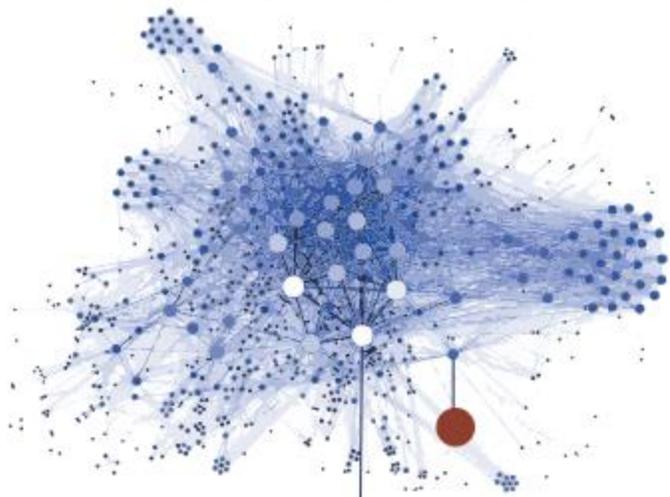




Ouvrir et gérer les données de la recherche
en Bourgogne Franche-Comté



dat@osu

Le portail de référencement des
données de recherche de l'OSU THETA

Sylvie Damy⁽¹⁾, Bernard Debray⁽²⁾,
Raphaël Melior⁽³⁾, Hélène Tisserand⁽³⁾, Gaëlle Laporte⁽³⁾

(1) Laboratoire Chrono-environnement

(2) Institut UTINAM

(3) OSU THETA





- Projet lancé fin 2013 (S. Damy et B. Debray)
- Objectif : améliorer les pratiques des chercheurs et des structures de recherche locales en terme de gestion des données de la recherche

Mise en place de l'OSU THETA (2007-2009) avec un axe fédérateur bases de données, interopérabilité

- Création de l'axe transversal bases de données (2012) de Chrono-environnement et projet OREAS (CDD ingénieur informaticien)
- Outils de simulation et bases de données en astronomie et en physique, lien avec l'observatoire virtuel (Utinam, SMPCA/LICB)
- Bases de données et réseaux en climatologie et environnement (Biogéosciences)





OSU THETA



Pluri-disciplinarité



Contexte local

Journée Bases de Données (mars 2013)

- Co-organisation OSU THETA – Axe transversal BD de Chrono-environnement
- 38 participants : Besançon, Dijon, Montbéliard + quelques extérieurs
- 12 présentations

Participation au FRéDoc (octobre 2013)

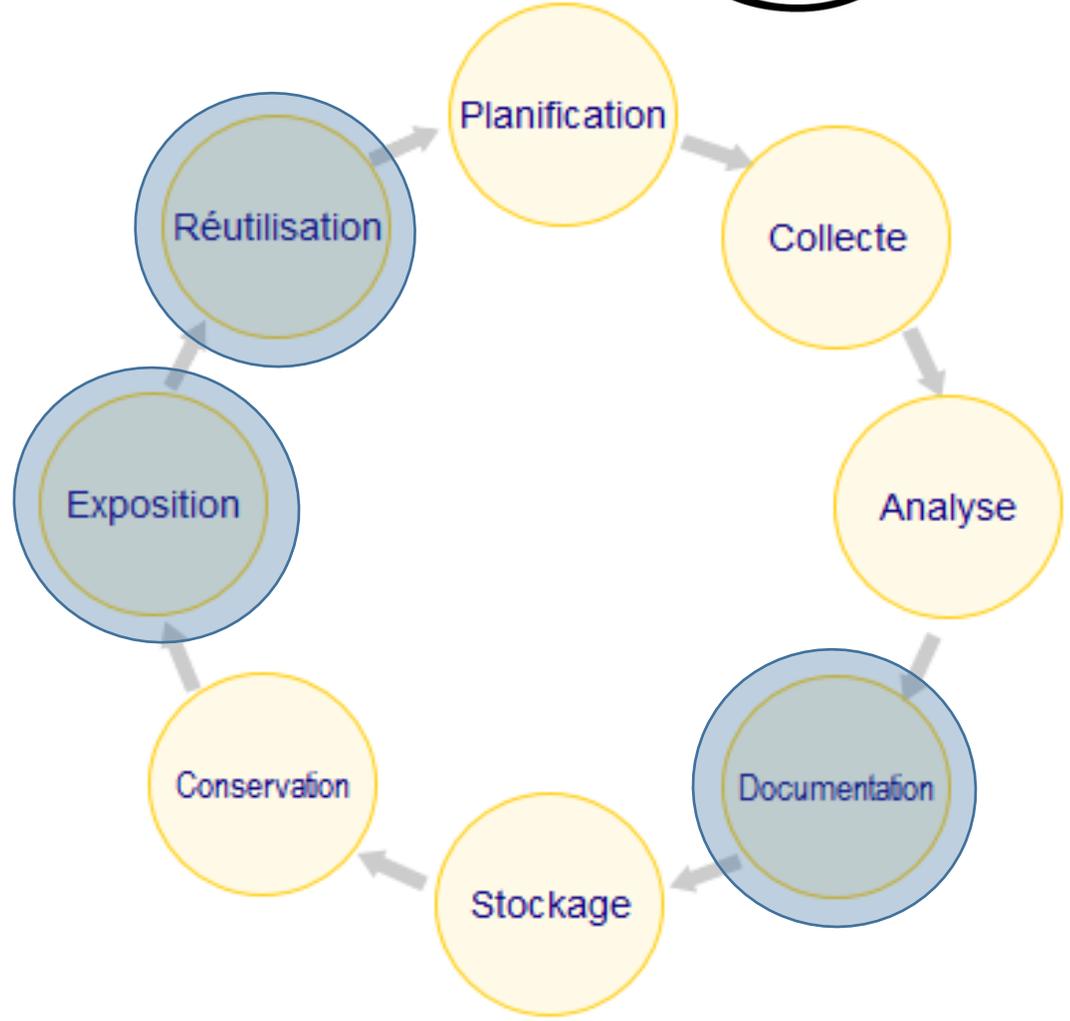
- « *Gestion et valorisation des données de la recherche* »



Le projet dat@osu

Référencement et valorisation des données produites dans les laboratoires et équipes de l'OSU THETA

Mise en place des outils et procédures permettant de décrire et faire connaître les données de recherche





Objectifs du projet

- rendre visible
 - partager
 - valoriser
 - pérenniser
- } les données

Valoriser les structures : observatoire, laboratoires, équipes, ...

