###### Projet de valorisation des données de décomposition des litières végétales dans les sols

**Table des matières**

[Introduction générale 4](#_Toc11355359)

[Chapitre 1 : Décomposition des litières végétales 5](#_Toc11355360)

[Introduction 5](#_Toc11355361)

[I. Fonctionnement du processus de décomposition 7](#_Toc11355362)

[1. Différents stades de la transformation de la matière organique 7](#_Toc11355363)

[2. Couplage des cycles du carbone et de l’azote 9](#_Toc11355364)

[II. Facteurs contrôlant la décomposition 11](#_Toc11355365)

[1. Biodégradabilité à différentes échelles 11](#_Toc11355366)

[a. Echelle des polymères 12](#_Toc11355367)

[b. Echelle cellulaire 14](#_Toc11355368)

[c. Echelle tissulaire 14](#_Toc11355369)

[d. Echelle des organes 15](#_Toc11355370)

[3. Ecologie et stratégie de vie des microorganismes du sol 16](#_Toc11355371)

[e. Fonctionnement général des microorganismes du sol 16](#_Toc11355372)

[f. Stratégie de vie des microorganismes 17](#_Toc11355373)

[Oligotrophes/Copiotrophes 17](#_Toc11355374)

[Modèle CSR 19](#_Toc11355375)

[Modèle GDM (Guild-based Decomposition Model) 20](#_Toc11355376)

[Biomasse zymogène/autochtone 20](#_Toc11355377)

[III. Influence de l’environnement 21](#_Toc11355378)

[1. Interactions plante-sol 21](#_Toc11355379)

[2. Conditions pédoclimatiques 21](#_Toc11355380)

[3. Pratiques culturales 23](#_Toc11355381)

[IV. Caractérisation analytique de la décomposition des résidus 24](#_Toc11355382)

[1. Caractérisation analytique des résidus 24](#_Toc11355383)

[a. Caractérisation C, N, P et S 24](#_Toc11355384)

[b. Fractionnement biochimique par la méthode Van Soest 25](#_Toc11355385)

[g. Sucres 26](#_Toc11355386)

[h. Phénols 26](#_Toc11355387)

[4. Cinétiques de minéralisation en conditions potentielles 27](#_Toc11355388)

[a. Carbone et azote 27](#_Toc11355389)

[b. Phosphore et soufre 29](#_Toc11355390)

[2. Cas particulier du marquage isotopique 29](#_Toc11355391)

[Conclusion 30](#_Toc11355392)

[Chapitre 2 : Gestion et partage de données scientifiques 31](#_Toc11355393)

[Introduction 31](#_Toc11355394)

[I. Les enjeux de l’accessibilité aux données scientifiques 33](#_Toc11355395)

[1. Enjeux patrimoniaux 33](#_Toc11355396)

[2. Enjeux économiques 33](#_Toc11355397)

[3. Enjeux scientifiques 34](#_Toc11355398)

[4. Enjeux sociétaux 34](#_Toc11355399)

[5. Enjeux éthiques 35](#_Toc11355400)

[II. Produire des données FAIR 35](#_Toc11355401)

[1. Findable 35](#_Toc11355402)

[2. Accessible 36](#_Toc11355403)

[3. Interoperable 36](#_Toc11355404)

[4. Reusable 36](#_Toc11355405)

[III. Gérer des données 37](#_Toc11355406)

[1. Etablir un plan de gestion des données (PGD) 37](#_Toc11355407)

[Documenter et organiser les données 39](#_Toc11355408)

[2. Nommage et organisation des fichiers et dossiers 39](#_Toc11355409)

[3. Documenter les données 40](#_Toc11355410)

[4. Identifier les données 42](#_Toc11355411)

[Identifiant pérenne objet (Figure 10) : 43](#_Toc11355412)

[Identifiant pérenne contributeur : 44](#_Toc11355413)

[IV. Partager et archiver des données 44](#_Toc11355414)

[1. Format des fichiers 45](#_Toc11355415)

[2. Choix de la licence 48](#_Toc11355416)

[Les licences CC BY : 48](#_Toc11355417)

[La licence CC 0 : 49](#_Toc11355418)

[3. Entrepôt de données et plateforme d’archivage 49](#_Toc11355419)

[V. Valoriser les données 53](#_Toc11355420)

[1. Data paper 53](#_Toc11355421)

[2. Web des données 55](#_Toc11355422)

[3. Application à l’agronomie : la modélisation 56](#_Toc11355423)

[VI. Etudes de cas à partir de data paper déjà publiés 56](#_Toc11355424)

[1. A global experimental dataset for assessing grain legume production, Cernay, Pelzer and Makowski, 2016 56](#_Toc11355425)

[a. Généralités sur le journal 57](#_Toc11355426)

[b. Contenu du data paper 58](#_Toc11355427)

[2. Data on spatio-temporal representation of mineral N fertilization and manure N application as well as ammonia volatilization in French regions for the crop year 2005/06, Génermont et al., 2018 59](#_Toc11355428)

[a. Généralités sur le journal 59](#_Toc11355429)

[b. Contenu 61](#_Toc11355430)

[3. Fifteen-year record of soil temperature at the Bear Brook Watershed in Maine, Patel et al., 2018 62](#_Toc11355431)

[Conclusion 64](#_Toc11355432)

[Perspectives 65](#_Toc11355433)

[Annexes 66](#_Toc11355434)

[Bibliographie 70](#_Toc11355435)

##### Introduction générale

Pour commencer, ce projet a été mis en place pour un apprenti ingénieur et est donc prévu pour une durée de 3 ans.   
Plusieurs chercheurs travaillant sur les problématiques liées à la décomposition des litières végétales partant bientôt ou déjà partis en retraite, la question de la sauvegarde de leurs données a rapidement vu le jour. Toutefois, pour la plupart des cas, aucun dispositif n’a été mis en place afin de récupérer et stocker les données acquises tout au long de leur carrière.   
Pourtant, depuis peu, l’INRA s’est orienté vers les sciences ouvertes afin de rendre accessible à tous les résultats de la recherche. De par cette initiative, de nouvelles méthodes sont disponibles afin de permettre le partage mais aussi le stockage de données de plus en plus importantes. Bien que les entrepôts de données, solution proposée par l’INRA à ce sujet, ne permettent pas une sauvegarde à long terme, ceux-ci garantissent tout de même une accessibilité à moyen terme pour environ une vingtaine d’années. Les sauvegardes à long terme sont bien sûres possibles mais onéreuses car elles nécessitent de puissants moyens de stockage ainsi que du personnel afin de vérifier et de mettre à jour au fil des années le format des fichiers qui leurs sont confiés.

D’autre part, l’enjeu d’un tel projet est également de compiler les données récoltées afin de créer, dans l’idéal, une base de données accessible par tous. Cette compilation ne peut qu’apporter de la valeur aux données et faciliter leur réutilisation. Dans un contexte où les financements de l’Etat sont de plus en plus rares, la réutilisation de données déjà acquises par un tiers pour effectuer ses recherches parait être une bonne alternative économique.

Bien qu’elles ne soient pas d’intérêt public, ces données sont importantes pour le devenir de notre planète dans le contexte actuel du changement climatique. Une meilleure connaissance de la décomposition des litières végétales pourrait en effet permettre de réduire les effets liés au réchauffement climatique mais également d’améliorer la qualité des sols, détériorés jusqu’alors par les activités humaines.

Comment valoriser des données de décomposition des litières végétales dans les sols ?

Pour cela nous étudierons d’abord la décomposition des litières dans un premier chapitre puis nous verrons la valorisation des données dans un second chapitre.

##### Chapitre 1 : Décomposition des litières végétales

##### Introduction

En 2016, les terres agricoles représentaient plus d’un tiers de la surface totale du territoire mondial (Banque mondiale, 2016). Ces terres sont utilisées pour la culture et l’élevage afin de nourrir la population mondiale. En effet la biomasse végétale, c’est-à-dire la masse de matière vivante (végétale) (dictionnaire environnement et développement durable, 2010), produite sur nos sols est aujourd’hui majoritairement exportée que ce soit pour l’alimentation humaine ou celle des animaux d’élevages. Les exports de biomasse végétale servent également à couvrir les besoins de l’industrie chimique ou encore de l’industrie des matériaux. De plus, face aux nouveaux challenges du réchauffement climatiques, les biomasses végétales sont également exportées pour la création de bioénergies. Toutefois, ces exportations ne sont pas sans conséquences sur la santé de nos sols qui mettent plusieurs milliers d’années à se constituer alors que leurs dégradation est très rapide. En effet, on estime à 40% la surface de sols cultivés déjà dégradée par les activités humaines dans le monde (ADEME, 2018). Or avec l’augmentation de la population mondiale et des revenus moyens, la préservation des sols devient centrale afin de pouvoir nourrir les habitants du globe.

Afin de limiter leur dégradation, il est indispensable de recycler une partie de la biomasse végétale produite pour que celle-ci alimente la litière du sol. En effet, le recyclage de la matière organique (matière issue des organismes vivants) permet tout d’abord la conservation de la biodiversité des sols en stimulant la faune et la flore qui rendent de nombreux services écosystémiques de par leur activité et leur diversité. Ces organismes assurent la structuration physique des sols et donc diminue les risques d’érosion. De plus, la diversité des organismes vivants des sols permet également d’assurer une régulation contre des organismes nocifs ou pathogènes. La protection des cultures s’exerce donc naturellement avec des besoins diminués en produits phytosanitaires. D’autre part, en cas de pollution, certains organismes sont capables de neutraliser les molécules toxiques et d’ainsi dépolluer leur environnement.  
Enfin, faune et la flore du sol jouent un rôle essentiel dans la fertilisation. En effet, les microorganismes sont capables de minéraliser la matière organique qui servira ensuite à la nutrition des plantes ce qui permet de boucler les cycles biogéochimiques (Agronutrition, 2014), (Lashermes, 2017a). D’autre part, l’apport de litière permet d’alimenter les compartiments labiles (temps de renouvellement allant de quelques jours à un an) et intermédiaires (temps de renouvellement allant de quelques années à quelques décennies) du carbone du sol mais également indirectement le compartiment stable, aussi appelée humus dont le temps de renouvellement peut aller de quelques décennies à quelques siècles (Derrien et al., 2016). En effet, les matières organiques dites « exogènes » (ou « fraîches ») constituées de résidus de récolte, d’effluents d’élevage ou encore d’exsudats racinaires alimentent les compartiments labiles et intermédiaires. Ces derniers alimentent le compartiment stable qui est constitué de matières organiques « endogènes » issue de la fraction non vivante des matières organiques (Annexe 1). C’est notamment grâce à ce réservoir stable que le stockage du carbone dans les sols est possible car la matière organique qu’il contient est peu dégradée ensuite par les microorganismes. Or, le stockage du carbone est une problématique d’actualité, notamment avec l’initiative 4 pour 1000. En effet, l’augmentation des stocks de carbone de 0.4% par an permettrait de réduire les effets du changement climatique grâce à la baisse du taux de dioxyde de carbone dans l’atmosphère (CGIAR System Organization, 2018).

La bonne gestion de la décomposition des résidus de culture étant donc nécessaire pour préserver nos écosystèmes, il est indispensable de savoir quels sont les mécanismes mis en jeu et comment les caractériser.

Pour cela nous commencerons par étudier le fonctionnement du processus de décomposition des litières végétales, les facteurs de contrôle de ce processus et enfin sa caractérisation analytique.

# Fonctionnement du processus de décomposition

## Différents stades de la transformation de la matière organique

Les matières organiques végétales se transforment au cours du temps selon différents stades : décomposition, minéralisation et enfin formation de molécules stables.   
La décomposition est possible grâce à l’action de la macrofaune et mégafaune (organismes supérieurs à 2mm) qui enfouissent et fragmentent la matière végétale, facilitant alors l’action de la microfaune et de la microflore (Figure 1). La microflore comprend les champignons qui décomposent plutôt des composés récalcitrants tels que la lignine alors que la microflore comprends les bactéries qui elles, décomposent plutôt les composés solubles (Trinsoutrot, 1999). Toutefois, il a été admis plus tard que les bactéries sont elles aussi capables de dégrader la matière organique récalcitrante (Bugg et al., 2011).  
Les microorganismes dégradent la matière organique végétale fraîche grâce à des actions enzymatiques. Les produits alors formés sont des composés organiques tels que par exemple du glucose, des acides gras, des acides aminés,… Les composés organiques issus de la décomposition sont ensuite recombinés grâce à d’autres produits microbiens agissant comme liants ce qui permet l’obtention de polymères stables. La source de production majeure des produits microbiens est la matière organique labile qui est le compartiment le plus efficacement utilisé par les microorganismes. La mort de ces derniers constitue également une source de matière organique fraîche qui est ensuite dégradée par les microorganismes vivants. La matière organique obtenue après décomposition est appelée matière organique humifiée et est liée à la matrice du sol grâce aux produits microbiens qui permettent la formation d’agrégats. Ce sont ces interactions entre matrice du sol et matière organique (interaction organo-minérales) qui permettent la stabilisation de la matière organique. D’autre part, la macrofaune, de par la formation de biostructures, est elle aussi responsable d’interactions organo-minérales et donc de la stabilisation du carbone (Derrien et al., 2016). Enfin, les microbes présents dans le tube digestif de la macrofaune favorisent également la minéralisation du carbone car ils sont directement en contact avec la matière organique. La digestion étant source d’altération de la matière organique, la minéralisation est d’autant plus simple pour les microorganismes (Derrien et al., 2016).

La micro et macroflore sont donc en relation étroite dans le processus de décomposition des matières organiques. Ces relations sont d’autant plus importantes que le sol est pauvre en matières organiques (Trinsoutrot, 1999). En outre, la régulation des microorganismes dépend de l’ensemble de la faune du sol : certains bactérivores entrainent la diminution de la population de microorganismes mais les rendent toutefois plus actifs. C’est donc tout le réseau trophique qui entre en jeu dans les mécanismes de dégradation des matières organiques (Derrien et al., 2016).   
La dégradation complète de la matière organique et des microorganismes du sol conduit à la formation de composés minéraux (tel que le CO2 issu de la respiration) : c’est la minéralisation.   
Ces matières minérales telles que l’azote ou le phosphore sont ensuite réutilisées par les plantes pour leur croissance. La minéralisation est donc source de fertilisation et le processus de « priming effect » augmente la disponibilité des nutriments pour la plante. Ce processus de surminéralisation a lieu lors de l’ajout de matière organique fraiche dans le sol. En effet, cet ajout stimule les microorganismes et leur apporte l’énergie nécessaire pour dégrader la matière organique humifiée. Cependant, ce processus conduit à une diminution des stocks de carbone (Quae 2017, chap 3).

Figure : Principales actions de la faune du sol sur la décomposition de la litière, (Sauvadet, 2016).

**Faune microbivore**

(Microfaune : protozoaires & nématodes)

**Microorganismes**

(bactéries & champignons)

**Prédation**

**Dégradation**

**Litière**

**chimique**

**Macro & Mégafaune**

(arthropodes, enchytréides, vers de terre,…)

**Dégradation**

**physique**

*Stimulation*

*activité microbienne*

*Production des enzymes*

*de dégradation chimique*

*Fragmentation physique de la litière*

*Stimulation des microorganismes*

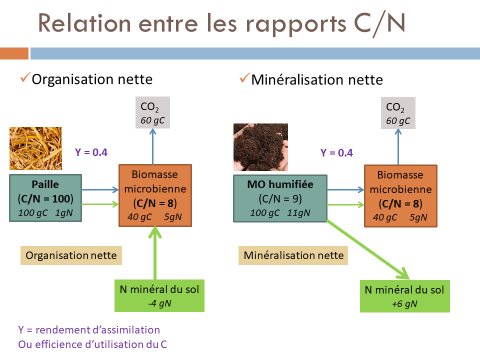
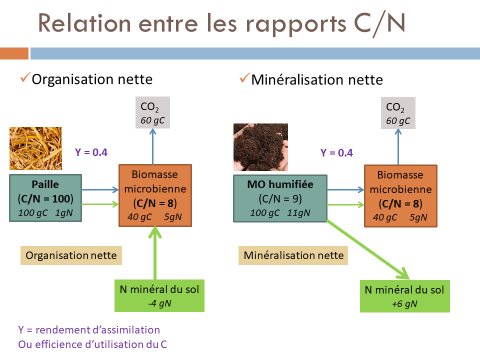
**Mise en contact microorganismes - litière**

**Modification de l’environnement microbien**

## Couplage des cycles du carbone et de l’azote

A travers ce processus complexe de décomposition de la matière organique, deux cycles d’éléments chimiques ont été principalement étudiés : celui de carbone et de l’azote afin de mieux comprendre leurs stockage/déstockage.  
Le stockage/déstockage d’azote minéral correspond aux phénomènes d’organisation et de minéralisation. On parle de minéralisation lorsque la dégradation de la matière organique par la biomasse microbienne entraine la production d’azote minéral, disponible ensuite dans le sol. Au contraire, on parle d’immobilisation lorsque les microorganismes prélèvent de l’azote minéral dans le sol afin d’avoir l’énergie nécessaire pour initier la décomposition.   
Les processus d’organisation/minéralisation dépendent de la teneur en carbone et en azote du résidu : plus la quantité de carbone apportée est importante et plus la biomasse microbienne a besoin d’azote pour l’assimiler. Les cycles du carbone et de l’azote ne sont donc pas interdépendants. Il a été estimé que pour un ratio C/N caractérisant le résidu inférieur à 24 une minéralisation nette du carbone était observée alors que c’est le contraire (organisation nette) pour un rapport C/N supérieur à 24 (Trinsoutrot, 1999) (Figure 2). Toutefois ces processus ne dépendent pas seulement de la composition des résidus mais aussi des besoins des communautés microbiennes. Ce dernier point sera développé plus en détail dans la seconde partie.

