollicitations

GT FAIRisation des données

- Pas de sollicitation
- Un inventaire des guides sur les principes FAIR au sein de différentes disciplines

L'évolution de l'ADCVL

03

La mise en place de correspondants données pour l'ADCVL

→ Premières réflexions sur le rôle des "correspondants "

- Un dans chaque unité de recherche ? un dans chaque labo ?
- Quel serait leur rôle ? Chargé de faire connaître l'ADCVL ? Chargé d'orienter les chercheurs vers l'atelier en cas de questionnement sur les données de recherche ? Charger de faire des informations de la part de l'ADCVL aux chercheurs ?

Les points incontournables

identifiés par GT mise en œuvre d'un réseau de correspondants données dans les laboratoires du réseau des ateliers :

- État des lieux préalable
- Appui politique indispensable
- Granularité du réseau adaptable
- Mode de désignation et valorisation du rôle de correspondant
- Modalités de formalisation de la mission
- Missions des correspondants
- Vie du réseau

L'écosystème Recherche Data Gouv

Écosystème Recherche Data Gouv

Journées RDG – 4 au 6 juin

- La communication
- L'évolution de l'écosystème, la mise en place d'outils pour accompagner son élargissement et les interactions entre les différents dispositifs

Bilans des réunions Recherche Data Gouv

- Réunion des responsables des ateliers de la donnée
- Réunion sur les relations entre la DDOR CNRS et les ateliers de la donnée ayant le CNRS en partenaire

Avez-vous des questions ?

Merci

 <https://adcvl.pu-cvl.fr/>

 @Atelier de la donnée Centre-Val de Loire