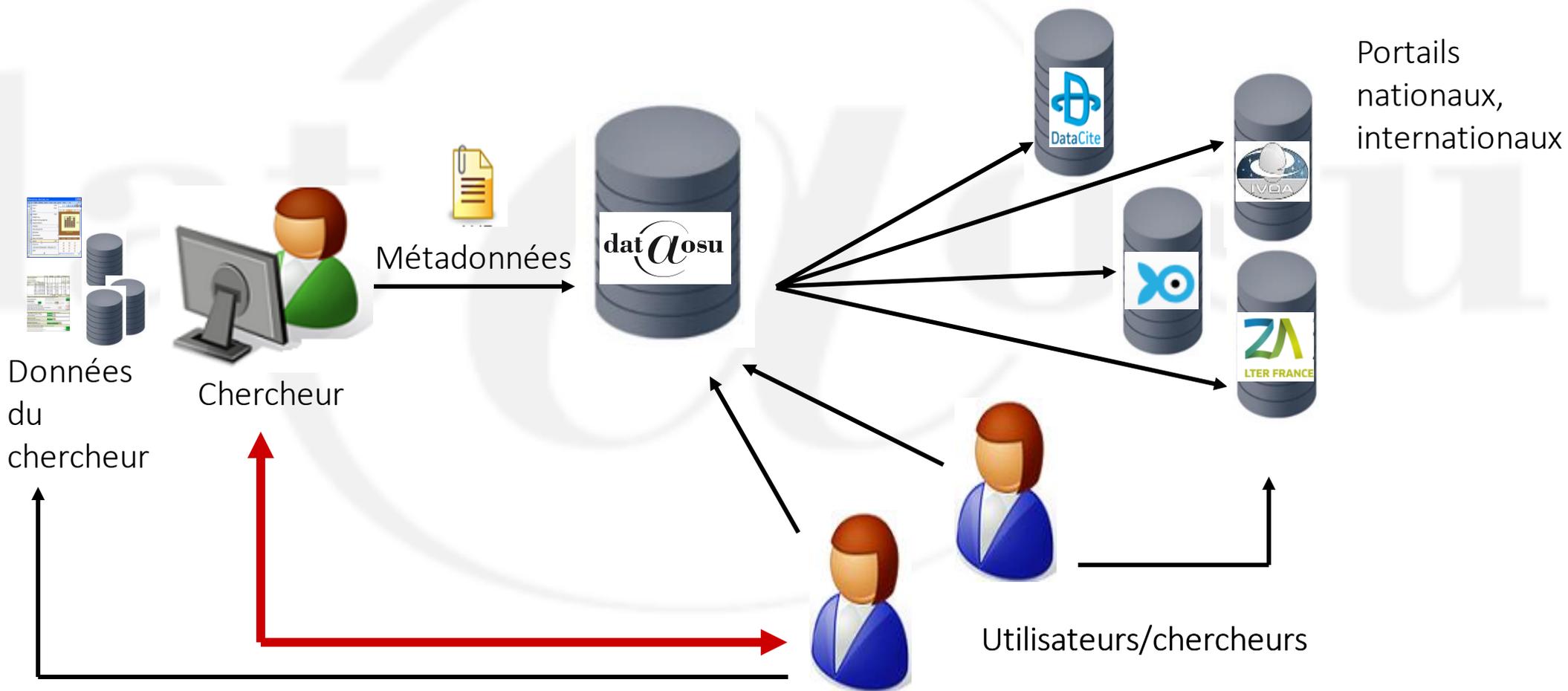
Homogénéisation de présentation

Meilleure diffusion et exploitation des données

Mutualisation des efforts et des savoir-faire

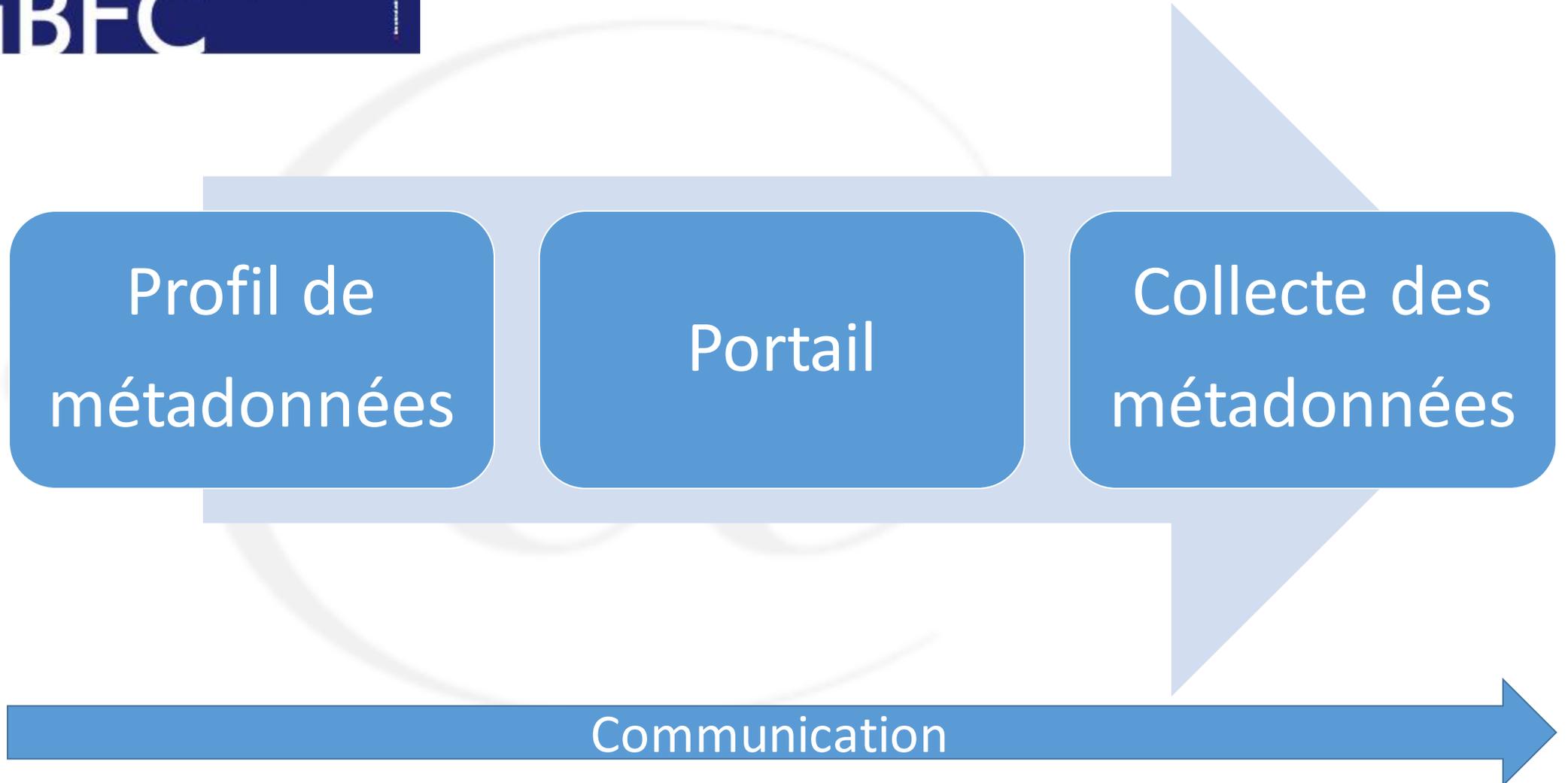


Le projet dat@osu





Les étapes du projet





Organisation du projet



S. Damy Chrono-environnement
B. Debray UTINAM
(G. Laporte OSU THETA
05/2014 - 03/2017)
R. Melior OSU THETA
H. Tisserand OSU THETA