Figure : Phénomènes d'organisation et de minéralisation nette selon le ratio C/N du résidu, (Lashermes, 2017b).



Cependant, bien que l’organisation (aussi appelée immobilisation) et la minéralisation dépendent de la teneur en azote du résidu, aucun effet n’a été observé à long terme à partir de ce seul paramètre. En revanche la disponibilité globale de l’azote durant la décomposition aurait un effet. La décomposition de la matière organique se réalise effectivement en 2 étapes : la dégradation des composés solubles qui intervient au début du processus et qui nécessite une quantité importante d’azote ainsi que la dégradation des composés plus récalcitrants qui intervient plus tard dans le processus et qui nécessite moins d’azote (Cotrufo et al., 2013), (Recous, 1995).  
Toutefois, en conditions limitantes en N, on observe une plus faible croissance des microorganismes car ceux-ci n’ont pas assez de nutriments pour permettre leur développement (Cotrufo et al., 2013). Cela entraine donc une plus faible quantité de carbone assimilée ainsi qu’une baisse de la respiration et donc une baisse de la minéralisation du carbone. De plus, dans ce cas-là ; l’azote est immobilisé directement après avoir été minéralisé, il n’y a donc pas d’accumulation (Recous, 1995). En revanche, on observe une diminution de la quantité d’azote organique au cours de la décomposition. Il semblerait donc que l’azote soit ensuite reminéralisé (Trinsoutrot, 1999).  
Cependant si le résidu a un rapport C/N plutôt bas permettant de couvrir les besoins en azote des microorganismes, la teneur en azote du sol n’influera pas sur la décomposition. Dans ce cas, l’activité bactérienne se verra stimulée ce qui va accélérer la décomposition des résidus et donc améliorer le stockage du carbone dans le sol (Derrien et al., 2016). L’augmentation de l’activité microbienne va aussi conduire à une importante minéralisation nette, notamment au début de la décomposition, durant les 2 premiers mois. Dans ce cas-là, l’essentiel de l’azote présent dans le résidu aura été assimilé par les microorganismes durant la décomposition et/ou humifié (Trinsoutrot, 1999).

L’azote minéral produit par les décomposeurs est sous forme d’ammonium (NH4+). Cet ammonium est ensuite transformé par des bactéries nitrifiantes pour donner des ions nitrates (NO3-). C’est cette dernière forme de l’azote qui est préférentiellement assimilée par les plantes. Cela explique pourquoi NH4+ est préférentiellement immobilisé par rapport à NO3-.



Figure 3 : couplage des cycles du carbone et de l'azote, (Amin, 2012)

# Facteurs contrôlant la décomposition

## Biodégradabilité à différentes échelles

Les cinétiques de décomposition varient selon la composition moléculaire et tissulaire ainsi que selon les organes de la plante concernés. Ces facteurs dépendent de la famille botanique (Recous et al., 2017) et plus précisément de l’espèce étudiée (Bertrand, 2013), de son stade de maturité (Recous et al., 2017), (Bertrand, 2013), de ses conditions de croissance (Bertrand, 2013) et des organes considérés (Bertrand, 2013), (Recous et al., 2017).  
Ainsi, on peut distinguer 2 grands groupes : les solubles qui sont des composés cytoplasmiques et les composés pariétaux qui correspondent principalement à la cellulose, à l’hémicellulose et à la lignine. La cellulose est un polymère de glucoses et l’hémicellulose un polymère d’arabinoses et de xyloses qui vient entourer la cellulose. Enfin la lignine est un polymère aromatique constitué de plusieurs composés phénoliques (National Science Fondation, University of Toledo, 2013) : on parle d’unités G, S et H (Figure 4). En moyenne, une plante contient 40 à 60% de cellulose, 20 à 40% d’hémicellulose et 10 à 25% de lignine (Jean-Luc WERTZ, 2010). Cette dernière vient s’enchevêtrer avec la cellulose et l’hémicellulose (Bertrand, 2013) ce qui permet la protection de ces polysaccharides (Trinsoutrot, 1999). Les tissus étant un groupement de cellules végétales de même origine embryonnaire, ces trois types de polymères se retrouvent au niveau tissulaire dans le xylème et le parenchyme par exemple.

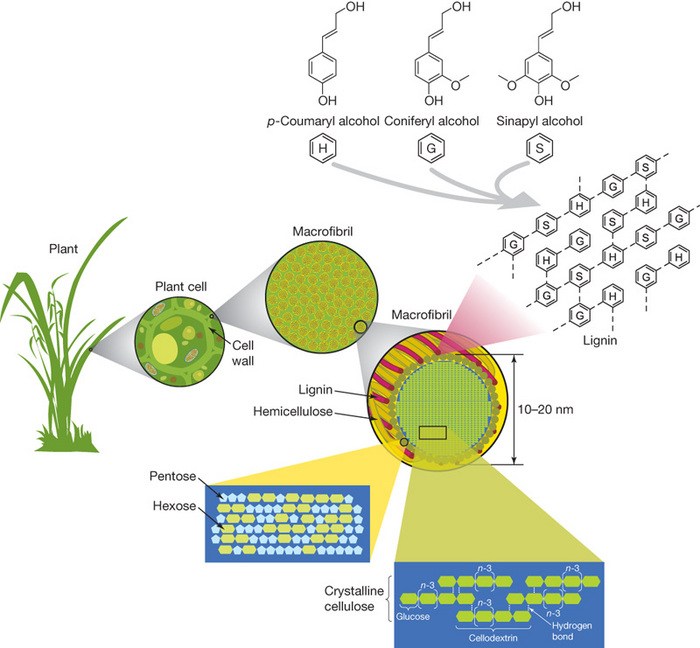


Figure 4 : Représentation de la structure de la plante, (Rubin, 2008)

### Echelle des polymères

Les sucres comme le glucose, l’arabinose et le xylose présentent une forte valeur énergétique pour les microorganismes alors que les composés phénoliques constituants la lignine sont pauvres en énergie et difficile à dégrader. Toutefois, la vitesse et la facilité de la dégradation dépendent également d’autres facteurs tels que la composition de l’hémicellulose par exemple. Ce polymère comprend une chaîne linéaire de xyloses sur lesquels viennent se ramifier des arabinoses. On appelle donc cette molécule arabinoxyloses. Le degré de ramification des arabinoses, représenté par le ratio xylose/arabinose (Machinet et al., 2009), joue un rôle dans les cinétiques de décomposition. En effet, plus les arabinoses seront ramifiés et plus ils seront difficilement dégradables (Bertrand, 2013). Les microorganismes commencent donc par dégrader les chaînes les moins ramifiées ce qui provoque l’augmentation du degré de substitution des arabinoxylanes au fur et à mesure de la dégradation enzymatique (Moorhead et al., 2014), (Machinet et al., 2009). Au cours du temps, les microorganismes ont donc de plus en plus de mal à dégrader la matière organique.

De par sa composition et sa structure la lignine est un composé difficilement dégradable. Ainsi, l’enchevêtrement avec les autres composés pariétaux la rend difficile d’accès et les microorganismes ont besoin d’utiliser de nombreuses enzymes différentes pour la dégrader (Trinsoutrot, 1999). De ce fait, cette dégradation leur est énergétiquement couteuse et les composés qu’ils en récupèrent sont pauvres en énergie.  
De plus, la ligne peut être sous deux états : condensée et non condensée. Les lignines condensées contiennent un nombre important d’unités G qui présentent des liaisons C-C difficilement dégradables contrairement aux unités S (Bertrand, 2013). De ce fait, elle sera moins accessible aux microorganismes et le degré de condensation va jouer sur sa vitesse de décomposition.  
D’autre part, la teneur de la lignine en polyphénols solubles aurait aussi un impact sur la minéralisation de l’azote à court terme : les polyphénols seraient capables de former des complexes avec les protéines ce qui diminueraient là encore l’accessibilité de la lignine et donc retarderait la minéralisation (Trinsoutrot, 1999).  
Cependant, certaines études récentes montrent que la récalcitrance de la lignine ne s’expliquerait pas par la matière elle-même mais plutôt par les associations existantes au cours de la décomposition entre lignine et produits microbiens (liaisons avec des cations métalliques) (Cotrufo et al., 2013).

En outre, les acides féruliques sont également des composés pariétaux impactant la décomposition de la matière organique. En effet, ces molécules sont des composés phénoliques capables de se lier aux polysaccharides par des liaisons esthers (Kroon et al., 2000),(Mastihuba et al., 2002). Ils servent alors d’amorce pour la polymérisation de la lignine (Lapierre, 2013). De plus ils sont aussi capable de se lier à la lignine par des liaisons éthers (Kheder, 2007) générant ainsi une nouvelle structure de branchement permettant de faciliter sa dépolymérisation (Lapierre, 2013).  
Les liaisons éthers et esters sont des liaisons covalentes difficilement dégradables qui nécessitent la synthèse d’hydrolases (Lapierre, 2013). De ce fait les acides féruliques contribuent à la diminution de la biodégradabilité de la paroi (Mastihuba et al., 2002) et pourrait également réduire la minéralisation à long terme (Bertrand, 2013).

Toutefois, selon Cotrufo, la récalcitrance chimique des résidus explique la décomposition de la matière organique sur le court et moyen terme mais non sur le long terme (Cotrufo et al., 2013)

La prise en compte des interactions entre les composés pariétaux semble essentielle afin de déterminer les cinétiques de décomposition de la matière organique. Ainsi les rapports Lignine/Glucose, Lignine/Arabinoxylane et plus généralement Lignine/Sucres semblent être plus pertinents que seule la teneur en lignine (Bertrand, 2013), (Machinet et al., 2009). D’autre part, ces rapports ne sont pas influencés par les pertes de matières sèches (Machinet et al., 2009). De plus les liaisons éthers et esters pourraient être de bons biomarqueurs de la stabilisation du carbone (Bertrand, 2013).

### Echelle cellulaire

Comme nous avons pu le voir dans la partie précédente, la lignine est un composé difficilement dégradable par les microorganismes. Les parois faiblement lignifiées sont donc préférentiellement dégradées par les microorganismes sur court terme (120j) (Bertrand, 2013). En effet, dans le cas où les parois sont très lignifiées la plupart du carbone issu de la litière est transformé sous forme de CO2 et non de métabolites microbiens ce qui ne permet pas la stabilisation du carbone (Cotrufo et al., 2013).   
Au contraire, la fraction soluble, est considérée comme la fraction la plus efficacement décomposable car elle contient des composés hydrophiles eux-mêmes facilement dégradables. Ainsi, un grand nombre de métabolites sont synthétisés à partir de cette fraction, permettant la stabilisation de la matière organique (Moorhead et al., 2014), (Cotrufo et al., 2015). Cependant les solubles sont aussi constitués de composés plus hydrophobes qui ne sont pas dégradés préférentiellement pas les microorganismes (Moorhead et al., 2014).  
La taille de la fraction soluble influencerait toutefois la minéralisation du carbone durant les premiers stades de la décomposition mais ce facteur n’a pas d’influence sur la minéralisation à long terme (Bertrand, 2013).

### Echelle tissulaire

La cohésion des cellules entre elles et même plus généralement la manière dont elles sont organisées dépendent du type tissulaire. Cela peut laisser penser que le type de tissu influe sur la décomposition des matières organiques. De plus nous verrons dans la partie suivante que la décomposition des résidus de culture est également régie par le type d’organe concerné. Or bien que la part de composés pariétaux varie selon les types d’organes, la fonctionnalité des tissus varie elle aussi et pourrait expliquer, au moins en partie, les différences de décomposition entre divers organes. Plus précisément, selon de rares études, le fait qu’un tissu soit conducteur ou non influerait sur le processus de décomposition (Bertrand, 2013). La densité des tissus impacterait également la décomposition des tiges (Freschet et al., 2012).

### Echelle des organes

La stabilisation de matière organique est plus importante lorsque celle-ci provient de résidus racinaires plutôt que de parties aériennes de la plante (Chenu, 2013), (Derrien et al., 2016). Cela est notamment vrai pour les monocotylédones qui ont une importante biomasse racinaire ainsi qu’une densité élevée de racines fines (Derrien et al., 2016). En effet, la densité de racines fines est corrélée à la quantité de litière et d’exsudats racinaires ainsi qu’à la « stabilisation des agrégats du sol ». Toutefois, bien que le colza soit une dicotylédone, la minéralisation du carbone des racines est inférieur à celle des tiges et des feuilles (Trinsoutrot, 1999). On peut expliquer ce phénomène par la plus forte proximité physique avec la zone d’interface résidus-microorganismes du sol mais aussi par la qualité chimique des différents organes (Chenu, 2013).  
Par exemple, la part de la fraction soluble qui varie selon les organes peut être un des facteurs explicatifs de cette différence de minéralisation. Celle-ci correspond à environ 20% en masse pour les racines alors qu’elle atteint les 50% pour les organes aériens (tiges et feuilles) (Bertrand, 2013). On observera donc une minéralisation plus importante durant les premiers stades de décomposition pour les tiges et les feuilles que pour les racines.   
Par ailleurs, la décomposition des feuilles enrichies en paroi cellulaire primaire est régulée par le degré de substitution des arabinoses alors que dans les racines les arabinoses sont liés à la lignine. Encore une fois, on observera donc une décomposition des polysaccharides moins importante dans les racines que dans les feuilles. Le fait que la décomposition soit plus lente dans les racines contribue à un meilleur stockage du carbone.  
La minéralisation brute des racines est toujours la même quelque soit la teneur en azote de celles-ci alors que ce n’est pas le cas pour les tiges et les siliques de colza : plus celle-ci contiennent de l’azote et plus la minéralisation brute est importante.  
On observe le phénomène inverse concernant l’organisation brute : celle-ci ne dépend pas de la teneur en azote pour les tiges alors qu’elle est plus importante pour les racines à faible teneur en azote (Trinsoutrot, 1999).

La longueur des racines va également impacter la vitesse de décomposition du carbone : les sols comportant des plantes à longues racines sont plus riches en matière organique. De plus, le phénomène de décomposition est plus long en profondeur ce qui laisse penser que le carbone est plus facilement stocké (Derrien et al., 2016).

## Ecologie et stratégie de vie des microorganismes du sol

Le terme générique « microorganismes du sol » englobe aussi bien les bactéries que les champignons. Toutefois ces organismes ne fonctionnent pas de la même manière : ils n’ont pas les mêmes besoins, ne sécrètent pas les mêmes enzymes et donc ne dégradent pas le même type de matière organique. Afin de s’y retrouver des classements basés sur la stratégie de vie des microorganismes ont été mis en place.

### Fonctionnement général des microorganismes du sol

Chaque espèce de microorganismes se caractérise par un rapport stœchiométrique C/N qui traduit leurs besoins. Ce ratio est en moyenne de 5 pour les bactéries alors qu’il est de 10 pour les champignons. En d’autres termes, avec 1g d’azote une bactérie assimilera 5g de carbone alors qu’un champignon en assimilera 10g. Ce ratio permet aussi d’établir la composition des microorganismes : une bactérie par exemple est composée de 5 fois plus de carbone que d’azote contre 10 pour un champignon (Figure 5).   
Comme il a déjà été expliqué brièvement plus haut, la matière organique apportée est elle aussi caractérisée par ce rapport stœchiométrique. La comparaison des ratios C/N caractéristique du résidu et du microorganisme va permettre de déterminer, à l’aide du rendement d’assimilation du microorganisme, s’il y a organisation/immobilisation nette ou au contraire minéralisation nette de carbone ou d’azote.

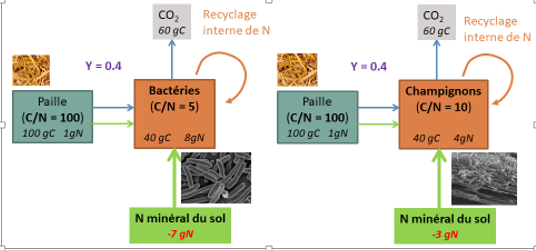


Figure 5 : Comparaison du fonctionnement d'une bactérie et d'un champignon en fonction de leur ratio C/N, (Lashermes, 2017b).

De par leur ratio C/N, les champignons sont plus adaptés aux situations où l’azote est limitant ainsi qu’à la décomposition des composés plus récalcitrants. De plus, ceux-ci peuvent modifier leur ratio en fonction de la disponibilité en azote ainsi que redistribuer les substances nutritives par translocations dans différentes parties du mycélium (Trinsoutrot, 1999 ; Moorhead, Sinsabaugh, 2006). D’autre part, afin de dégrader la matière organique, les microorganismes utilisent deux types d’action : l’hydrolyse et l’oxydation. Les enzymes oxydatives permettent notamment la dégradation de la lignine et sont principalement secrétées par les champignons ce qui explique qu’ils soient capables de dégrader les composés récalcitrants.  
Toutefois, sur le long terme, la disponibilité en azote pourrait inhiber la dégradation de composés lignifiés. Effectivement, une grande quantité d’azote modifierait le mode d’action des microorganismes en bloquant la production d’enzymes oxydatives et donc la dégradation des composés phénoliques contenus dans la lignine (Bertrand, 2013). Ce phénomène dépend de la stratégie de vie des microorganismes qui explique l’évolution des populations microbiennes en fonction de la disponibilité des ressources.

### Stratégie de vie des microorganismes

La diversité écologique et les stratégies de vie des microorganismes sont encore mal connues. Les animaux et plantes sont catégorisés depuis 1967 par leur stratégie de vie qui peut être K ou r. Les stratèges r sont caractérisés par un cycle de vie court, une croissance rapide et une faible compétitivité alors que les stratèges K ont un cycle de vie plus long, une croissance moins rapide et sont plutôt compétitifs. La classification écologique des microorganismes reprend encore aujourd’hui ces grands principes.

#### Oligotrophes/Copiotrophes

Plusieurs scientifiques distinguent deux grands groupes écologiques de microorganismes : les oligotrophes et les copiotrophes. Cette distinction est faite à partir de traits physiologiques : les microorganismes sont différents de par leur cinétique de croissance, leur affinité pour le substrat, l’efficacité qu’ils ont à utiliser les ressources et leurs caractères génomiques. Selon leur catégorie écologique, ils répondront donc différemment aux changements environnementaux (Ho et al., 2017).  
Les copiotrophes sont caractérisés par une constante de Michaelis Menten (Km) ainisi qu’un coefficient de demi-saturation (Ks) (traduisant le taux de croissance par rapport à la nutrition) élevés. Ces microorganismes sont plus sensibles à la baisse de disponibilité d’un substrat et ne sont pas compétitifs lorsque les concentrations en nutriments sont faibles. Toutefois, lorsque leurs besoins nutritifs sont couverts les copiotrophes présentent de forts taux de croissance   
A l’inverse, les oligotrophes sont caractérisés par un Km et un Ks plus faibles. Ils présentent une croissance plus lente mais sont capables de se développer malgré de faibles concentrations en substrat car ils l’utilisent plus efficacement. On ne les retrouve pas dans des milieux riches car ils ne sont pas compétitifs, ils sont surtout présents dans des environnements pauvres en nutriments et sont capables de dégrader la matière organique récalcitrante plus efficacement que les copiotrophes.  
Afin de classer les microorganismes dans un de ses deux groupes, le taux de minéralisation du carbone s’avère être un bon paramètre (Fierer et al., 2007). Lorsque celui-ci est corrélé positivement avec l’abondance d’une communauté donnée, cette dernière sera plutôt de type copiotrophe et vice versa pour une corrélation négative. En effet, le taux de minéralisation dépend de la disponibilité en matière organique dont dépend le type de communautés présentes.  
Les bactéries seraient donc majoritairement copiotrophes à l’inverse des champignons. Ces derniers sont en effet moins compétitifs et ne sont donc peu présents dans des milieux riches en azote. Toutefois ce classement n’est pas strict, un microorganisme est oligotrophe ou copiotrophe par rapport à un autre. Bien que par rapport aux bactéries les champignons sont oligotrophes, il est possible de les classer entre eux : certains champignons sont copiotrophes par rapport à d’autres.   
La stratégie de vie globale des champignons est assez particulière : ces derniers se confinent sur une ressource jusqu’à épuisement puis produisent ensuite des spores asexuels qui seront ou dispersées afin de coloniser un nouveau milieu ou laissées sur place le temps que la ressource redevienne disponible.  
Les champignons copiotrophes dégradent les composés labiles de la matière organique, tout comme les bactéries.   
Les champignons oligotrophes sont capables de dégrader la matière organique récalcitrante pour pouvoir accéder ensuite à de la matière organique labile (van der Wal et al., 2013). Afin de s’affranchir des faibles quantités d’azote disponibles, les champignons oligotrophes adoptent différentes stratégies comme le transfert de l’azote du sol à la litière via leurs hyphes (Frey et al., 2000 ; van der Wal et al., 2007). Certains sont aussi capables de se nourrir des nutriments présents dans l’air et l’eau de pluie (Ho et al., 2017).  
Toutefois, les réponses des communautés aux changements environnementaux ne dépendent pas seulement de la stratégie de vie mais également de la source de carbone, du type de sol,…   
Les changements de communautés observés sont généralement sur le shéma oligotrophes – copiotrophes dans le cas d’un substrat pauvre (Fierer et al., 2007). Dans le cas de longues expérimentions il est possible d’observer ensuite une recolonisation des oligotrophes (Ho et al., 2017). Inversement, dans le cas où le substrat est riche, on observe d’abord des copiotrophes puis des oligotrophes au fur et à mesure qu’il s’appauvrit (Fierer et al., 2007). Cependant, certaines bactéries copiotrophes sont présentes de manière simultanée avec les oligotrophes car elles se nourrissent de leurs produits de dégradation.

#### Modèle CSR

En 1977, Grime propose un autre modèle : le modèle CSR pour compléter le modèle r/K (Tardy, 2015). Les microorganismes sont alors divisés en trois communautés (Ho et al., 2017) (Figure 6) :

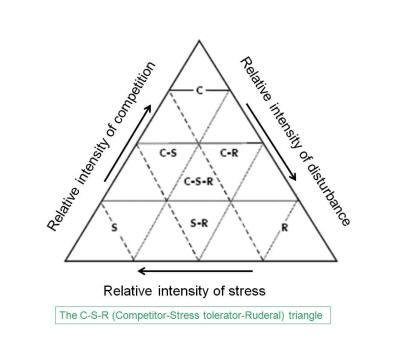
* C pour Compétiteurs : les microorganismes appartenant à cette catégorie sont capables d’utiliser rapidement les ressources lorsqu’ils sont dans un environnement propice (sans stress ni perturbation).
* S pour Stress tolérants : les microorganismes appartenant à cette catégorie sont capables de résister et persister en condition de stress élevé.
* R pour Rudéraux : les microorganismes appartenant à cette catégorie sont capables de s’établir et de recoloniser un environnement qui fait face à des perturbations fréquentes.

Figure 6 : Triangle CSR, (Oudot-Canaff Jehanne, 2016)

#### Modèle GDM (Guild-based Decomposition Model)

Ce modèle tient compte du fait qu’il existe trois populations de bactéries différentes qui décomposent trois pools de carbone. La première catégorie (guilde 1) est appelée opportunistes. Les microorganismes appartenant à cette guilde sont caractérisés par une croissance rapide et une forte affinité avec le substrat (Moorhead, Sinsabaugh, 2006), ils correspondraient à des stratèges r. Ces microorganismes consomment uniquement des composés solubles et des métabolites intermédiaires grâce à la sécrétion d’enzymes hydrolytiques. Les opportunistes sont dominants au cours des premiers temps de la décomposition mais disparaissent ensuite au profit d’autres populations pouvant les éradiquer en sécrétant des antibiotiques au fur et à mesure que la disponibilité des ressources diminue.   
La deuxième guilde est celle des décomposeurs. Ils apparaissent donc après les opportunistes et dégradent la cellulose et la lignocellulose à l’aide d’enzymes hydrolytiques et oxydatives. Ils ont un taux de croissance moins rapide mais sont efficaces quant à l’utilisation des nutriments. Cette guilde regroupe principalement des champignons qui possèdent un ratio C/N élevé leur permettant de se nourrir de composés plus pauvres en azote. D’autre part, ils sont capables de mettre en œuvre des processus de translocation qui leur permet d’être plus compétitifs lorsque la disponibilité des ressources est faible.  
Enfin la dernière guilde, qui intervient à la fin du processus de décomposition comprends des microorganismes mineurs capables de dégrader la matière organique humifiée ainsi que les composés les plus récalcitrants tels que la lignine et les tannins grâce à leur équipement enzymatiques. En effet, ils possèdent des enzymes oxydatives puissantes, indispensables pour dégrader ce type de matière organique. Leurs taux de croissance sont faibles en raison de la complexité et de l’irrégularité de la structure des composés. De plus, les enzymes nécessaires à la dégradation de ce type de matière organique sont coûteuses à synthétiser.

#### Biomasse zymogène/autochtone

Pour terminer, dans certains modèles tel que CANTIS, les microorganismes sont également catégorisés en fonction des pools de matière organique qu’ils sont capables de dégrader mais en deux groupes au lieu de trois : la biomasse zymogène et la biomasse autochtone. La biomasse zymogène comprend les microorganismes qui dégradent la matière organique soluble, issue d’un apport de matière organique fraîche. Alors que la biomasse autochtone comprend les microorganismes présents initialement dans le sol qui dégradent la matière organique humifiée.

# Influence de l’environnement

## Interactions plante-sol

Comme nous l’avons abordé plus haut, les plantes produisent des exsudats racinaires qui constituent une entrée supplémentaire de carbone et qui permettent également la formation d’agrégats. Ces agrégats sont à l’origine de la stabilisation physique du sol (Waligora, 2010 ; Cotrufo et al., 2013). Par ailleurs les exsudats racinaires sont constitués de produits issus de la photosynthèse tels que des sucres et des protéines, permettant l’alimentation des microorganismes (Waligora, 2010). En outre, comme beaucoup d’autres, ce processus est réversible : les exsudats racinaires peuvent aussi déstructurer les associations organo-minérales et ainsi permettre aux microorganismes d’accéder à des composés déjà stabilisés (Recous et al., 2017).  
Enfin comme nous l’avons également déjà évoqué, l’enchevêtrement des racines facilite le piégeage des particules du sol. Toutefois, des racines trop denses sont responsables de la formation de galeries ce qui augmente la porosité du sol. Or une augmentation de la porosité favorise la minéralisation de la matière organique (Derrien et al., 2016).

## Conditions pédoclimatiques

Il existe différents phénomènes pédo-climatiques expliquant la stabilisation de la matière organique du sol tel que le type de sol et notamment les minéraux qu’il contient. En effet, la matière organique soluble peut s’associer avec les fractions limoneuses et argileuses du sol (Cotrufo et al., 2015). Néanmoins, le type d’argile présent dans le sol n’influerait pas sur la stabilisation de la matière organique. Le stockage du carbone sur les minéraux est permis grâce à la formation d’agrégats, difficiles à détruire notamment lorsqu’ils sont en profondeur. Toutefois, il existe aussi des microagrégats, es trouvant généralement plus en surface, où le temps de résidence du carbone y est plus court. En effet, ces agrégats sont plus sensibles aux pratiques culturales par exemple ce qui entraine des rétroactions.   
La dynamique de formation de ces agrégats est en fait liée à la cinétique de décomposition. En effet, la biodégradation de la matière organique stimule les microorganismes qui vont alors produire des produits microbiens stimulant à leur tour l’agrégation en agissant comme liants. Ces agrégats permettent une protection physique : les enzymes sécrétées par les microorganismes ne pourront pas atteindre la matière organique piégée dans les agrégats.

D’autre part, la matrice porale du sol joue un rôle très important dans la décomposition de la matière organique. On distingue plusieurs sortes de porosité dans un sol en fonction de l’origine : texturale ou structurale et en fonction de la taille : macroporosité et microporosité. La macroporosité correspond aux vides existants entre les agrégats alors que la microporosité correspond aux vides à l’intérieur des microoagrégats. Ces pores permettent la circulation d’eau et de gaz : en effet, les pores de grandes tailles alternent entre un remplissage d’eau et d’air alors que les micropores se remplissent principalement d’eau. La taille de ces pores ainsi que leur organisation influe sur la distribution des microorganismes et donc sur la dégradation de la matière organique.  
Par exemple, les microorganismes occupant les pores du sol, la manière dont il sont connectés entre eux va déterminer l’accès des microorganismes à la matière organique et va donc influer sur la décomposition : plus ils sont connectés entre eux et plus la minéralisation est importante (Nunan et al., 2017).   
De plus, lors d’une expérience sur du fructose, la vitesse de minéralisation était corrélée à la taille des pores : plus la matière organique était placée dans des pores de grandes tailles et plus la vitesse de minéralisation était importante (Cotrufo et al., 2013 ; Chenu, 2013).   
En outre, le pH influence lui aussi la décomposition des résidus de culture en agissant directement sur les populations microbiennes (Marschner et al., 2005). Un pH élevé ainsi que des concentration importantes d’azote favoriseraient la croissance des bactéries alors qu’un pH plus faible avec de plus faibles concentrations en azote favoriserait plutôt le développement de champignons (Allison, 1973). Or la décomposition de la matière organique contribue à la modification du pH car elle source de libération d’ions dans le milieu. En effet, l’azote est transformé en ammoniaque par ammonification, ce qui produit des ions OH- qui font augmenter le pH. La nitrification, au contraire entraine la libération d’ions H+ acidifiant le sol (Bolan et al., 2003). La part de ces processus va donc jouer sur le pH du sol.  
Enfin, le climat impacte les vitesses de décomposition de la matière organique mais ne joue pas sur l’importance relative des différents processus (Chenu, 2013). Toutefois, les communautés microbiennes sont dépendantes de la température et de l’humidité du sol, l’optimum de température se situant entre 20 et 45°C selon les espèces. Un climat chaud et humide est propice au développement microbien (Vigil, Spark, 2004). Mais un climat trop humide avec des conditions anaérobie peut ralentir la décomposition de la matière organique (Amin, 2012). Des observations viennent appuyer ces postulats : une minéralisation accrue a été observé lors d’expériences en champ sur du colza lorsque la température augmentait en été. De ce fait le stock d’azote du sol était augmenté lui aussi. De plus, durant ces mêmes expériences on observe une diminution des quantités d’azote en automne/hiver qui serait dû à l’importance de la pluviométrie et donc au lessivage des nutriments (Trinsoutrot, 1999). Il est donc nécessaire d’adapter les pratiques culturales afin de minimiser ces phénomènes.

## Pratiques culturales

La mise en place de couvert végétal permet de minimiser l’érosion et le ruissellement ce qui contribue au piégeage des particules du sol et donc à la stabilisation du carbone (Trinsoutrot, 1999 ; Derrien et al., 2016). De plus il permet aussi de minimiser l’évaporation en période estivale et constitue une source de matière organique pour les microorganismes. Par ailleurs, l’optimisation du contact entre le résidu et le sol optimiserait l’immobilisation d’azote. Il est donc conseillé de broyer et d’enfouir sur quelques centimètres le résidu après récolte. Cette pratique permettrait aux microorganismes d’être en interaction plus directe avec la matière organique. Cela participerait également à minimiser les quantités d’azote nitriques (NO3- ) dans le sol avant le lessivage hivernale (Trinsoutrot, 1999).  
D’autre part, dans une optique de fertilisation azotée, il est recommandé d’utiliser des mélanges de culture légumineuses/graminées. Les graminées sont des plantes qui capturent efficacement l’azote et la destruction du couvert de légumineuse permet d’avoir un engrais vert pour la culture suivante (Recous et al., 2017). De plus, il est conseillé d’adapter au mieux ses pratiques de fertilisation azotée afin de diminuer les risques de lixiviation. En effet, si la quantité d’azote disponible est nettement supérieure aux besoins des microorganismes, le surplus ne sera pas utilisé par les microorganismes et donc lixivié.  
La rotation culturale permet aussi une meilleure gestion des sols notamment de par la diversité des cultures qu’elle impose. Les plantes n’ayant pas toutes les mêmes besoins et les mêmes apports, cette pratique permet d’éviter la raréfaction d’un nutriment en particulier. Là aussi, le fait d’insérer une période de culture de légumineuses n’est que meilleure pour la fertilité du sol à condition d’ajuster sa fertilisation azotée.  
Enfin l’aération du sol semble inévitable pour leur bonne santé. En effet, un sol compacté sera peu poreux et donc ne permettra pas une bonne circulation de l’eau, de l’air et des nutriments indispensables à la vie des microorganismes et des plantes. Toutefois le labour peut également détruire les agrégats du sol et donc changer le devenir du carbone qui était stabilisé à l’intérieur. De plus cette pratique contribue aussi à la destruction des réseaux de mycelium responsables de la stabilité des sols. L’aération est donc bénéfique si elle est se met en place le plus naturellement possible par la pédofaune et notamment grâce aux nombreuses galeries des vers de terre. Afin d’éviter le compactage du sol il est par ailleurs conseillé de limiter les interventions avec des engins agricoles notamment en période où la pluviométrie est élevée.