M. Pochard SCD UFC

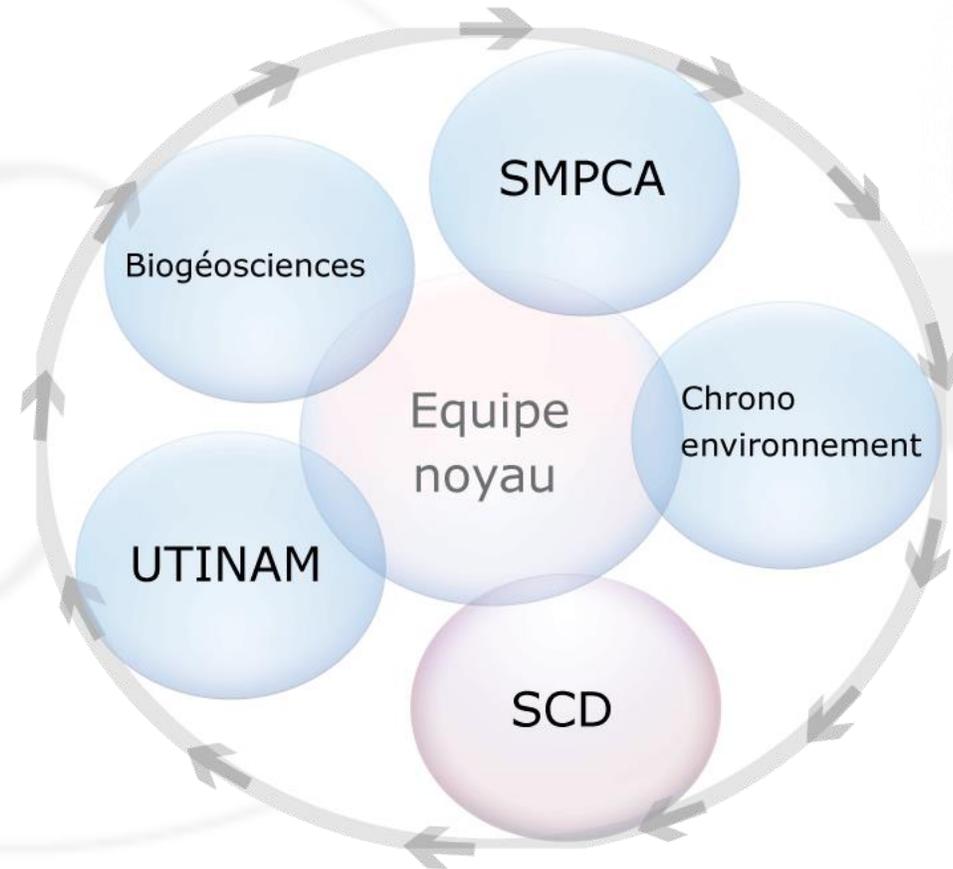




Organisation dans les laboratoires

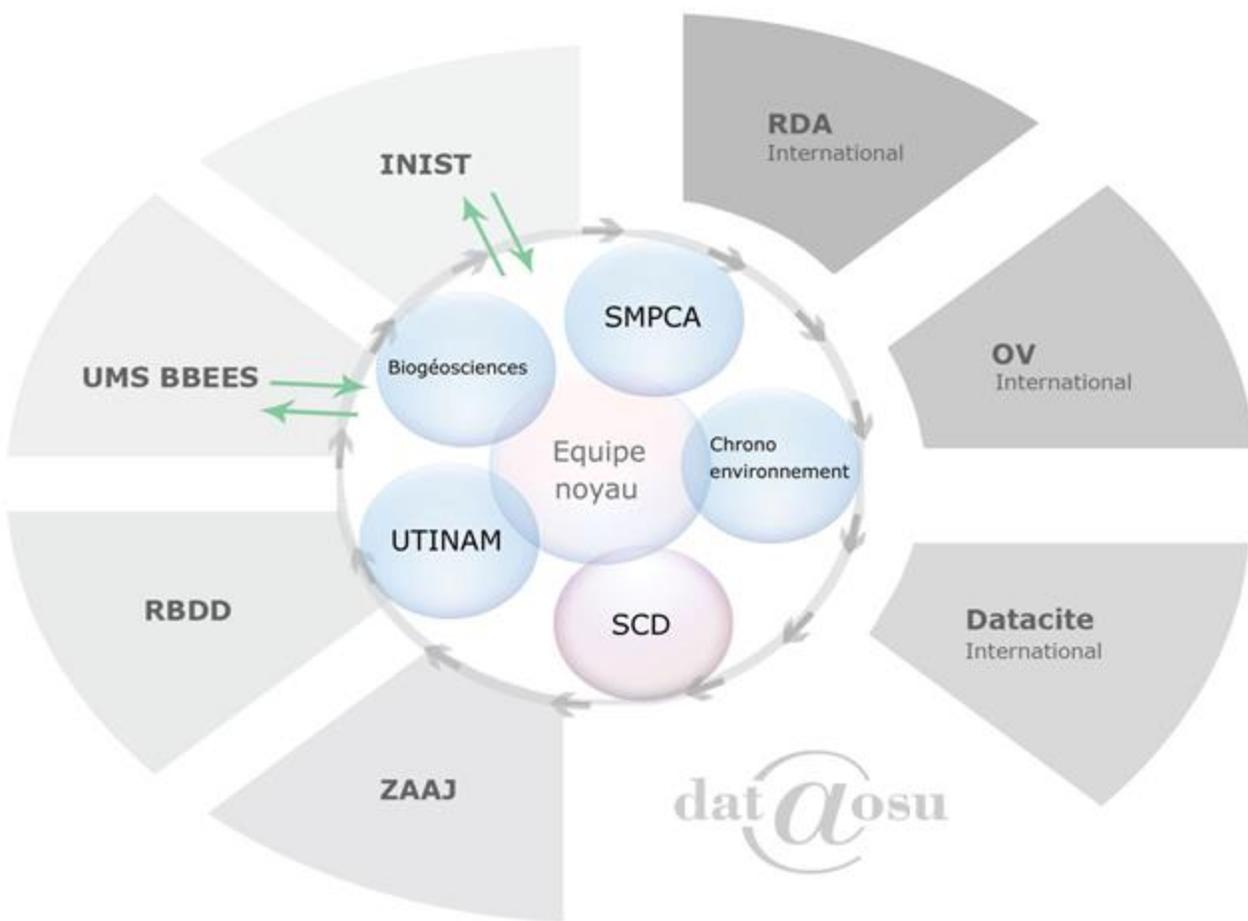
Les correspondants

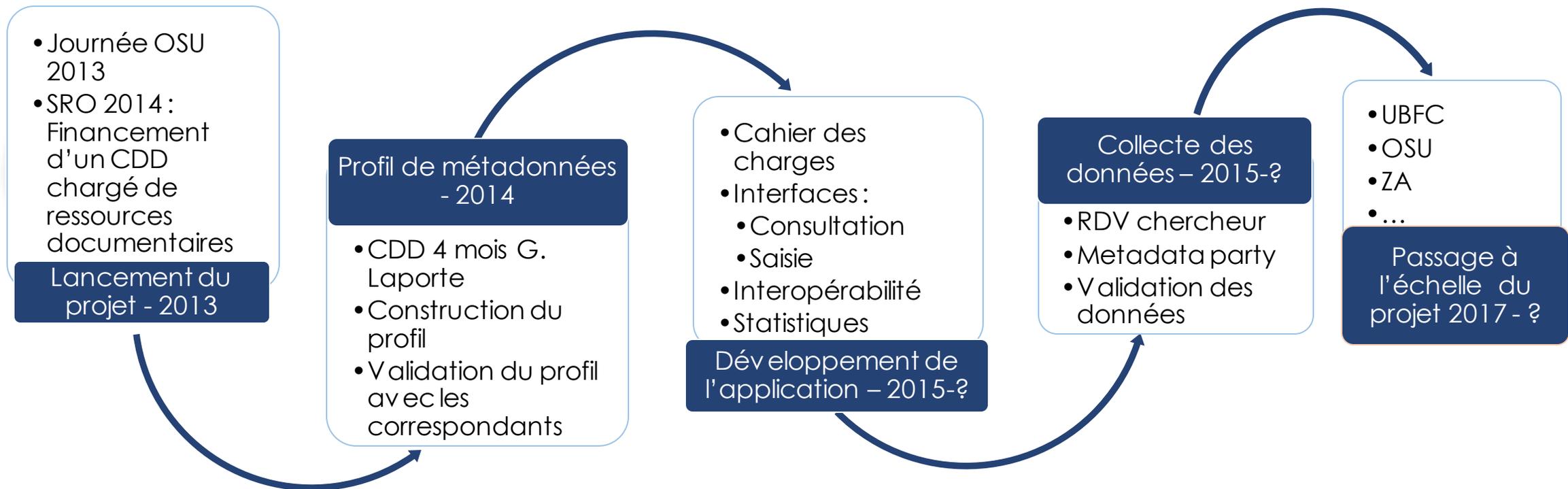
- F. Raoul *Chrono-environnement*
- C. Reylé *OSU THETA, UTINAM*
- V. Boudon *LICB-SMPCA*
- B. Pohl *Biogéosciences*





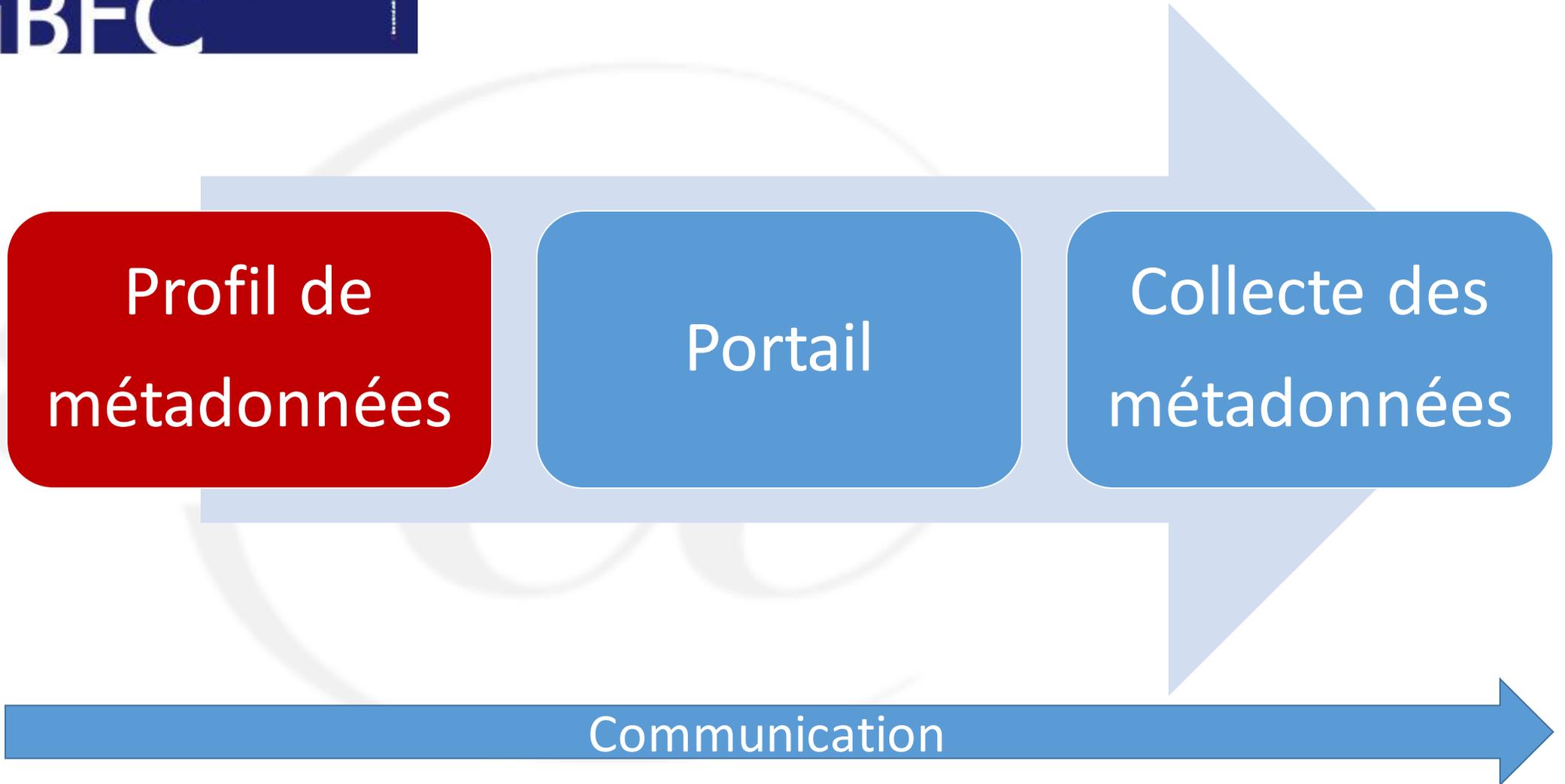
Organisation avec l'extérieur







Les étapes du projet



Profil de
métadonnées

Portail

Collecte des
métadonnées

Communication



Profil de métadonnées

- 1- Etat de l'art concernant l'ouverture des données de la recherche ou *open research data*
 - Directives
 - Projets, groupes de travail
 - Standards de métadonnées
- 2 - Conception du profil de métadonnées

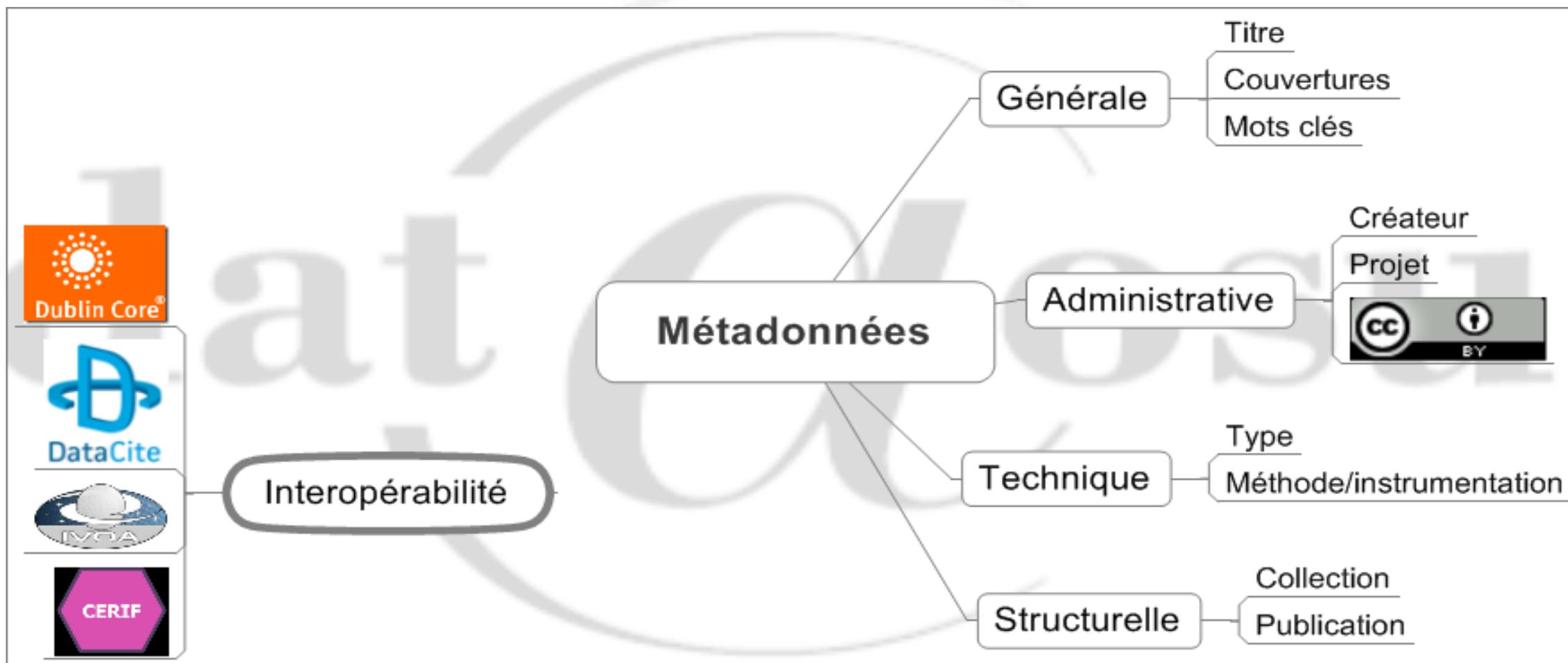


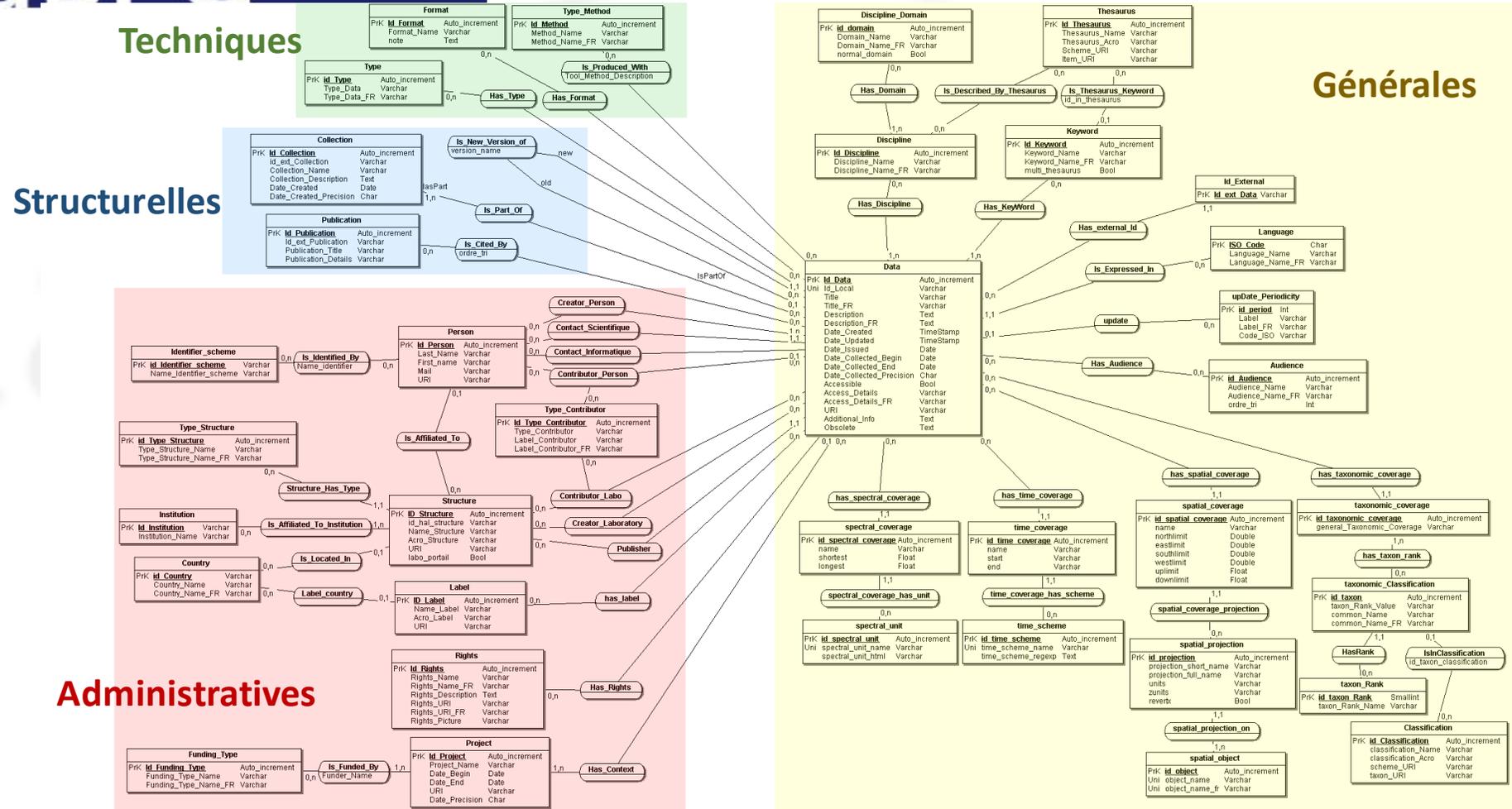
FAIR metadata

Les métadonnées doivent être :

- trouvables (Findable)
- accessibles (Accessible)
- interopérables (Interoperable)
- réutilisables (Re-usable)