# Caractérisation analytique de la décomposition des résidus

Afin de prédire la décomposition des résidus, il est nécessaire d’analyser leur composition chimique et de mettre ensuite en place les dispositifs nécessaires afin de pouvoir mesurer la minéralisation du carbone dans les bonnes conditions d’incubation.

## Caractérisation analytique des résidus

### Caractérisation C, N, P et S

Les teneurs en carbone et azote total sont généralement obtenues par combustion analytique selon la méthode de Dumas. Le principe de cette méthode repose sur une combustion totale en présence d’oxygène. A l’issue de la combustion les composés formés sont du CO2, de l’eau et du NO2 qui sont ensuite dosés à l’aide d’un détecteur universel (Gerhardt, 2015) .   
Toutefois, il est également possible de mesurer l’azote total par la méthode de Kjeldahl qui consiste à minéraliser l’azote organique contenu dans l’échantillon par le biais d’une oxydation à l’acide sulfurique. L’azote ainsi obtenu se trouve sous forme d’ammonium (NH4+) et est ensuite dosé (Perrin, 2011).   
La teneur en phosphore totale peut être déterminée par la méthode d’Odynsky (1936).   
Celle en soufre est déterminée par oxydation alcaline selon la méthode décrite par Tabatabai et Bremner. Le principe de cette méthode repose sur l’oxydation des composés soufrés du sol en sulfate qui est lui-même ensuite réduit en hydrogénosulfure par la procédure Johnson-Nishita avant d’être dosé par colorimétrie (Tabatabai, Bremner, 1970).  
Carbone, azote, phosphore et soufre totaux se mesurent en g pour 100g de MS.

La teneur d’un résidu en azote minéral et en carbone soluble peut être obtenue par extraction dans l’eau d’abord à température ambiante pendant 30 minutes puis à ébullition pendant 1h. L’azote minéral peut ensuite être mesuré par colorimétrie en flux continu et le carbone soluble par oxydation en milieu persulfaté couplé à une détection infrarouge (Machinet, 2009).  
La teneur en phosphore minéral peut être obtenue par la méthode de Watanabe et Olsen (détaillée plus bas).  
La teneur en soufre minéral s’obtient généralement par une extraction au CaCl2 suivie par la procédure Johnson-Nishita et un dosage colorimétrique.

### Fractionnement biochimique par la méthode Van Soest

La méthode Van Soest permet le fractionnement de la matière organique en 4 familles biochimiques : les solubles (SOL), l’hémicellulose (HEM), la cellulose (CEL) et enfin la lignine (LIG) (Figure 7).  
Cette méthode a fait l’objet d’une norme AFNOR en 2004: XP U44-162. La première étape consiste à sécher l’échantillon de matière organique à 40°C (Thuriès, 2009) jusqu’à obtenir un poids constant (AFNOR, 2004). Ensuite une extraction à l’eau bouillante est réalisée afin de « solubiliser partiellement des pectines, gélatiniser l’amidon, et dénaturer des protéines » (Lashermes et al., 2007). L’étape suivante consiste à extraire la fraction soluble composée de lipides, de pectines, d’acides organiques, de sucres peu polymérisées, de certains tanins, de protéines solubles, d’azote non protéique et d’amidon. Cette dernière est extraite à l’aide d’un détergent de pH neutre (NDF) qu’on laisse agir pendant 1h à 100°C. Un détergent acide est utilisé par la suite (ADF), toujours pendant 1h à 100°C, afin de solubiliser les hémicelluloses. Enfin, on utilise une solution d’acide sulfurique concentré à 72% pour l’extraction de la cellulose. Cette dernière étape doit être réalisée à 20°C pendant 3h. Il ne reste donc plus que la fraction LIG ainsi que quelques minéraux à ce stade-là. Pour déterminer les quantités de SOL, HEM, CEL, LIG obtenus, une calcination à 480°C pendant 6h est réalisée à l’issu de laquelle les cendres sont pesées. Le résultat final s’exprime en pourcentage de la matière organique sèche totale initiale (en g MS / 100g de MS pour la caractérisation de la matière sèche).  
Cette méthode peut être réalisée pour la matière sèche, le carbone et l’azote. Dans les deux derniers cas, le résultat s’exprimera en g de C (ou N) / 100g de MS ou encore en g C (ou N) / 100g de C (ou N) litière.  
A partir du fractionnement biochimique de Van Soest, il est possible de quantifier la teneur en parois en soustrayant au Van Soest total la fraction soluble.

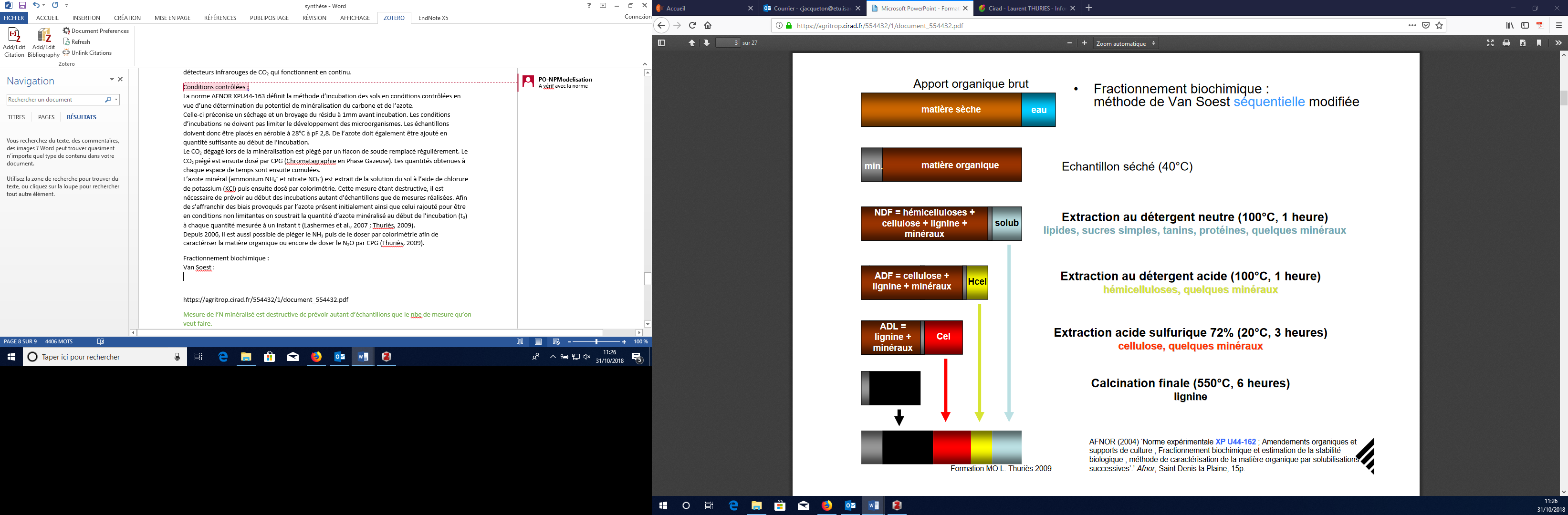


Figure 7 : Illustration du fonctionnnement de la méthode de Van Soest, (Thuriès, 2009)

Toutes les méthodes présentés ici ont des avantages comme des inconvénients c’est pourquoi plusieurs méthodes de caractérisation chimique des résidus sont généralement utilisées.

### Sucres

Les sucres totaux sont généralement déterminés par hydrolyse acide avec dosage chromatographique en prenant le 2-deoxy-D-ribose en tant que standard interne.  
La méthode de Weende permet de caractériser la teneur en cellulose brute d’une litière (Figure 8). Le principe repose sur une hydrolyse acide (acide sulfurique) puis une hydrolyse alcaline. Le résidu obtenu est ensuite séché puis incinéré. La cellulose brute correspond à la différence entre le poids du résidu avant et après incinération (Gall et al., 2012), (Möller, 2014).



Figure 8 : Schéma reprenant la caractérisation des polymères selon les méthodes utilisées, (Gidenne, 2010)

### Phénols

Les polyphénols peuvent être obtenus par extraction au méthanol suivi d’un dosage colorimétrique. Le dosage effectué est généralement un dosage utilisant le réactif de Folin-Ciocalteu. Lorsque les phénols s’oxydent ce réactif est réduit et prend une coloration bleue. L’absorption maximum de cette coloration est proportionnelle à la quantité de polyphénols contenue dans le végétal. L’étalon utilisé peut être l’acide tannique ou encore l’acide gallique (Boizot, Charpentier, 2006).

La lignine VSC est un indicateur servant à déterminer la teneur en lignine. La méthode utilisée ici peut être une oxydation cuprique ou une thioacidolyse. Ces deux méthodes permettant la libération de trois types de monomères : Vanillyl, Syringyl et cynnamyl qui sont des phénols dérivés de la lignine. Ces monomères sont ensuite quantifiés par chromatographie gazeuse par exemple. Leur somme correspond à la teneur en lignine (Carter, Stewart, 1995), (Recous, Le Roux, 2008).

Enfin, la Lignine de Klason (KL) correspond à la matière organique restante après attaque à l’acide sulfurique (H2SO4). En effet, ce dernier hydrolyse les polysaccharides contenus dans la paroi cellulaire (Machinet, 2009).

## Cinétiques de minéralisation en conditions potentielles

### Carbone et azote

Il est possible de réaliser la décomposition des résidus en conditions réelles en champ grâce à des détecteurs infrarouges de CO2 qui fonctionnent en continu ou encore en disposant des pièges à soude sur le site expérimental ainsi que des cylindres pour l’extraction de l’azote minéral, tel que c’est pratiqué en laboratoire.  
La norme AFNOR XPU44-163 définit la méthode d’incubation des sols en conditions contrôlées en vue d’une détermination du potentiel de minéralisation du carbone et de l’azote.  
Celle-ci préconise un séchage et un broyage du résidu à 1mm avant incubation. Les conditions d’incubations ne doivent pas limiter le développement des microorganismes. Les échantillons doivent donc être placés en aérobie à 28°C à pF 2,8. De l’azote doit également être ajouté en quantité suffisante au début de l’incubation et l’humidité est réajustée au long de l’expérimentation si besoin par des ajouts d’eau déminéralisée.  
Le CO2 dégagé lors de la minéralisation est piégé par un flacon de soude étanche remplacé régulièrement et ensuite dosé par CPG (Chromatagraphie en Phase Gazeuse). Il est aussi possible de déterminer la teneur en CO2 par colorimétrie en flux continu à l’aide d’un auto-analyseur ou encore par dosage. Les quantités obtenues à chaque pas de temps sont alors cumulées. Afin de connaitre la quantité de carbone organique présente au départ dans le sol il est d’abord indispensable de le séparer du carbone minéral. Pour cela les carbonates sont neutralisés par de l’acide dont l’excès est ensuite titré.  
L’azote minéral (ammonium NH4+ et nitrate NO3-) est extrait de la solution du sol à l’aide de chlorure de potassium (KCl) puis ensuite dosé par colorimétrie selon les normes NF EN ISO 11732 et NF EN ISO 13395 respectivement. Cette mesure étant destructive, il est nécessaire de prévoir autant d’échantillons que de mesures réalisées dès le lancement de l’expérimentation. Afin de s’affranchir des biais provoqués par l’azote présent initialement ainsi que celui ajouté pour être en conditions non limitantes, la quantité d’azote minéralisé au début de l’incubation (t0) est soustraite à chaque quantité mesurée à un instant t (Lashermes et al., 2007 ; Thuriès, 2009).  
Il est également possible de piéger le NH3 puis de le doser par colorimétrie ainsi que de doser le N2O par CPG (Thuriès, 2009).   
Ces mesures permettent de déterminer le potentiel de minéralisation, c’est-à-dire la « proportion maximale d’azote ou de carbone organique du produit testé, susceptible de se minéraliser »(AFNOR, 2004).  
Selon la norme, 9 mesures doivent être réalisées pour déterminer le potentiel de minéralisation du carbone : au 1er, 3ème, 7ème, 14ème, 21ème, 28ème, 49ème, 70ème et 91ème jour d’incubation. Concernant l’azote seulement 7 sont nécessaires : à la mise en place de l’expérimentation, au 7ème, 14ème, 28ème, 49ème, 70ème et 91ème jour d’incubation.  
Afin de connaître les quantités de carbone organique restantes après une expérimentation, il est possible d’effectuer une combustion par voie humide : le carbone organique est oxydé grâce à un mélange de bichromate de potassium, d’acide sulfurique et d’acide phosphorique. A l’issu de cette combustion le carbone organique est transformé en CO2 et pourra donc être dosé par colorimétrie (Lashermes, 2017a). Une combustion oxydative est aussi utilisée afin de déterminer la teneur en azote et carbone total. La combustion oxydative permet le passage de l’azote et du carbone sous forme minérale (diazote et dioxyde de carbone). Ces gaz sont ensuite séparés et détectés par conductibilité thermique.

Enfin concernant la biomasse microbienne, la technique la plus utilisée est celle de fumigation-extraction où l’on soumet le sol à des vapeurs de chloroforme entrainant la mort des microorganismes. L’échantillon soumis à ces vapeurs est appelé l’échantillon fumigé. Un autre échantillon est non fumigé. Les quantités de carbone et d’azote sont extraites pour les deux échantillons. Le carbone/l’azote de la biomasse microbienne correspond à la différence entre les quantités extraites à partir de l’échantillon fumigé et celui non fumigé (Alavoine, Lashermes, 2018).

### Phosphore et soufre

Concernant le phosphore, il existe 3 méthodes normalisées en France : la méthode Dyer, la méthode Joret-Hébert et méthode Olsen. Cette dernière est celle utilisée par l’INRA et fait l’objet d’une norme : NF ISO 11263. Le réactif d’extraction est une solution de bicarbonate de sodium à pH 8.5. Le phosphore est ensuite dosé par colorimétrie à 882 nm (Gatoux, 2011).  
Les méthodes utilisées pour la dynamique du soufre sont les mêmes que celles utilisées pour la caractérisation du soufre minéral dans un résidu à savoir une extraction au CaCl2 ou encore au KH2PO4.  
Pour la détermination du soufre microbien, la méthode généralement utilisée est la méthode de (Saggar et al., 1981) qui reprends les grandes lignes de la fumigation-extraction.

## Cas particulier du marquage isotopique

Cette technique consiste à marquer les résidus avec des isotopes stables (qui ne se désintègrent pas) afin de pouvoir suivre le cheminement de l’élément marqué. Les isotopes stables les plus utilisés dans notre cas sont le carbone 13 et l’azote 15 et leurs teneurs sont généralement mesurées par un analyseur élémentaire couplé à un spectromètre de masse.  
Dans la plupart des cas, les résidus marqués ont grandi dans des chambres de croissance spéciales avec des solutions enrichies en isotope stable.  
On parle d’excès isotopique ou d’abondance isotopique quand on mesure la quantité de l’élément marqué par rapport aux isotopes totaux. Cette donnée s’exprime en pourcentage d’isotope stable.

##### Conclusion

La décomposition des litières végétales est un processus long et complexe faisant intervenir plusieurs cycles d’éléments chimiques notamment ceux du carbone et de l’azote. Les acteurs de ce processus sont vivants et interagissent entre eux ce qui le complexifie d’autant plus. Toutefois, la décomposition des litières dépend également de leur composition qui elle-même dépends de divers facteurs tels que les organes concernés ou encore les espèces étudiées. Enfin l’environnement joue lui aussi un rôle clé dans la décomposition des litières végétales notamment de par les conditions pédoclimatiques du site étudié mais aussi de par les pratiques culturales mises en place et le type de plante cultivées au moment de la décomposition. D’autre part, toute la complexité du processus repose notamment sur le nombre importants de rétroactions possibles liées à l’équilibre fragile des différents facteurs de contrôle cités ici.  
Dans le domaine de la recherche, afin de pouvoir prédire approximativement la décomposition d’un résidu donné celui-ci est caractérisé à minima par fractionnement biochimique ainsi que de par ses teneurs en C et N et les cinétiques de minéralisation de chaque type de résidu sont mesurées à intervalles réguliers. Des marquages isotopiques sont également réalisés pour obtenir plus d’informations sur le devenir d’un ou plusieurs éléments chimiques en particulier. Ces procédures malheureusement encore peu normalisées et différentes d’un laboratoire à l’autre ont permis de produire un grand nombre de données dont la plupart ont déjà été valorisée dans des articles scientifiques. Toutefois nous nous inscrivons aujourd’hui dans des objectifs de science ouverte visant à partager les données et à leur rendre une seconde vie, c’est pourquoi nous allons voir les modalités d’une telle démarche de valorisation dans le chapitre suivant.