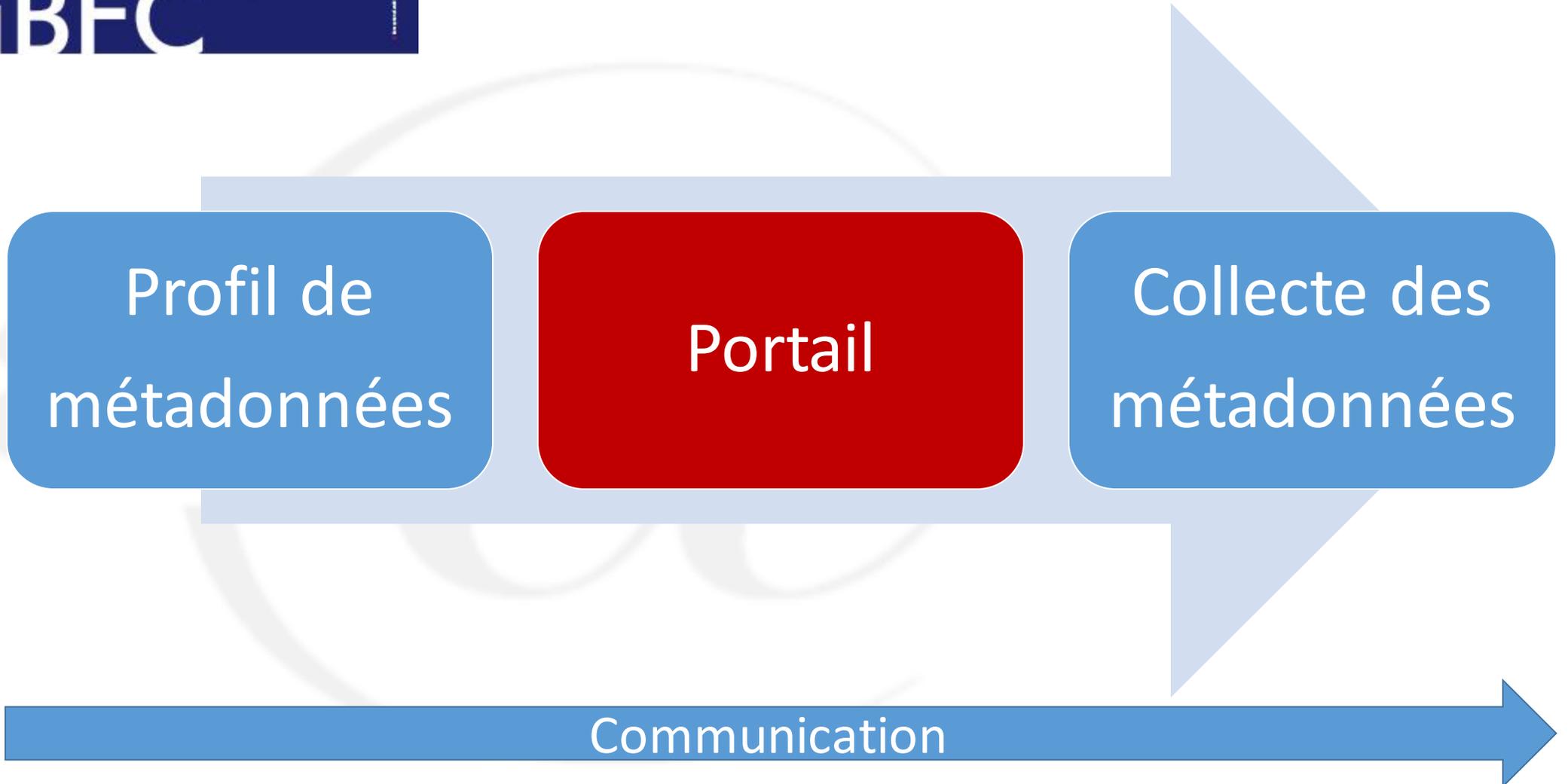
- **Dublin Core** : schéma de métadonnées générique qui permet de décrire des ressources numériques ou physiques et d'établir des relations avec d'autres ressources (1995)
- **Datacite** : Consortium international centré sur la valorisation des jeux de données et autres données non textuelles dans les services d'information (2009)
- **CERIF** : Format informatique utilisé pour représenter les données d'un système d'information de recherche (Commission européenne)
- **OV : Resource Metadata (2007)** : Méta-données génériques de haut niveau pour décrire des ressources.
- **EML** : Ecological Metadata Language. Spécification de métadonnées spécifique aux disciplines écologiques.
- **ISO 19 115**: Norme de référence pour l'information géographique.







Les étapes du projet



Profil de
métadonnées

Portail

Collecte des
métadonnées

Communication



Le portail Dat@OSU

Application destinée à être utilisée :

- en consultation : tout public
- en saisie : chercheurs
- en contrôle/validation : documentaliste

Interopérable : « moissonnable » par des portails de référence des données de la recherche

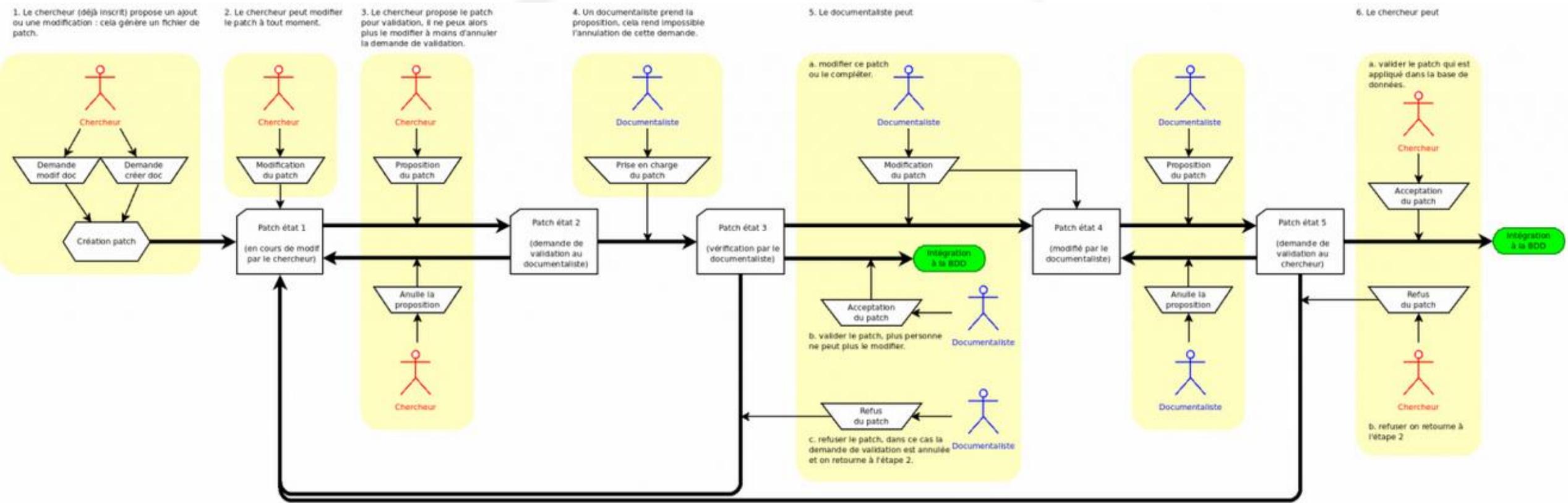


Le portail Dat@OSU

- Réalisation du cahier des charges
 - novembre 2014 – mars 2015 (étude de portails existants)
 - en fonction du profil de métadonnées
- Développement des interfaces
 - consultation
 - saisie
 - gestion / administration



Création d'une fiche de métadonnées



Jeu de données : Prairies permanentes de l'Arc Jurassien (étude synchronique BioGEcoP)

Métadonnées générales

Identifiant : local : FR-18008901306731-2017-02-16-02

Description : Végétation, communautés d'orthoptères, descripteurs du sol et de l'utilisation agricole de 48 prairies de l'Arc Jurassien français.

Disciplines :

agriculture (biologie appliquée - écologie), biodiversité, conservation (biologie appliquée - écologie), botanique, biologie végétale (biologie appliquée - écologie), écologie (biologie appliquée - écologie), entomologie (biologie appliquée - écologie), sciences de l'environnement (sciences de l'univers)

Mots clés : communauté végétale, écosystème prairie, pâturage, pré, sol, végétation

Dates : Obtention des données : entre mars 2012 et oct. 2013

Fiche de métadonnées : Création : 16 févr. 2017 Publication : 1 oct. 2017

Rythme de mise à jour : pas de mise à jour Langue : Anglais (eng)

Audience : Recherche

Contact accès aux données



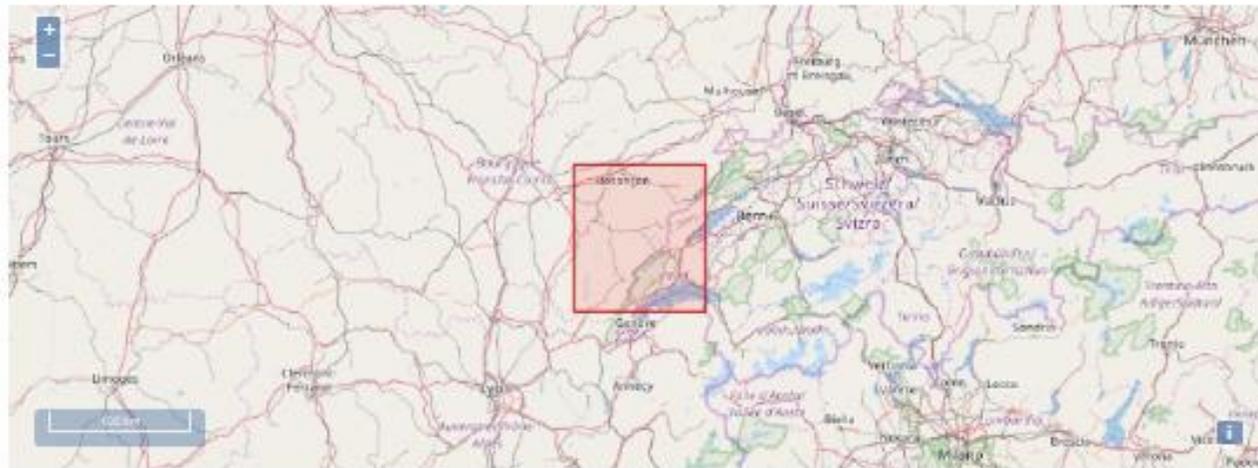
Citation

François Gillet, Leslie Mauchamp, Arnaud Mouly (2017): Prairies permanentes de l'Arc Jurassien (étude synchronique BioGEcoP). Chrono-environnement. FR-18008901306731-2017-02-16-02

Couvertures

Couverture spatiale :

- **Périmètre BioGEcoP étude synchronique** : sur Terre, latitude entre 47° 16' 36" N et 48° 20' 3" N, longitude entre 5° 33' 24" E et 6° 47' 15" E, entre 391m et 1195m d'altitude



Couverture temporelle :

- (sans titre) : entre 2001 yr common era et 2012 yr common era



Data

d

u

Quelques particularités

Sur terre

Couverture spatiale

Dans le ciel

Couvertures

Couverture spatiale :

- **Franche-Comté** : sur Terre, latitude entre 47° 47' 59" N et 46° N, longitude entre 5° E et 7° E

Couverture temporelle :

- **Dernier millénaire** : entre 1144 et 1899

Couverture taxonomique :

- **Species** : *Abies alba* (sapin pectiné)

Couvertures

Couverture spatiale :

- **Allsky** : dans le ciel, latitude entre 90° N et 90° S, longitude entre 360° et 0°, projection J2000

Couverture temporelle :

- **Binary star observations** : 1690 et après

Couverture spectrale :

- **Visible** : entre 300 nm et 1 µm
- **X-Ray** : entre 120 eV et 120 keV



Quelques particularités

Recherche à facettes

audience:Décisionnaire

7 résultats Tri : Nb de fiches :

- sciences de l'environnement
- météorologie
- géosciences
- écologie
- géographie physique
- histoire
- multidisciplinaire
- océanographie

[+]

- Chrono-environnement
- Biogéosciences
- FEMTO-ST
- THETA
- UTINAM

Label

- SNOT
- ZAAJ
- ZATA

Langue des données

- anglais
- français

Projet

- Dat@OSU : Gestion et valorisation des données de la recherche des laboratoires de l'Observatoire des Sciences de l'Univers Terre-Homme-



Dendrochronologie, dendroarchéologie, dendroclimatologie du sapin en Franche-comté
Jeu de données Disciplines : archéologie histoire sciences de l'environnement Date de collecte : depuis 1990 et en cours Mots clés : Dendrochronologie dendroclimatologie
Base de données de séries individuelles de largeur de cerne de sapin (Abies alba), acquises lors des études dendroarchéologiques de sites archéologiques et monuments historiques en Franche-Comté pour la période du dernier millénaire.



Suivi météorologique sur la tourbière du Forbonnet
Jeu de données Disciplines : météorologie sciences de l'environnement Date de collecte : depuis nov 2008 et en cours Labels : SNOT
Mots clés : climat direction du vent pluviométrie surveillance température température de l'air tourbière vitesse du vent
Le laboratoire Chrono-Environnement assure le suivi météorologique d'une station expérimentale sur la tourbière de Frasne (Tourbières du Forbonnet, Doubs) depuis novembre 2008. Les paramètres météorologiques (température / humidité de l'air, pluviométrie, vitesse et direction du vent, niveau de ...



Mesures mobiles de température de l'air dans l'agglomération de Dijon
Jeu de données Disciplines : géosciences météorologie Date de collecte : depuis mars 2016 et en cours
Mots clés : climat climatologie climat urbain îlot de chaleur urbain surveillance température
Mesures mobiles par vélo de température de l'air dans l'agglomération de Dijon à très haute résolution (une mesure tous les 5m) pour caractériser l'îlot de chaleur urbain le long de transects constants

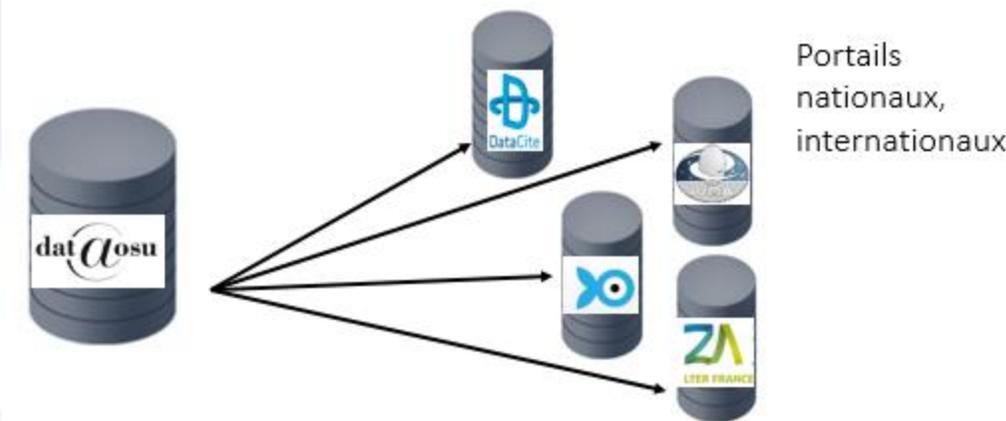


Dat@OSU
Base de données Disciplines : multidisciplinaire sciences de l'information Date de collecte : depuis 2015 et en cours
Mots clés : Données de la recherche Données ouvertes Métadonnées Science ouverte
Dat@OSU est une plateforme de métadonnées proposée par l'Observatoire des Sciences de l'Univers Terre Homme Environnement Temps Astronomie (OSU THETA). Elle propose une description des ensembles de données numériques issues des recherches scientifiques menées dans les différents laboratoires et ...



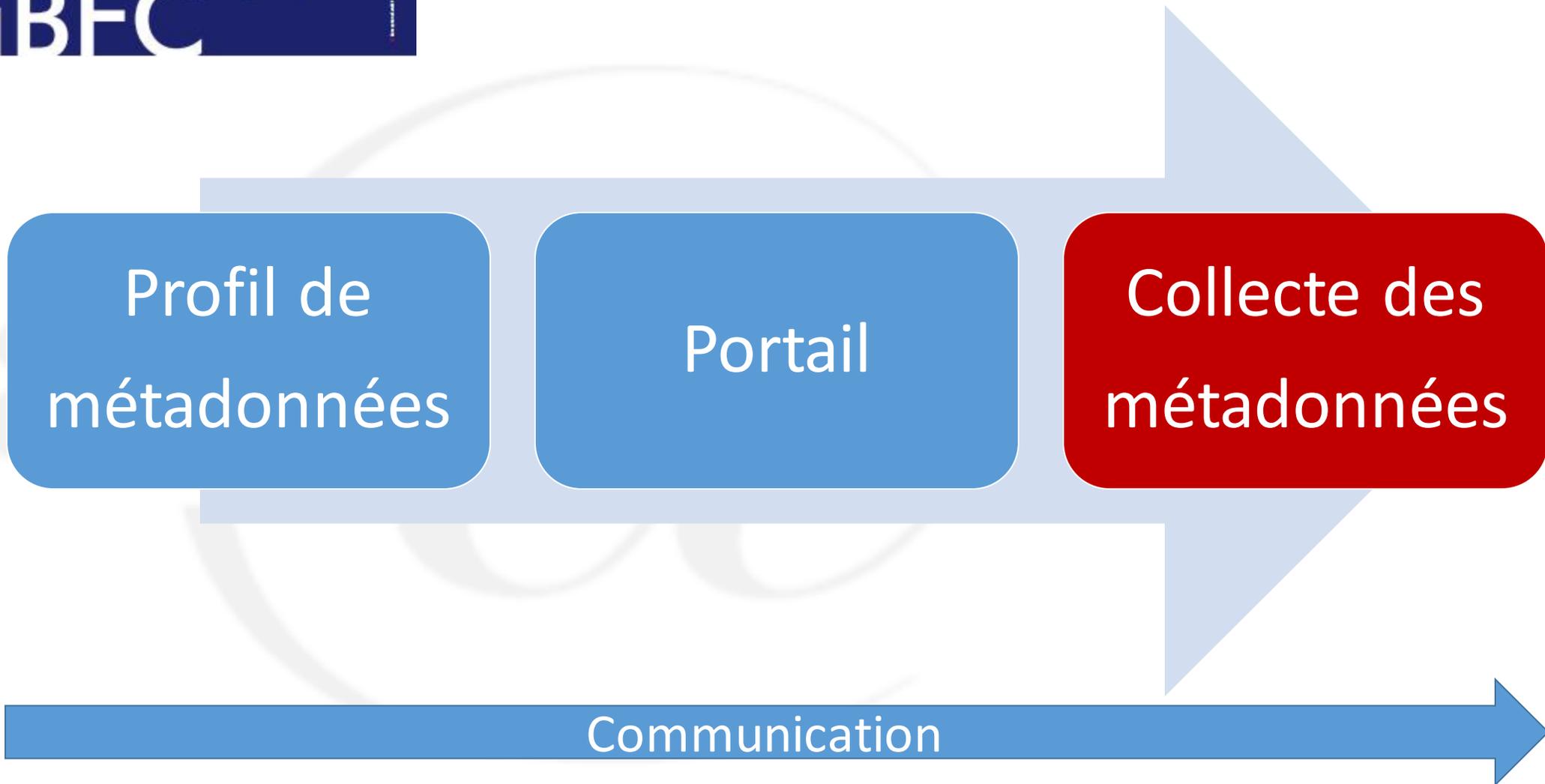
Carte bruit Besançon 2008
Jeu de données Disciplines : acoustique santé publique et environnement sciences de l'environnement Date de collecte : entre janv. 2006 et juil. 2007 Labels : ZAAJ
Mots clés : bruit bruit urbain épidémiologie
Carte du bruit environnemental avant tram (trafic routier, ferroviaire, rues piétonnes, fontaines à vasque)