##### Chapitre 2 : Gestion et partage de données scientifiques

##### Introduction

Le contexte national et européen des infrastructures de recherche favorise de plus en plus des politiques de gestions des données. En effet, pour être inscrit sur les feuilles de route nationale et européenne, les instituts sont dans l’obligation d’avoir une telle politique. Les données produites doivent répondre à des objectifs fixés dont notamment celui de l’ouverture des données.   
De telles politiques sont donc mises en place au niveau institutionnel, au sein de l’Inra. La charte de l’institut repose sur les six valeurs suivantes : « ouverture, partage, transparence, accès aux données, soutenabilité financière et conformité. », (INRA, 2016a). En effet, en signant la déclaration de Berlin en 2004 qui incite à rendre l’information scientifique librement accessible, l’Inra est entré dans une démarche Open Science. Cette démarche permet d’ouvrir les données de la recherche et de les rendre accessibles à tous. L’objectif est de capitaliser les efforts expérimentaux de chacun et d’éviter les duplications. Cela permettrait d’augmenter l’efficacité de la recherche. La loi Lemaire « Pour une République Numérique » du 7 octobre 2016 incluant la problématique des données de la recherche, renforce la démarche Open Science déjà mise en place. Découle de cela une charte publiée par l’Inra pour le libre accès à ses publications et données scientifiques en février 2017. Cette charte comprend sept points de recommandations destinés aux chercheurs afin de les inciter à entrer dans la démarche Open Science (INRA, 2016b).  
D’autre part, depuis quelques années, avec la révolution du numérique et l’acquisition de jeux de données toujours plus conséquents, l’Inra s’est fixée des objectifs Open Sciences. Ces objectifs comprennent :

* L’organisation des infrastructures de recherche afin qu’elles soient connectés
* L’organisation des données pour faciliter leur partage et leur réutilisation notamment grâce à la création d’un portail de données et de collaborations avec le CEA, le CNRS ainsi que l’Inria et l’Irstea.
* Favoriser les approches prédictives en biologie et en écologie
* Proposer de nouveaux modes de diffusion de la connaissance notamment en adaptant les revues propriétaires à la science ouverte mais également en encourageant des modèles alternatifs de publications.
* Adapter le métier et l’environnement du chercheur au numérique en demandant des compétences en analyse de jeu de données massifs par exemple ou encore en améliorant le processus de dématérialisation.

En outre, Horizon 2020, le programme de financement de la recherche et de l’innovation de l’Union européenne pour la période 2014-2020 s’inscrit lui aussi dans cette démarche de science ouverte. En effet, toutes les publications issues de projets financés par ce programme doivent être en libre accès. De plus, il existe un pilote ORD : Open Research Data qui tend à rendre accessibles le plus de données possibles acquises dans le cadre d’un projet H2020 (Ministère de l’enseignement supérieur, de la recherche et de l’innovation, 2014).  
Enfin, en 2018, apparaît le plan national pour la science ouverte qui « rend obligatoire l’accès ouvert aux publications ainsi qu’aux données issues de recherches financées sur projet ». Ce plan est composé de 3 axes décrivant des mesures afin de « généraliser l’accès ouvert aux publications », « structurer et ouvrir les données de la recherche » ainsi qu’afin de « s’inscrire dans une dynamique durable, européenne et internationale » (Ministère de l’enseignement supérieur, de la recherche et de l’innovation, 2018).  
Cette synthèse présentera donc les modalités liées à l’Open Science, comment s’inscrire dans cette démarche de valorisation des données. Pour cela, une première partie abordera les enjeux de l’accessibilité aux données tandis que les parties suivantes se verront plus techniques et permettront de donner un cadre de gestion des données au sens large, toujours dans un objectif d’open science. En effet, ces parties traiteront de la production de données FAIR, de la gestion des données, de leur partage et de leur archivage et enfin des différentes formes de valorisation envisageables. La sixième et dernière partie comprends trois études de cas de data paper afin d’avoir des exemples concrets du cycle de vie des données à travers la structure d’un document de valorisation.

# Les enjeux de l’accessibilité aux données scientifiques

L’open data est un concept récent parfois encore mal connu du personnel scientifique. Les enjeux de l’ouverture des donnés sont multiples, que ce soit pour les personnes ou les instituts producteurs de données ou pour les utilisateurs des données (Pôle Données de la Recherche IST, 2018e), (Aumont, 2017).

## Enjeux patrimoniaux

Inciter les chercheurs à rendre leurs données accessibles permet tout d’abord d’éviter de perdre des données. En effet, de nombreuses données sont encore dans des formats non pérennes voir peut-être non lisibles car trop vieux pour être remis à jour. De plus, de nombreux chercheurs n’ont plus connaissance des lieus de stockage des données acquises au long de leur carrière ou n’ont pas toujours pris le temps nécessaire pour stocker leur données correctement. En effet, il est estimé qu’environ 80% des données sont perdues 20 ans après publication (Roberge, 2015). D’autre part, le problème du stockage des données est d’autant plus important pour des chercheurs ayant travaillé dans plusieurs instituts différents au cours de leur carrière. De plus, dans le cas où les supports de stockage sont encore disponibles et lisibles par d’autres scientifiques, les données sont parfois trop peu décrites à l’aide de métadonnées pour être comprises et réutilisées par d’autres. Enfin les données apparaissant dans les publications sont généralement des données traitées et il est souvent difficile de retrouver les données brutes initiales à partir de celles-ci. Or, l’accessibilité à des données plus anciennes est impérative dans certains domaines comme la météorologie par exemple ou tous les domaines concernant l’évolution des sols.

## Enjeux économiques

Tout d’abord, il est inutile que les fonds de l’Etat financent des expérimentations supplémentaires sur un projet alors que les données produites par d’autres instituts pourraient être en partie réutilisées car complémentaires. L’Open data (données ouvertes) permet donc « d’accélérer l’innovation et le retour sur Investissement dans la R&D » (Pôle Données de la Recherche IST, 2018e). A titre d’exemple, on peut citer l’institut européen de bio-informatique (EMBL-EBI), organisation intergouvernementale fournissant gratuitement des données dans le monde entier. Les avantages sont estimés à 1 milliards de livres sterling par an pour les utilisateurs et leurs bailleurs de fonds ce qui correspond à 20 fois le coût opérationnel de l’institut (Beagrie, Houghton, 2016).

## Enjeux scientifiques

D’une part, l’accessibilité aux données garantit une certaine qualité. En effet, de nombreuses questions se posent actuellement sur la qualité des productions scientifiques et notamment sur le niveau de signification statistique des conclusions tirées. Il semblerait que seulement 20% seraient statistiquement fiables (Aumont, 2017). En publiant les données brutes, la recherche peut s’affranchir de ce risque, les données étant la preuve de l’exactitude de la publication. Plus les données seront vues et réutilisées par des collaborateurs et plus le risque d’erreurs ou de conclusions hâtives sera évité.   
De plus, en amont, le chercheur qui publiera ses données sera plus attentif. Toutefois, publier ses données entraîne de nombreux changements techniques pour la communauté scientifique et de nouvelles responsabilités.   
Pour l’institut de recherche qu’est l’INRA, l’accessibilité aux données permet d’assurer l’intégrité de ses recherches et produits.

D’autre part, l’accessibilité des données peut être une passerelle entre les instituts de recherche, producteurs de données, et les instituts techniques, utilisateurs de données. En effet, chaque année, 7% des adresses mail des chercheurs ne sont plus fonctionnelles ce qui rend plus difficile l’accès aux informations. De plus, l’Open data permettrait également de faciliter l’usage des données notamment par ces communautés techniques.

Enfin, l’ouverture des données les rend citables et améliore la vitesse d’accessibilité à ces dernières. Découle de cela un taux de citation pouvant être plus important ce qui est dans l’intérêt des producteurs de données qui voient leur travail reconnu.

## Enjeux sociétaux

Comme il a été écrit précédemment, l’Open data permet une traçabilité des données et garantit une certaine qualité. L’accessibilité aux données contribue donc à améliorer l’image de la recherche auprès des citoyens ainsi qu’à augmenter leur confiance et donc leur participation à la science (sciences participatives).  
En termes d’éducation, les jeux de données publiés pourraient permettre à des étudiants de les intégrer à leurs productions pour l’argumenter ou encore de s’entraîner sur de vraies données.  
L’Open data permettrait donc de nouveaux liens entre citoyens et instituts de recherche.

## Enjeux éthiques

L’ouverture des données nécessite de bien réfléchir en amont à la question « Quelles données publier ? ». En effet, d’un point de vue éthique et juridique les chercheurs se doivent de respecter les droits d’auteurs, la vie privée et selon le domaine il est parfois possible de rencontrer des obligations de secret ou de sécurité.  
Selon le type de données, l’ouverture ne sera donc pas toujours possible notamment pour des données personnelles ou encore des données concernant le domaine de la santé par exemple.

# Produire des données FAIR

FAIR est un acronyme signifiant : « Findable Accessible Interoperable et Reusable » soit en français : Trouvables Accessible Interopérable et Réutilisable. Ces quatre principes s’appliquent aux données produites par la recherche dans le cadre de la commission européenne et plus précisément de l’objectif Horizon 2020 (H2020). Ceux-ci sont censés permettre l’ouverture des données de la recherche. Chacun de ces quatre principes a été décliné en différents niveaux selon le collectif de travail FORCE 11.

## Findable

La première étape pour avoir des données ouvertes est de les rendre trouvables, que ce soit par des humains ou par des machines. 4 niveaux de caractéristiques des données « findable » ont été établies :

F1 : Les données et métadonnées sont associées à un identifiant unique et pérenne.

F2 : Les données sont décrites par des métadonnées riches.

F3 : Les données et métadonnées sont enregistrées ou indexées dans un dispositif permettant de les rechercher.

F4 : Les métadonnées spécifient l’identifiant des données.

## Accessible

Selon ce principe, les données et les métadonnées doivent être stockées durablement et leur accès facilité. Il est possible de spécifier des conditions d’accès (ouvert ou restreint) et d’utilisation (licence).

A1 : Les données et métadonnées doivent être accessibles par leur identifiant via des protocoles de communication standardisés.

A1.1 : Le protocole utilisé est ouvert, libre, et universellement implémentable.

A1.2 : Le protocole permet des procédures d’authentification et d’autorisation lorsque cela est nécessaire.

A2 : Les métadonnées sont accessibles même lorsque les données ne le sont plus.

## Interoperable

Les données doivent être combinables avec d’autres données, téléchargeables, utilisables et intelligibles par des humains comme par des machines. Les systèmes informatiques doivent pouvoir dialoguer entre eux et échanger des données sans ambigüité (INIST, CNRS, 2018). Le web des données est la continuité du concept d’interopérabilité (DoRANum, 2017).

I1 : Les données et métadonnées utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances.

I2 : Les données et métadonnées utilisent du vocabulaire qui suit les principes FAIR.

I3 : Les données et métadonnées incluent des liens vers d’autres données ou métadonnées.

## Reusable

Le but ultime des principes FAIR est d’optimiser la réutilisation des données. Pour cela les données et métadonnées doivent être doivent être correctement décrites.

R1 : Les données et métadonnées ont des attributs multiples et pertinents.

R1.1 : Les données et métadonnées sont mises à disposition selon une licence claire et accessible.

R1.2 : Les données et métadonnées sont associées à leur provenance.  
R1.3 : Les données et métadonnées correspondent aux standards des communautés indiquées.

# Gérer des données

Les principaux risques dans la gestion des données reposent sur l’obsolescence matérielle, logicielle ou encore du format de fichier ainsi que la perte de son contenu. On peut également citer d’autres risques plus secondaires tels qu’un mauvais lieu de rangement ou encore un mauvais nommage des fichiers. Cette partie comprend donc des explications sur la manière dont il est possible de s’affranchir de ces risques afin que les données respectent les principes FAIR.   
La première étape dans la gestion durable des données consiste à établir un plan de gestion des données (Marie Claude Quidoz, 2018).

## Etablir un plan de gestion des données (PGD)

Un plan de gestion des données est un document formel et évolutif qui décrit de façon synthétique toutes les étapes du cycle de vie de la donnée : de la collecte à l’archivage et au partage, basé sur les principes FAIR. Ce document est considéré comme une bonne pratique de gestion et permet de gagner en temps et en efficacité ainsi que de prévoir au mieux le budget à allouer aux données. Le plan de gestion n’est pas figé mais évolutif au cours du projet et se remplit pendant et après le projet.

Dans le cadre d’un projet financé par H2020, tout scientifique doit participer à l’action pilote ORD : « Open Research Data » soit en français « libre accès aux données issues de la recherche » sauf exception si les données ne peuvent être partagées pour des raisons de sécurité ou autre. Dans ce cas, ce choix devra être motivé. Toutefois, il est possible de participer au pilote ORD (voir même vivement conseillé) bien que la totalité des données ne soit pas rendue accessible. Le principe de l’ORD concernant l’accessibilité des données est « as open as possible, as closed as necessary ».   
Tout participant à l’action pilote ORD doit obligatoirement fournir un PGD à la commission européenne dans les 6 mois qui suivent le commencement du projet. Il est également recommandé de soumettre à titre volontaire un plan de gestion des données pour les projets se désengageant du pilote. Une nouvelle version du DMP est nécessaire en cas de modifications significatives du projet tel que l’acquisition de nouvelles données, la déposition d’une demande de brevet, des arrivées ou départs de membres,… Dans le cas d’une évaluation périodique, le PGD mis à jour doit-être fournit. Dans tous les cas, un examen final à lieu où la version finale du PGD doit être soumise (EUROPEAN COMMISSION - Directorate-General for Research & Innovation, 2016) (Figure 9).

Afin d’aider les participants au projet pilote ORD, un modèle de plan de gestion de données applicable à tout projet H2020 est proposé par la commission européenne. Ce modèle est basé sur les principes FAIR et comporte 6 parties composées chacune d’une série de questions. Un résumé synthétique du modèle est disponible en annexe 2. Un seul DMP doit être réalisé par projet même si celui-ci contient plusieurs jeux de données. Néanmoins tous les jeux de données doivent apparaître dans le PGD.

Afin d’aider la communauté scientifique française dans ses démarches un outil collaboratif est mis à disposition par l’Inist-CNRS : DMP OPIDoR (Data Management Plan pour une Optimisation du Partage et de l’Interopérabilité des Données de la Recherche). Cet outil est basé sur le code open source DMPRoadmap. Des instituts peuvent ajouter leur propre PGD comme c’est le cas de l’INRA (Doranum, 2018). DMP OPIDoR permet de partager le PGD avec des collaborateurs et offre des sauvegardes automatiques sur un serveur en France. Toutefois, pour avoir toutes les versions du PGD il est fortement recommandé de l’exporter (Filatre, Hensens, 2017).

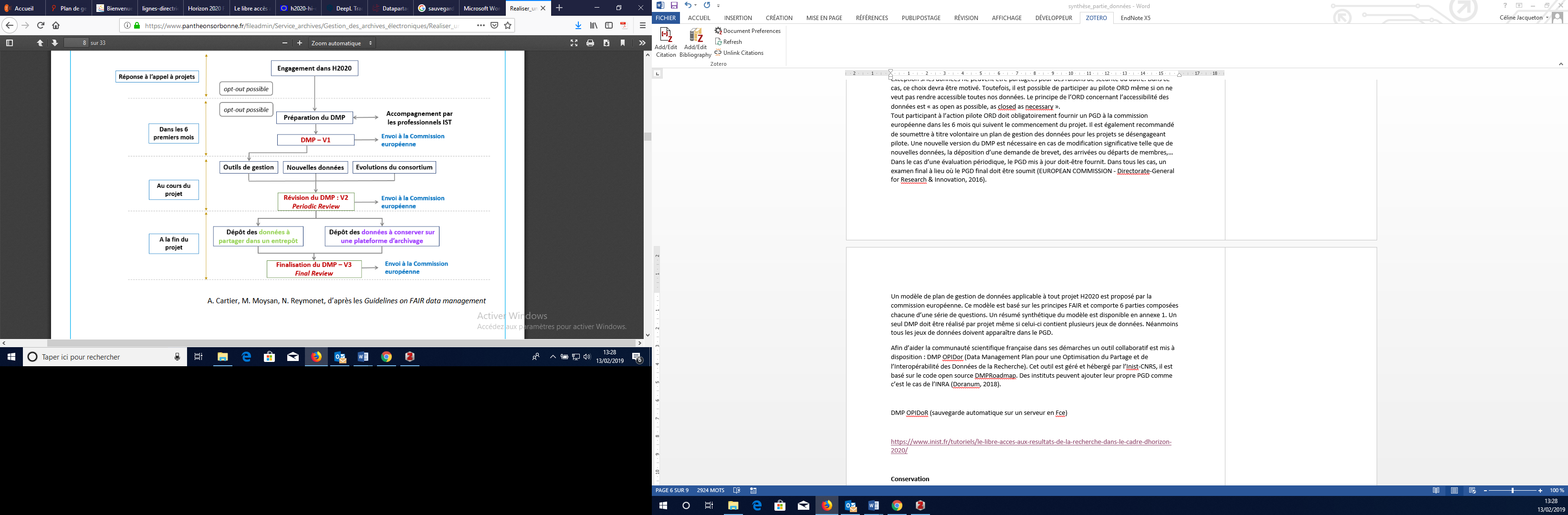


Figure 9: Marche à suivre en terme de plan de gestion des données dans le cadre d'une participation au pilote ORD. A. Cartier, M. Moysan, N. Reymonet, d’après les Guidelines on FAIR data management.

## Documenter et organiser les données

Cette partie est très importante, que ce soit pour l’auteur ou l’utilisateur des données. En effet une bonne gestion des règles de nommage, d’organisation et de renseignements des métadonnées permet de gagner en intelligibilité et d’accroître largement le potentiel de réutilisation des données.

## Nommage et organisation des fichiers et dossiers

Le nommage et l’organisation des fichiers (ou dossiers) de données permettent un gain de temps considérable, que ce soit pour l’auteur pendant le projet qui peut plus facilement localiser et distinguer les fichiers que pour l’utilisateur lorsqu’il veut réutiliser les données. Une organisation structurée et un nommage clair permettent donc de faciliter le partage des jeux de données.  
En outre, cela permet également de réduire les risques de suppression ou de déplacements malencontreux et donc de préserver l’intégrité des données (Pôle Données de la Recherche IST, 2016).

Concernant le nommage, il est tout d’abord conseillé de choisir une convention de nommage adaptée aux fichiers et aux contenus. Cette convention doit être appliquée pour tous les fichiers et toujours de la même façon. Il est préférable d’écrire cette convention quelque part afin de pouvoir s’y référer lors du nommage de nouveaux fichiers ou pour mieux (re)comprendre les titres des fichiers nommés ultérieurement.  
Quelques règles de bases :

* Le nom de fichier doit être court : l’INRA recommande au maximum 25 caractères (Pôle Données de la Recherche IST, 2018c). Les noms de fichiers peuvent inclure des abréviations, dans ce cas il est préférable de choisir des abréviations usuelles telles que par exemple AP pour Avant-Projet, CR pour Compte Rendu, VP pour Version Provisoire, VF pour Version Finale,... Il est aussi possible et conseillé d’y faire figurer les initiales du chercheur ainsi que le numéro de version ou encore la date. Attention cependant à ce dernier point, les dates doivent avoir un format consistant de type AAAA-MM-JJ selon la norme ISO 8601 (Unil, 2016). Les fichiers doivent être identifiables par leurs noms même en cas de changement du lieu de stockage.
* Il est impératif d’éviter les caractères spéciaux ainsi que les accents, les cédilles ou encore les espaces. Ces derniers peuvent être remplacés par des underscores ou encore des tirets (Unil, 2016). Attentions certaines applications ne font pas la différence entre majuscules et minuscules donc supprimer les espaces en les remplaçant par des majuscules en début de mot n’est pas toujours pertinent (Pôle Données de la Recherche IST, 2018c).
* Dans le cas d’une numérotation il est préférable d’utiliser des zéros devant le numéro. Par exemple il vaut mieux écrire 001, 010, 100 plutôt que 1, 10, 100. Cette règle permettra d’assurer le bon ordre leur dans leur affichage (Unil, 2016).
* Il est également conseillé de ne pas garder toutes les versions intermédiaires d’un même fichier afin d’éviter les confusions.

Concernant l’organisation de l’arborescence, il est recommandé d’établir un plan de classement pour les projets impliquant plusieurs personnes. Idéalement, tous les collaborateurs doivent participer à son élaboration (Laperdrix, 2016). Pour l’élaboration d’un tel plan, il est conseillé d’aller du plus général au plus spécifique et de mettre en place une organisation durable, qui ne serait pas impactée par des changements quelconques (Arseneau, 2015).  
Enfin, afin d’éviter tout problèmes de stockage et de perte de fichiers, il est conseillé d’utiliser la règle du 3-2-1 à savoir : produire 3 copies, sur 2 supports différents dans au moins 1 lieu distant (Marie Claude Quidoz, 2018).

## Documenter les données

Afin de documenter les données, des métadonnées sont utilisées. Etymologiquement métadonnées signifie « données sur les données », c’est en fait un ensemble de données structurées permettant de décrire, expliquer, localiser les données afin que celles-ci soient plus compréhensibles. Elles permettent d’appliquer les principes FAIR et donc de faciliter (DoRANum, 2017), (INIST, CNRS, 2018) :

* L’accessibilité : elles favorisent la recherche des données notamment grâce à la recherche multicritères.
* L’interopérabilité : elles doivent permettre aux données d’être interprétables par des humains comme par des machines
* La pérennité : les métadonnées doivent permettre de documenter les éventuelles migrations d’un format de document à un autre
* La réutilisation : elles permettent une meilleure gestion et organisation des productions scientifiques notamment grâce à l’identification de l’auteur et la définition des usages.

Les métadonnées répondent généralement aux questions : « Quoi ? », « Où ? », « Quand ? », « Comment ? », « Pourquoi ? » et définissent (INIST, CNRS, 2018) :

* Le contexte de production : protocoles expérimentaux, procédés de traitement et d’analyse, auteur, date, localisation, objectifs de la collecte des données, financeur.
* Le contenu intellectuel : définitions, titre, résumé, domaine de recherche, mots-clés, type de ressource.
* Les caractéristiques techniques des données : format de fichiers, taille, organisation, liens.
* Les propriétés et droits d’usage : détenteur des droits, licence, droits d’exploitation, conditions d’accès, outils spécifiques.

Ces métadonnées peuvent être embarquées ou bien externes, c’est-à-dire dans un catalogue d’accompagnement par exemple ou encore un annuaire d’entrepôt. Elles peuvent être gérées par plusieurs approches différentes :

* Un tableau de type Excel par exemple : cette approche, bien que largement utilisée, est déconseillée car elle ne prend pas en compte la sémantique et ne permet pas de faire des requêtes facilement. Un tableau construit dans un format csv est préférable car il permet l’interopérabilité (lisible par humains et machines).
* Une structure en arbre de type xml : cette approche permet une meilleure organisation des métadonnées ainsi que la possibilité de lier des concepts avec d’autres.
* Un graphique sémantique avec des triplets RDF : les métadonnées sont organisées de façon structurées et cette approche permet la prise en compte de la sémantique.

Il existe des standards de métadonnées ainsi que des modèles de métadonnées LovInra compatibles avec l’AgroPortal (Puren, 2017). Ces standards donnent une trame à suivre pour la description des données. On les choisit en fonction de la destination des données (certains entrepôts ont leurs propres standards) ou encore par domaines spécifiques. Pour aider les chercheurs dans leur démarche, le site internet du DCC (Digital Curation Center) et celui du RDA (Research Data Alliance) proposent des listes de standards de métadonnées et d’outils d’aide à la description. Le site FAIRsharing propose également des standards de métadonnées selon la discipline étudiée. Il est nécessaire d’indiquer les standards utilisés pour les jeux de données dans le DMP.  
Ces standards peuvent être sous forme de schémas ayant pour avantages d’indiquer les relations existants entre les différentes métadonnées. Dans le cas d’un schéma de métadonnées il est souvent nécessaire d’utiliser le format xml permettant une structuration en arborescence.

Les schémas sont composés de 4 parties décrivant chaque élément :

* la signification (auteur, date, titre,…)
* le contenu (nombre, texte)
* la formulation (« Nom, Prénom », texte libre, norme ISO 8601)
* la valeur

Un schéma définit également le caractère obligatoire des métadonnées (obligatoire, recommandé et facultatif), les règles d’ajout d’éléments ainsi qu’éventuellement des règles plus spécifiques (obligation de renseigner tel champ si tel autre champ est rempli par exemple). Il est possible d’utiliser le schéma DataCite par exemple ou encore Dublin Core qui et générique et qui a fait l’objet d’une norme ISO (certifié ISO 9001).

Au sein de ces schémas les métadonnées peuvent être décrites à l’aide de vocabulaire contrôlé, de thésaurus ou encore d’ontologies qui sont respectivement des listes de termes prédéfinis, des répertoires structurés de termes et « un ensemble structuré de termes et concepts représentants le sens d’un champ d’informations » (Wikipédia). Ces outils améliorent la visibilité des données ainsi que leur réutilisation. En agronomie, il est possible d’accéder à ces outils de description déjà existants en passant par exemple par l’Agroportal.

## Identifier les données

Dans le cadre du projet H2020, il est obligatoire de « faire référence à un système standard d’identification », (COMMISSION EUROPÉENNE - Direction Générale de la Recherche et de l’Innovation, 2016). On utilise donc un identifiant pérenne (PID) afin d’identifier ses données ou même plus largement les contributeurs et auteurs du jeu de données. Le PID sont des identifiants uniques et pérennes permettant d’associer chaque jeu de données à une ressource ou entité. Ce sont en quelques sortes les équivalents des URL dans le web. Toutefois, les URL ne sont pas des identifiants stables contrairement aux PID car la ressource n’est plus accessible dès lors qu’elle est déplacée ou renommée. La gestion active des PID par des infrastructures humaines et techniques permet d’éviter ce problème. Il existe deux types d’identifiants pérennes : objets et contributeurs.

#### Identifiant pérenne objet (Figure 10) :

Ces identifiants concerne des objets c’est-à-dire des publications et des jeux de données. Ils permettent la découverte, le partage, la citation et la réutilisation des données. De plus, ils font également le lien entre les données ayant permis la publication et la publication elle-même. Il existe une grande variété d’identifiants objets : handle, PURL (Persitstant URL), ARK (Archive Resource Key), ePIC (european Persistant Identifier Consortium), DOI (Digital Object Identifier). Il existe également des identifiants appelés URI qui étendent le principe des URL et permette de prendre en compte la sémantique. Toutefois ces identifiants ne seraient pas tout à fait persistants.   
Il est recommandé d’utiliser plutôt des identifiants de type Handle, ePIC et DOI car ces derniers sont globaux et acceptés dans un grand nombre d’entrepôts. De tels identifiants garantissent l’interopérabilité et l’accessibilité des données.   
Par ailleurs, tous les identifiants pérennes garantissent des fonctionnalités de base qui sont l’identification et l’accès permanent au jeu de donnée. Toutefois, le système Handle serait plus robuste et le DOI offre des services supplémentaires aux fonctionnalités de base. De ce fait, c’est le DOI qui est le plus souvent utilisé (INIST, CNRS, 2018), (DoRANum, 2018b). C’est également ce dernier qui est recommandé par l’INRA.

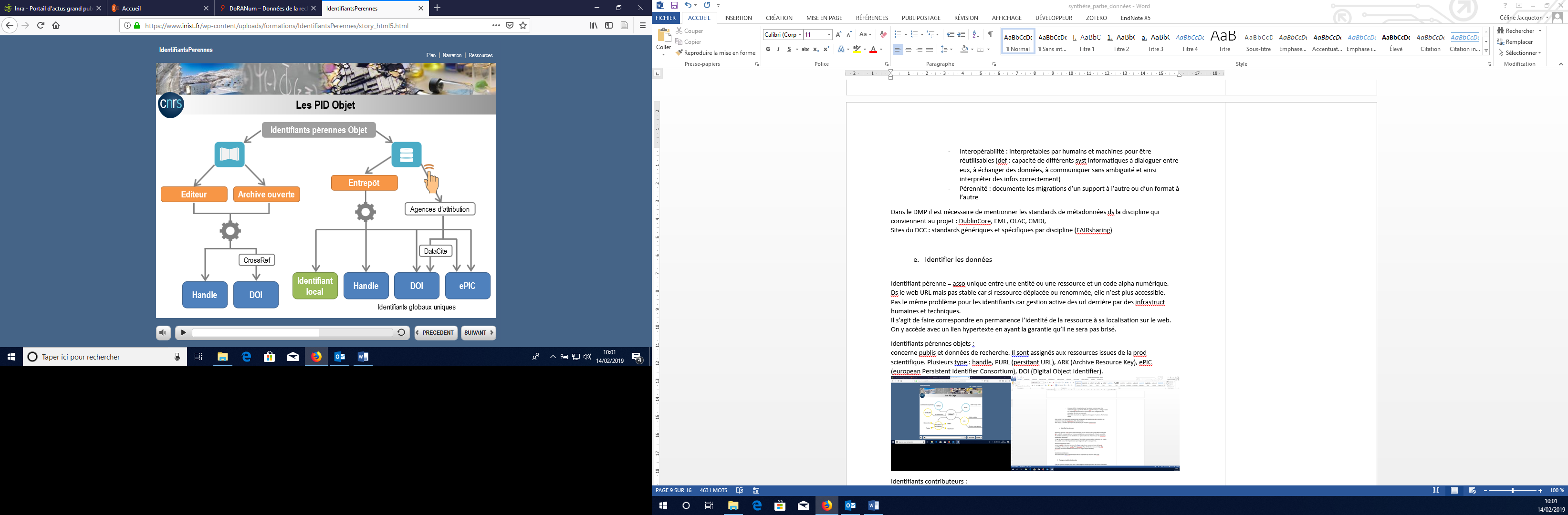


Figure 10 : Schéma des principaux identifiants objets pérennes, (INIST, CNRS, 2018).

#### Identifiant pérenne contributeur :

Ces identifiants sont associés aux auteurs d’une production scientifique ou encore à un organisme de recherche assurant cette production. Cela permet notamment d’être plus visible et offre les mêmes avantages que les identifiants objets en termes de citation et de partage. Comme pour les identifiants objets, il en existe plusieurs types (Figure 11). L’obtention de ces identifiants se fait souvent de manière automatique sauf pour le researcher ID du WOS (Web Of Science) ainsi que les identifiants liées aux archives ouvertes (idHAL et ArXiv Author ID). Il est également possible d’avoir un identifiant global grâce aux plateformes ISNI (International Standard Name Identifier) et ORCID.

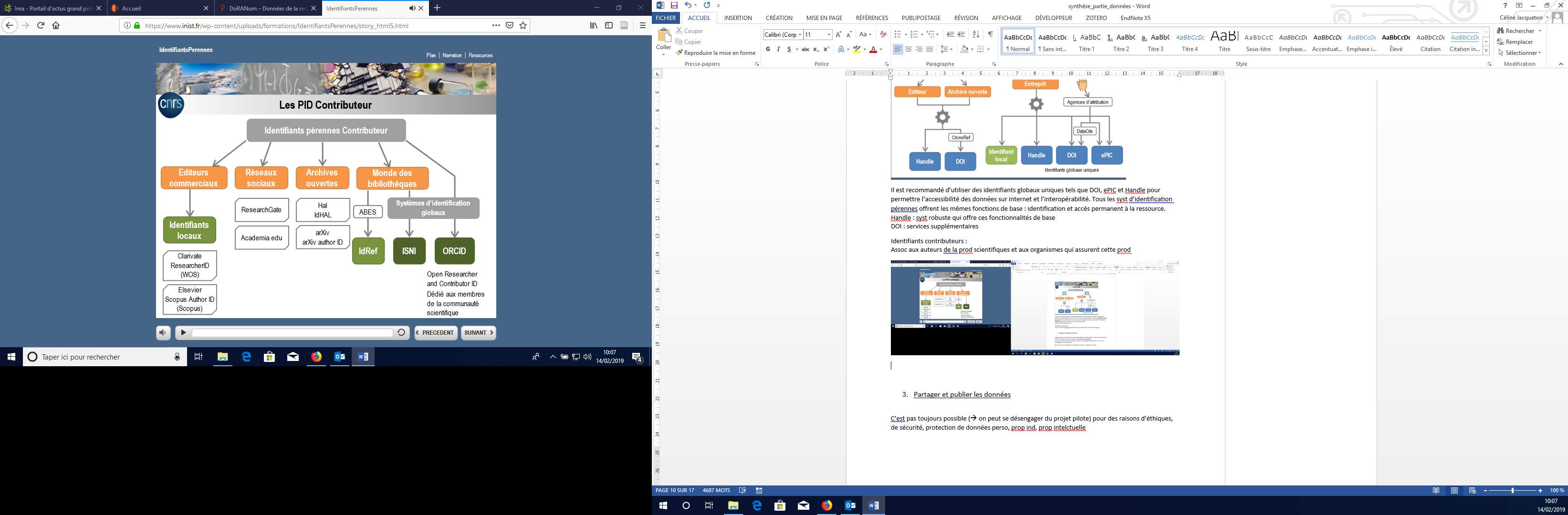


Figure 11: Schéma des principaux identifiants pérennes contributeurs, (INIST, CNRS, 2018).