- Mise en place d'un serveur OAI-PMH
 - Moissonnage par ISIDORE (TGIR Huma-Num)
 - Travail en cours pour transférer les métadonnées vers :
 - Datacite
 - ...





Les étapes du projet



Collecte des métadonnées

Saisie d'une fiche de métadonnées complète et valide dans le portail :

- temps : 1 à 2 heures
- utilisation de référentiels : mots-clés (thésaurus), personnes, structures, formats, droits ...



Nécessité d'un soutien au chercheur : documentaliste

Sensibilisation des chercheurs : animation

- *metadata parties*
- présentation dans les laboratoires
- intervention dans les écoles doctorales



Collecte des métadonnées

Qualité des métadonnées

Référentiels : mise en place et maintenance

Données saisies : *curation* et validation (documentaliste spécialiste IST) - veiller à ce que les métadonnées soient :

- complètes
- correctes
- homogènes



Conclusion

Etat du projet :

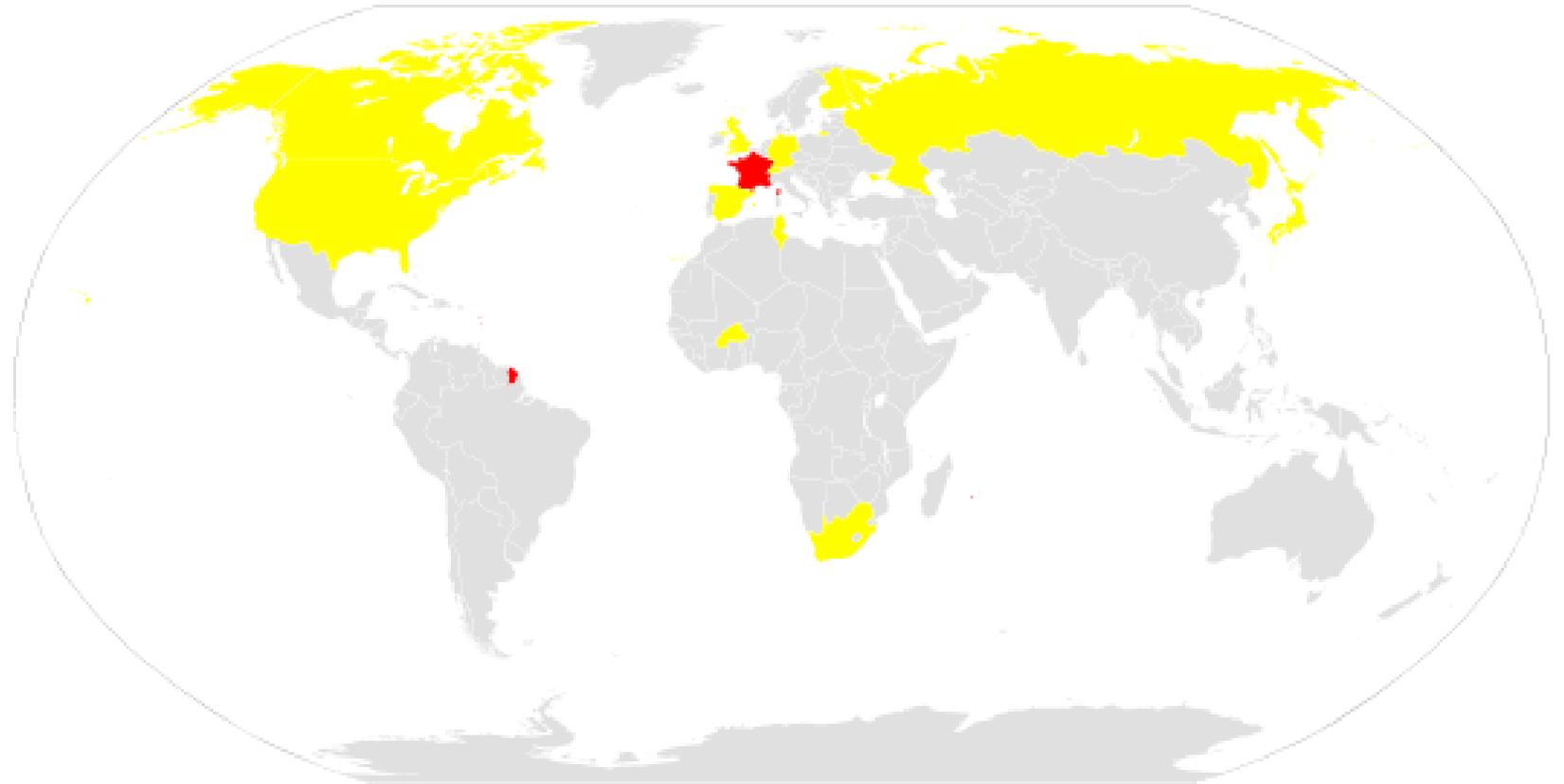
- Interfaces de consultation et de saisie finalisées (ouverture le 26 avril 2016)
- Planification de rencontres avec les chercheurs et de présentation dans les laboratoires
- Poste de documentaliste attribué au projet par le CNRS-INSU (1^{er} juillet 2017)
- 332 fiches complètes : 258 contributeurs, 32 structures, 13 pays

Contributeurs

- 126 personnes en tant que créateurs
- 259 personnes ont contribué
- 8 structures en tant que créateurs
- 20 éditeurs distincts
- 32 structures ont contribué
- 13 pays avec des contributions
- 375 publications

Pays les plus représentés :

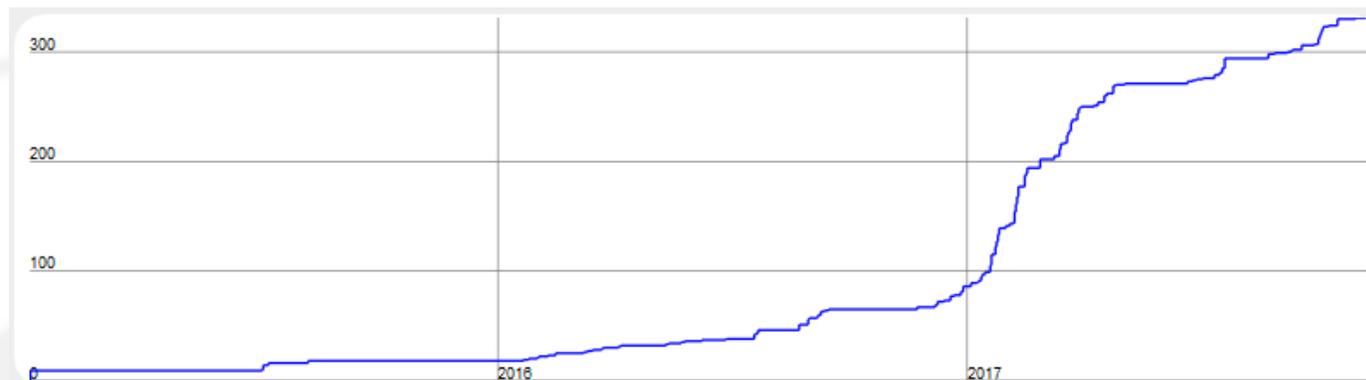
France	332	100%
Allemagne	7	2%
Suisse	3	1%
Canada	2	1%
États-Unis	2	1%
Russie	2	1%
Tunisie	2	1%
Afrique du Sud	1	0%
Burkina Faso	1	0%
Espagne	1	0%