\*ABES : Agence Bibliographique de l’Enseignement Supérieur

# Partager et archiver des données

La gestion de données se fait bien souvent dans un but de partage, que ce soit avec d’autres contributeurs, la communauté scientifique ou alors la communauté citoyenne. Dans le cadre de l’Open Science et du projet H2020, les données doivent être autant ouvertes que possible. Toutefois, comme il a été écrit plus haut, le partage des données n’est pas toujours envisageable pour des raisons éthiques, de sécurité ou encore de propriétés individuelle ou intellectuelle. En cas de doutes, l’INRA a développé un outil comprenant huit questions afin de conseiller les chercheurs sur la diffusion de leurs données (Andro et al., 2018) (http://www.bibliotheque-numerique.fr/DonneesDiffusables.php).  
Si le partage de données est partiellement ou entièrement impossible, les raisons doivent être explicitées dans le DMP. Le désengagement du projet pilote n’est pas nécessaire mais s’il y a lieu, les raisons du désengagement doivent elles aussi être explicitées dans le DMP.  
L’objectif du partage des données est l’accès direct et immédiat par un tiers aux données, facilitant la réutilisation (INIST, CNRS, 2018).  
Se pose également à cette étape la question de quelles données partager ? Sous-entendu : données brutes ou données traitées ? Cette question doit être étudiée en amont dans le DMP en fonction du public à qui seraient destinées les données et de la réutilisation qu’il pourrait en faire.

L’archivage des données, également réfléchit en amont dans le PGD, a aussi pour objectif de permettre l’accessibilité des données. Attention toutefois à ne pas confondre stockage et archivage : l’archivage se réalise à la fin du projet, avant qu’il ne soit totalement terminé et ne porte que sur la dernière version des fichiers de données contrairement au stockage qui a lieu au cours du projet et qui permet de conserver à court terme les différentes versions.  
L’objectif principal de l’archivage de données est la conservation de l’information sur le long terme. A cette étape, les risques principaux sont donc liés à la lisibilité, l’intégrité et l’intelligibilité des données. L’archivage pérenne nécessite donc de mettre en place une veille sur les supports et formats de fichiers, les systèmes d’exploitation et les technologies matérielles.

## Format des fichiers

Il existe une grande diversité de formats de fichiers liés généralement à la nature des données. Les principaux risques dans le choix du format sont ceux associés à la lisibilité du fichier. En effet, les mises à jour logicielles peuvent ne plus supporter le format, le format peut évoluer et les versions précédentes peuvent ne plus être disponibles.

Pour avoir des données FAIR il est donc conseillé de choisir des formats :

* Ouverts donc sans restriction d’accès ou de mise en œuvre
* Non propriétaires. En effet, bien que le risque d’obsolescence soit également réduit avec des formats ouverts propriétaires ceux-ci comporte un autre risque : l’éditeur peut décider de fermer le format (Pôle Données de la Recherche IST, 2015).
* Largement utilisés par la communauté scientifique (M.-C Jacquemot-Perbal, Ciolek-Figiel, 2016).
* Normalisés (ISO, W3C) (CINES, 2017).

Tim Berners-Lee propose un programme Open data 5 étoiles concernant les formats de fichiers destinés à la publication de données ouvertes dans le cadre de la normalisation W3C (Wolrd Wide Web Consortium) (Figure 12).



OL : Open Licence  
RE : Real Estructure ?  
OF : Open Format  
URI : Uniform Resource Identifier  
LD : Linked Open Data

Figure 12 : Programme Open data 5 étoiles, (Michael Hausenblas, 2012).

La première étoile correspond au fait de rendre disponibles ses données sur le web avec une licence ouverte. C’est donc une manière très simple de publier pour le chercheur toutefois les données sont à l’intérieur du document et donc non moissonables ou extractables par des machines.

La deuxième étoile correspond à des formats permettant la structuration des donnés et donc la lisibilité par les machines. Toutefois les données sont toujours à l’intérieur du document et il est possible de les extraire seulement avec un logiciel propriétaire. Néanmoins il est facile de publier des données de ce type.

La troisième étoile correspond à des fichiers de formats ouverts, non propriétaires permettant une plus large utilisation par autrui. Toutefois pour publier ce type de format il peut être nécessaire de passer par des convertisseurs si les données sont dans un format propriétaire à l’origine. De plus, un format type csv entraîne une perte d’information par rapport au xls ou xslx telles que les formules ou les commentaires. Pour palier cela, les informations perdues peuvent être listées au préalable dans un fichier de type « lisez-moi » (Filatre, Hensens, 2017).

La quatrième étoile correspond à des formats où les fichiers sont « dans » le web, c’est-à-dire qu’elles peuvent avoir des liens avec d’autres ressources qui pointent vers elles (woddiscovery, 2010). Cela est rendu possible grâce à un URI (Uniform Resource Identifier) qui est un identifiant universel étendant le principe de l’URL sans qu’il y ait forcément besoin d’une page web derrière. Ces identifiants ne sont pas tout à fait pérennes mais des mécanismes permettent de faire les conversions nécessaires. L’avantage de ce format est de pouvoir lier ces données à d’autres. Toutefois, les graphes RDF sont plus difficiles à comprendre pour un utilisateur lambda qu’un fichier xls par exemple. De plus un tel format demande plus de réflexion au chercheur et plus de travail.

Enfin la cinquième étoile correspond à « Linked Open Data » c’est-à-dire que les données sont reliées à d’autres données par des alignements dans le LOD Cloud (web de données ouvertes). Ainsi les données sont d’autant plus visibles et leur valeur s’en voit augmentée.

Les standards du W3C sont ouverts et extensibles (INES et al., 2011) mais ne sont pas toujours largement utilisés et sont réservés, à partir de la quatrième étoile à des personnes ayant des bases solides en informatique.

Les formats généralement les plus conseillés sont :

* Le PDF/A : Le PDF Archivage est un format normalisé ISO adapté à l’archivage, largement utilisé et ouvert. Bien que ce soit un format propriétaire, il est documenté. De plus, les métadonnées sont intégrées dans le document sous un format XMP. Ce format est considéré comme pérenne, en effet, ISO ne peut pas annuler ce standard (PDF Tools AG, 2016). Toutefois ces types de fichiers pouvant contenir aussi bien du texte que des images, la migration vers un autre format est plus difficile (Gautier Poupeau, 2010).
* Le csv semble également être un format considéré comme pérenne (Archives d’Etat, 2011) (Arms et al., 2013). Néanmoins, ce format peut entraîner des pertes d’informations notamment au niveau des commentaires et des formules. Il peut donc être nécessaire de le compléter avec un autre fichier où figureraient ces informations (Marie Claude Quidoz, 2018).
* Le xml, recommandé en 1998 par une publication du W3C (INES et al., 2011) est largement utilisé par la communauté scientifique et est lui aussi considéré comme pérenne (Rouchon, 2010) notamment s’il est associé avec un fichier xsd (Marie Claude Quidoz, 2018).

## Choix de la licence

La licence est un élément juridique définissant les modalités de partage (droits de copies et de diffusion) et de réutilisation des données (INIST, CNRS, 2018). Elle assure la préservation des droits d’auteur en imposant la citation de la source pour toute réutilisation.  
La communauté européenne recommande une licence Creative Commons notamment les licences CC 0 et CC BY. Ces licences sont gratuites et permettent aux utilisateurs des données de pouvoir les reproduire, les distribuer et les communiquer gratuitement (| Creative Commons France, [sans date]).  
Toutefois, l’attribution de ces licences est définitive. Il est possible de stopper la diffusion de l’œuvre sous licence Creative Commons ou encore de changer de licence mais dans ce cas toutes utilisations précédant ce changement resteront soumises aux conditions de la première licence. Il faut également faire attention aux DRM (Data Rights Management) pouvant être mis en place par certains éditeurs afin de restreindre l’accès et l’utilisation de l’œuvre. En effet, ces DRM sont des protections techniques, incompatibles avec les licences Creative Commons (Fily, 2015).

#### Les licences CC BY :

Avec une licence de ce type, l’utilisateur doit citer l’auteur de l’œuvre et indiquer les éventuelles modifications apportées.   
Il est possible d’ajouter différentes clauses aux licences CC BY :

* BY NC : Ce type de licence empêche que l’œuvre soit réutilisée à des fins commerciales. Pour toutes les licences Creative Commons, le partage de fichiers n’est pas considéré comme une réutilisation commerciale (Fily, 2015).
* BY SA : Ce type de clause oblige le partage de l’œuvre, même modifiée, dans les mêmes conditions c’est-à-dire sous une licence identique à celle choisie initialement par l’auteur. Avec cette clause il est donc impossible pour l’utilisateur de combiner ensuite deux œuvres ayant des licences différentes (Fily, 2015).
* BY ND : Cette licence autorise le partage et la diffusion de l’œuvre initiale mais n’autorise pas sa modification sauf autorisation préalable. De ce fait, une telle clause interdit la traduction de l’œuvre (Fily, 2015).

Ces trois clauses peuvent ensuite être combinés : il existe des licences BY NC ND qui n’autorisent ni diffusion commerciale ni modification de l’œuvre initiale et des licences BY NC SA qui n’autorisent pas l’utilisation commerciale non plus et qui obligent à partager d’éventuelles modifications sous les mêmes conditions de licence.  
Dans le cadre d’une démarche OpenScience, l’INRA, comme le CIRAD, déconseillent l’utilisation des clauses ND et NC (Pôle Données de la Recherche IST, 2018b), (Fily, 2015). Le CIRAD va même plus loin en déconseillant également la clause SA à cause des possibilités réduites de compilation de différents jeux de données.

La licence la plus utilisée dans le cadre d’une démarche Open Science semble être la licence CC BY 4.0. Cette licence permet le partage et la réadaptation du jeu de données, que ce soit à des fins commerciales ou non.

#### La licence CC 0 :

Cette licence permet de placer les données dans le domaine public, encourageant une large diffusion et réutilisation, même à des fins commerciales. Le titulaire des droits renonce alors à tous ses droits sur l’œuvre (INIST, CNRS, 2018). De ce fait, aucune restriction d’usage ne s’applique sur les données et l’auteur n’est pas obligatoirement cité (Fily, 2015).  
  
Plusieurs outils sont disponibles afin d’aider les auteurs dans le choix de la licence adéquate à leur jeux de données. Le site de Creative Commons propose par exemple une interface comportant plusieurs questions pour guider les auteurs dans leur choix (| Creative Commons France, [sans date]). Un assistant de licence EUDAT B2SHARE a également été développé (COMMISSION EUROPÉENNE - Direction Générale de la Recherche et de l’Innovation, 2016).

## Entrepôt de données et plateforme d’archivage

Un entrepôt de données est une base de données structurée où les auteurs peuvent déposer leurs données afin de les rendre visibles, accessibles, moissonables et citables. Le dépôt des données doit être réalisé le plus tôt possible, en même temps que la publication s’il y en a une. De plus, le choix de l’entrepôt et son type (institutionnel, thématique,…) doit être précisé dans le DMP. Avant de déposer ses données dans un entrepôt il est indispensable de s’assurer que toutes les étapes précédentes aient bien été respectées, à savoir : le nommage et l’organisation des fichiers, la documentation des données, le choix d’une licence, le choix d’un format de fichier pérenne et ouvert et l’attribution d’un identifiant pérenne et unique. De plus il faut également vérifier que le volume des fichiers est conforme à la taille maximale prise en charge par l’entrepôt choisi (DoRANum, 2018a).  
Si tel est le cas, il faut ensuite déterminer les procédures d’accès. L’accès aux données peut être immédiat ou non. En effet, il existe des procédures de mise sous embargo permettant de bloquer l’accès aux données pendant un certain temps afin de soumettre la publication associée aux données par exemple. D’autre part il est également possible de restreindre l’accès à des groupes spécifiques comme par exemple les membres du projet ou encore les membres de l’institution (INIST, CNRS, 2018).

Afin de choisir l’entrepôt le plus adéquat au dépôt des données, l’INRA propose le schéma suivant (Figure 13).

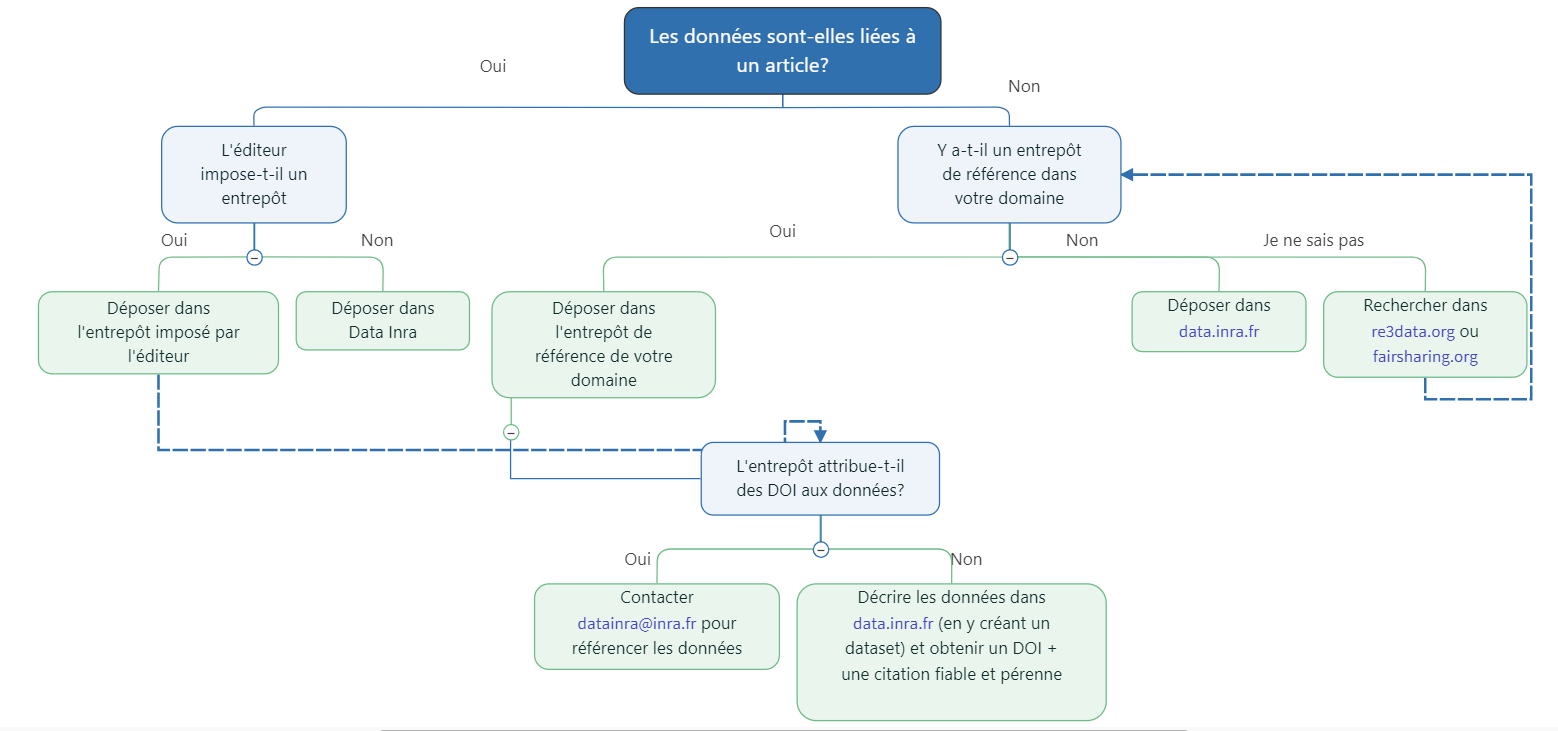


Figure 13 : Schéma d'aide au choix d'un entrepôt, (Pôle Données de la Recherche IST, 2018b).