Conclusion

Nombre de fiches
De métadonnées



État des données		
Nombre total	332	100%
Données ouvertes	35	11%
Données en accès restreint	297	89%
Données obsolètes	1	0%

Type de données		
Jeu de données	305	92%
Base de données	27	8%

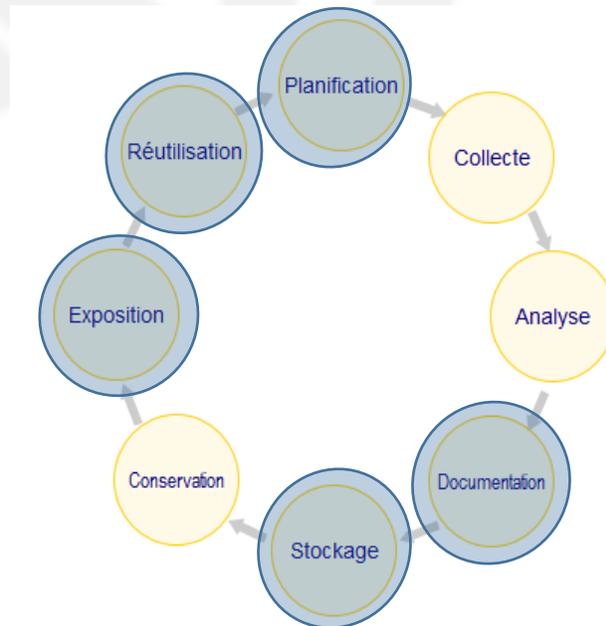
Disciplines représentées		
sciences de l'environnement	268	81%
paléontologie	252	76%
archéologie	205	62%
géographie physique	39	12%
écologie	28	8%
limnologie	28	8%
géologie	25	8%
géosciences	23	7%
biodiversité, conservation	14	4%
crystallographie	13	4%

Labels représentées		
ZAAJ	107	32%
ZABR	7	2%
PNA Milan royal	1	0%
RBV	1	0%
SNO Karst	1	0%
SNOT	1	0%
ZAA	1	0%
ZATA	1	0%

Types de financement		
Région Franche-Comté	12	4%
Projet Européen	11	3%
Gouvernement étranger	6	2%
Agence de l'eau	5	2%
CNRS	4	1%
ANR	3	1%
Contrat industriel	3	1%
OSU	2	1%
Agence nationale	1	0%
Autre établissement public	1	0%

Structures représentées		
Chrono-environnement	289	87%
Biogéosciences	33	10%
UTINAM	24	7%
SRA Franche-Comté	19	6%
ChrEco	18	5%
INRAP	16	5%
SRA Lorraine	16	5%
ArScAn	12	4%
IUCr	11	3%
AORoc	10	3%

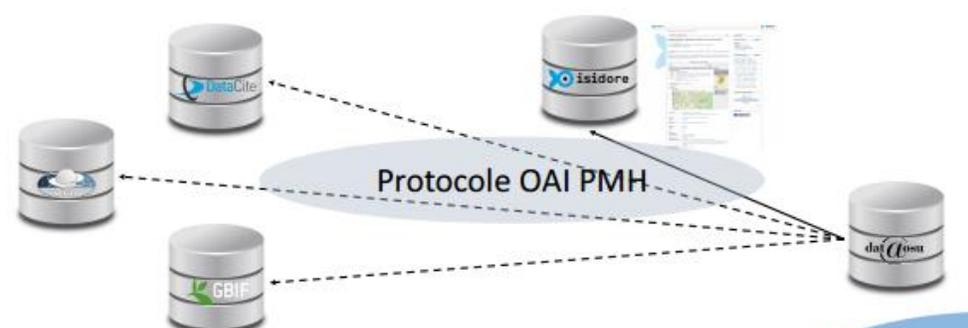
- Export des métadonnées (interopérabilité) – OAI-PMH
 - Datacite
 - *Registries* de l'Observatoire Virtuel
 - Zones Ateliers, GBIF
 - Pôle atmosphère
- Gestion des données de la recherche : prochaines étapes
 - attribution d'un identifiant pérenne (DOI)
 - plan de gestion des données (DMP)
- Diffusion et/ou passage à l'échelle de Dat@OSU



Dat@OSU : une plateforme de référencement et de valorisation des données de la recherche à l'OSU THETA

Sylvie DAMY (Chrono-environnement)
 Bernard DEBRAY (UTINAM)
 Gaëlle LAPORTE (OSU THETA)
 Raphaël MELIOR (OSU THETA)
 Hélène TISSERAND (OSU THETA)

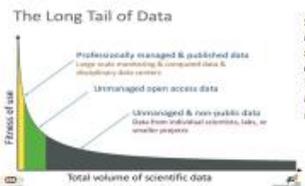
La plateforme Dat@OSU propose une description homogène des données numériques issues des recherches scientifiques menées dans les différents laboratoires et équipes de l'OSU THETA et ce quelque soient les disciplines traitées.



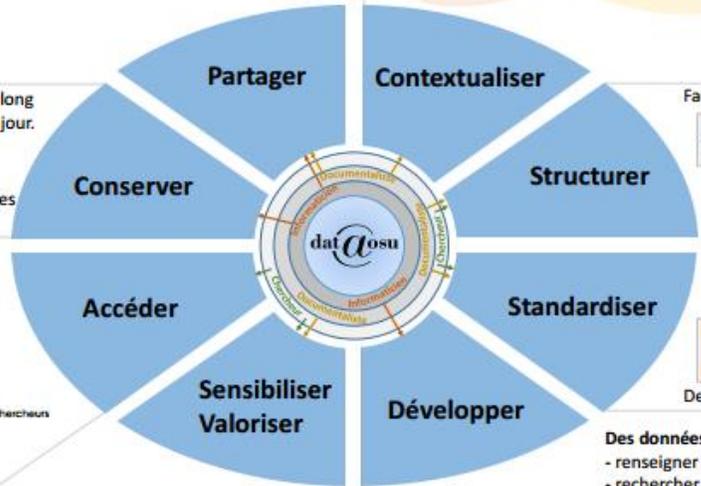
L'interopérabilité de Dat@OSU permet la ré-exposition des métadonnées sur d'autres portails.



La conservation et la diffusion des données peuvent être très utiles à la communauté scientifique.



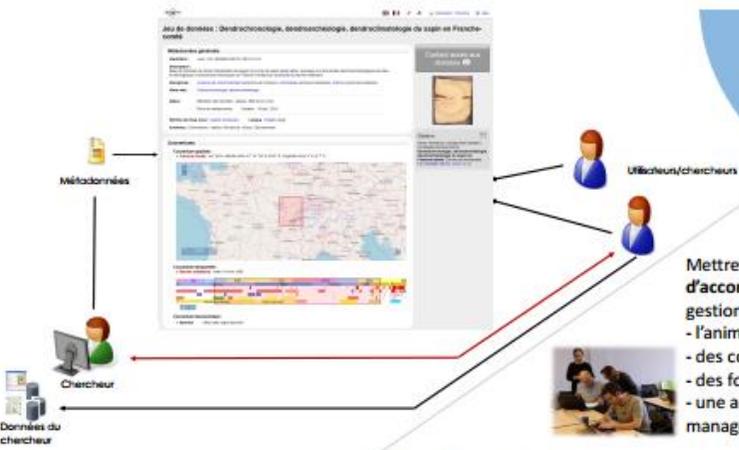
90% des données de la recherche seraient stockées sur les disques durs locaux donc non réutilisables (de Waard A. Making Research Data Discoverable and Usable, 2013)



Faciliter la recherche, l'usage et l'exploitation des données par l'usage de métadonnées.



Définition d'un profil de métadonnées.



Mettre en place des dispositifs d'accompagnement au chercheur dans la gestion et la description de ses données, via :

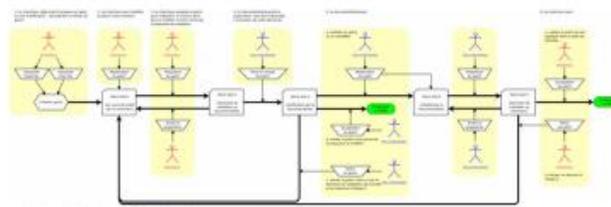
- les animations d'atelier « metadata party »
- des conseils individuels
- des formations pour l'école doctorale
- une aide ponctuelle à la rédaction de data management plan (DMP)

Valoriser le portail au travers d'un catalogue des services français dédiés aux données scientifiques.



Des données à portée de clic. Une interface et des outils permettant de :

- renseigner des fiches de métadonnées
- rechercher des fiches
- consulter des fiches



Data

Le poster



Merci à tous ceux qui nous ont soutenus