En effet, si les données sont liées à une publication, il faut savoir si l’éditeur impose un entrepôt. Si ce n’est pas le cas alors les données peuvent être déposées dans Data Inra qui est un entrepôt institutionnel.   
Si les données ne sont pas liées à une publication il est conseillé de les déposer dans un entrepôt spécifique au domaine scientifique. Le cas échéant il est possible de consulter des annuaires d’entrepôts tels que re3data, fairsharingOAD, OpenDOAR ou encore CoreTrustSeal pour les entrepôts certifiés. Il existe en effet des entrepôts certifiés conformes aux directives d’OpenAIRE et répondant à des critères de qualité en ce qui concerne le stockage et l’intégrité des données. Ces entrepôts garantissent l’interopérabilité, l’accessibilité et le fait que les données soient moissonnables (INIST, CNRS, 2018).  
Toutefois, pour le choix de l’entrepôt, il faut également prendre en compte les recommandations du financeur s’il y en a. De plus il semble important de vérifier que l’entrepôt soit connu de la communauté scientifique et que le moteur de recherche soit performant afin de garantir la visibilité de ses données (DoRANum, 2018a) (Figure 14).

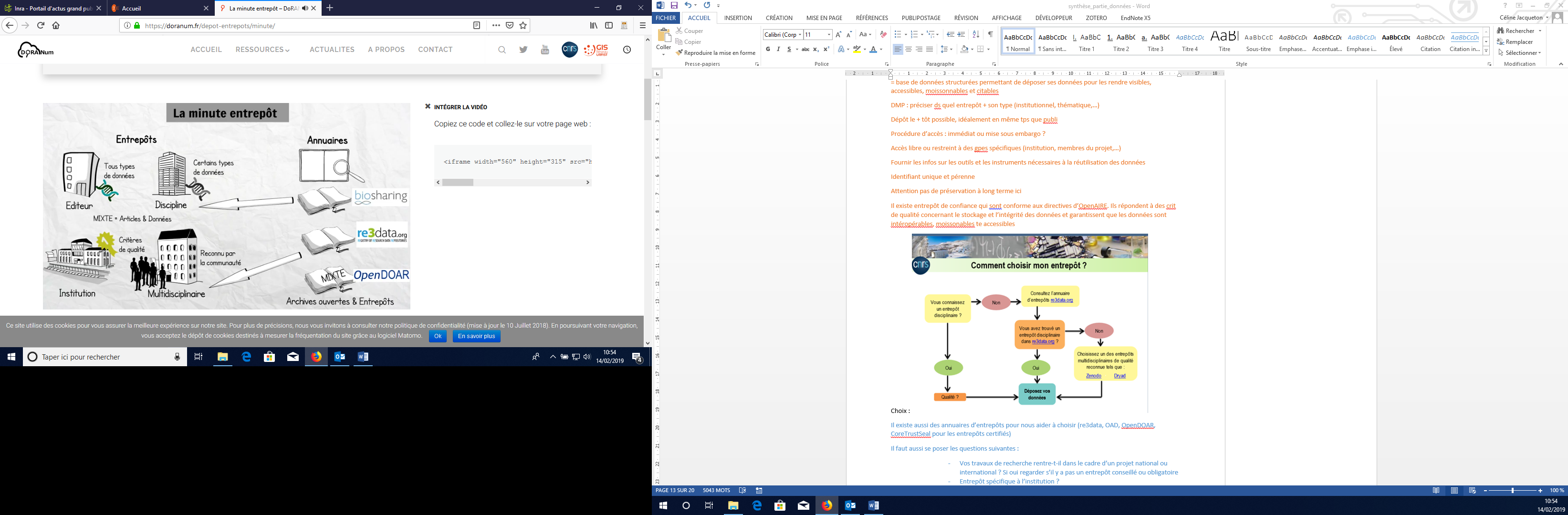


Figure 14 : Schéma présentant la démarche du choix d'un entrepôt, (DoRANum, 2018a).

Certains entrepôts permettent également l’archivage pérenne des données. Cependant, si ce n’est pas le cas de l’entrepôt choisit il faut soumettre les fichiers de données à une plateforme d’archivage. Il est principalement attendu d’une plateforme d’archivage qu’elle garantisse la sécurité et l’intégrité physique des données en mettant en place des veilles technologiques. Elle doit également pouvoir traiter de gros volumes de données.   
Il est recommandé de choisir une plateforme suivant le modèle OAIS (Open Archival Information System) qui est un modèle conceptuel de référence, aujourd’hui devenu une norme ISO (ISO 14721). Ce modèle est abstrait et définit « l’architecture logique et les fonctionnalités d’un système d’archivage » (CINES, 2016). Il recense six fonctions d’archivage (INIST, CNRS, 2018), (CINES, 2016) :

* Entrée : réception et contrôle des objets à archiver. Les données sont ensuite transmises à l’entité « stockage » et leur documentation à l’entité « gestion de données ».
* Stockage
* Gestion de données
* Administration : pilote le système et établit les règles internes.
* Planification de la pérennisation : assure la veille technologique et prépare et planifie les éventuelles évolutions.
* Accès : interface directe avec les utilisateurs leur permettant de trouver l’objet qu’ils cherchent.

Et quatre acteurs principaux :

* L’archive : acteur interne, opérateur du système d’archivage.
* Le management : décideur politique.
* Les producteurs : fournissent les objets à archiver.
* Les utilisateurs : peuvent accéder aux objets archivés.

Ainsi que différent type d’informations :

* Information de représentation qui peut être une information de structure ou une information sémantique
* Information de pérennisation : informations de provenance, de contexte, d’identification et d’intégrité
* Information d’empaquetage : relie les informations de pérennisation et de représentation
* Information de description : transmise à l’entité « gestion de données », elle permet de constituer le référentiel de base de données.

Il existe également une norme AFNOR NF Z42-013 décrivant les exigences opérationnelles et techniques à mettre en œuvre au sein d’un système d’archivage électronique ainsi qu’une norme de certification ISO 16-363.  
En complément des différentes certifications existantes, un outil en ligne permettant l’évaluation d’une archive numérique est disponible : Data Seal of Approval. De plus, il existe des normes sur l’expression des métadonnées en vue d’un archivage telles que METS et PREMIS.  
Enfin le CINES propose également un outil permettant de tester la validité des formats de fichiers en vue d’un archivage. De plus cet organisme dispose d’une plateforme d’archivage pérenne, PAC, certifiée et conforme aux normes OAIS et AFNOR Z42-013 (Pôle Données de la Recherche IST, 2018a) (INIST, CNRS, 2018).

# Valoriser les données

## Data paper

Un data paper est un document associé à un identifiant décrivant un ou plusieurs jeux de données publié sous la forme d’un article évalué par les pairs dans un journal scientifique. La publication d’un data paper a pour but « d’informer la communauté scientifique de l’existence et la disponibilité d’un jeu de données », (Cirad, 2018) et permet donc d’augmenter la visibilité du jeu de données ainsi que sa réutilisation (Pôle Données de la Recherche IST, 2018f).   
Toutefois, ce concept assez récent n’est pas toujours nommé de la même manière selon les journaux scientifiques : data notes, data article, data descriptor, data report,… (Pôle Formation IST, 2018).  
D’autre part, la structure du data paper est également variable selon les revues (Figure 15).

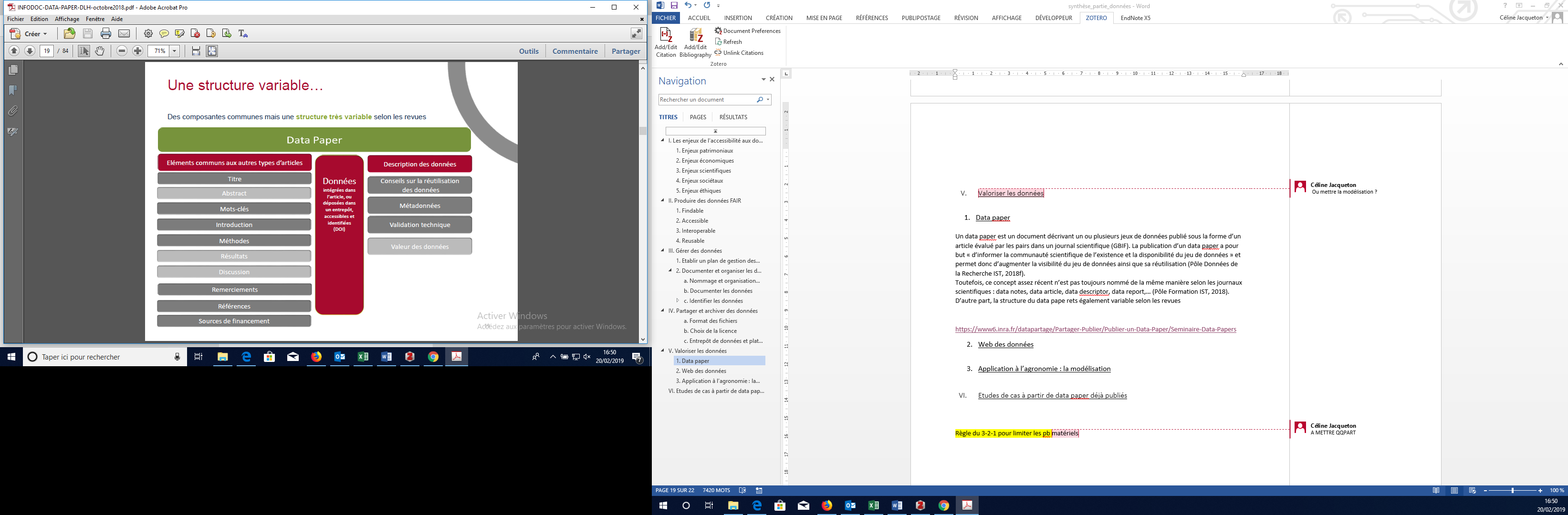


Figure 15 : Structure générale d'un data paper, (Pôle Formation IST, 2018).

En effet, l’abstract n’est pas obligatoirement présent, de même pour l’introduction. Cependant les méthodes sont toujours expliquées et l’on y trouve également des remerciements, des références ainsi que les sources de financement. Contrairement à l’article scientifique, la partie la plus importante du data paper porte sur la description des données et notamment sur leur réutilisation potentielle. Souvent le contenu comprend un tableau récapitulatif des données après l’abstract. Certains data paper contiennent le jeu de données alors que d’autres renvoient à celui-ci via un lien pérenne. Dans ce cas, le jeu de donnée renvoie lui aussi au data paper également via un lien pérenne.

Beaucoup de revues scientifiques classiques publient des data paper mais il existe également des revues dédiées en libre accès. Certaines de ces revues offrent des modèles ou des outils afin d’aider la rédaction et guident les chercheurs dans leur démarche en proposant des entrepôts. Les revues de publication peuvent être choisies selon :

* leur notoriété
* la langue de publication souhaitée
* leur fonctionnement notamment en termes de délais et de qualité de révision par les pairs
* leur visibilité : open accès, référencement dans les bases de données, moteurs de recherche performants,…
* les outils mis à disposition

Le choix de la revue va impacter sur la localisation des données (intégrées au data paper ou non), les modalités de diffusion des données et la présentation de ces dernières ainsi que des métadonnées qui les accompagnent.   
L’INRA met à disposition un fichier excel recensant un grand nombre de revues publiant des data paper avec leurs modalités de fonctionnement et les frais liés à la publication.

Pour qu’un data paper soit accepté, les critères classiques concernant la qualité générale du manuscrit sont évalués à savoir les citations, le respect des instructions ainsi que plusieurs points concernant les données :

* leur importance et leur originalité
* leur valeur de réutilisation
* leur qualité et leur fiabilité
* leur accès
* la qualité et la rigueur de la méthode de collecte
* le choix des métadonnées et des formats

Le plus gros frein à la publication d’un data paper réside essentiellement dans son coût.

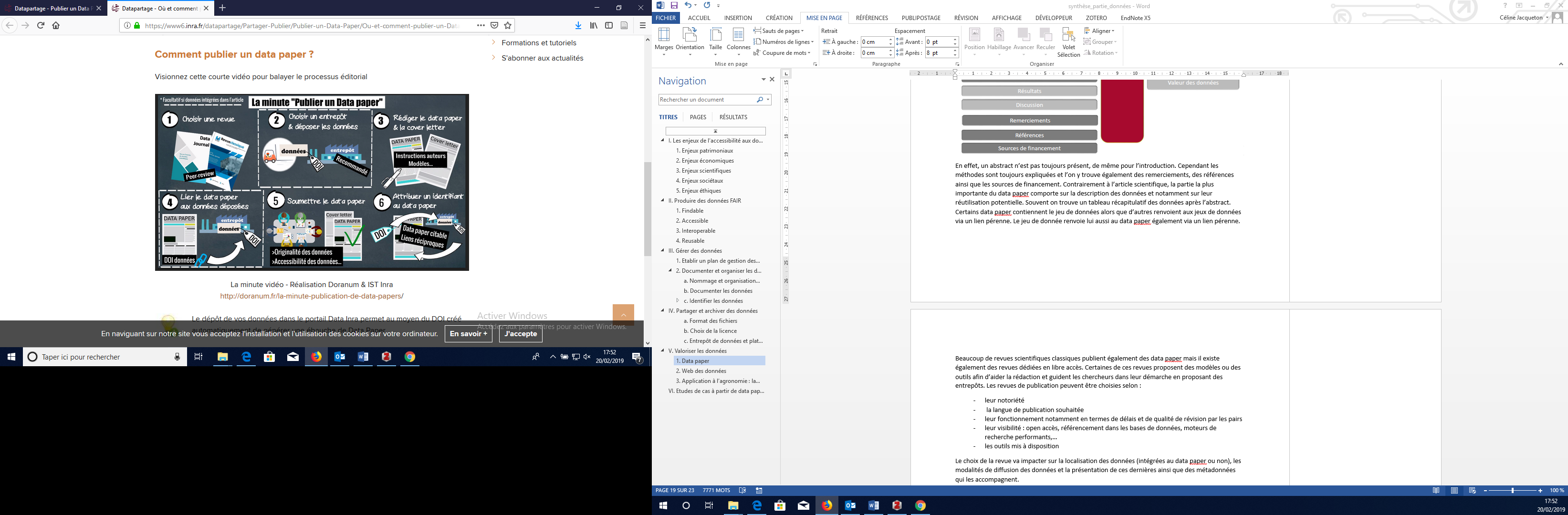


Figure 16 : Démarche de publication d'un data paper, (Pôle Données de la Recherche IST, 2018g).

## Web des données

Le web sémantique ou web des données est une évolution du web proposée par Tim Berners-Lee en 2001 et standardisé par le W3C. Ce concept permet de favoriser la découverte et la réutilisation de données en proposant des standards d’échange (respectant les principes FAIR) en termes de formats et de protocoles s’appuyant sur le modèle RDF (Resource Description Framework). Le web sémantique est en fait un web intelligent qui comprend les informations plutôt que de les stocker, permettant de mieux répondre aux attentes des utilisateurs (Dunod, 2008).

Publier dans le web sémantique apporte donc une plus-value aux données qui sont d’autant plus interopérables et accessibles. De plus le web sémantique étant considéré par certains comme l’évolution future du web d’aujourd’hui, l’utiliser d’ors et déjà peut être une bonne forme de valorisation des données.

Le web sémantique s’appuie sur les principaux standards suivants (Bibliothèque nationale de France, 2017), (Pôle Données de la Recherche IST, 2018d) :

* Les URI (Uniform Resource Identifier) dont nous avons déjà parlé plus haut. Ces identifiants permettent d’accéder directement à la ressource mais étant une extension du principe des URL ils ne sont pas tout à fait pérennes. Toutefois il existe des mesures permettant une meilleure gestion de leur pérennité, on les appelle alors « URI http ».
* Le modèle RDF : ce modèle est un standard pour l’échange de données et a pour but de faciliter leur diffusion. L’élément de base est le triplet, il se compose de : sujet, prédicat, objet. Le RDF permet de lier les données et fournit une organisation qui peut être requêtée par les machines.
* L’OWL (Web Ontology Language) : ce langage permet la représentation des connaissances sur des classes et/ou sur des propriétés afin de les intégrer dans des environnements informatiques et pouvoir exercer des raisonnements dessus.
* SKOS (Simple Knowledge Organization System) : c’est un « modèle de données standards pour les systèmes d’organisation de connaissances » (Pôle Données de la Recherche IST, 2018d) tels que les vocabulaires contrôlés ou les thésaurus.
* SPARQL (Simple Protocole and RDF Query Language) : c’est un langage de requête pour un graphe RDF. Il permet de rechercher, modifier, ajouter ou supprimer des données RDF sans pour autant les télécharger.

Il existe un guide des bonnes pratiques pour publier des données sur le web disponible à l’adresse suivante : <https://www.w3.org/TR/dwbp/#ProvideMetadata> dont les grandes lignes reprennent ce qui a été déjà dit dans ce document en termes de gestion et de partage des données

## Application à l’agronomie : la modélisation

# Etudes de cas à partir de data paper déjà publiés

## A global experimental dataset for assessing grain legume production, Cernay, Pelzer and Makowski, 2016

### Généralités sur le journal

Ce data paper est issu du journal « Scientific data » qui est en open access et proposé par le groupe de publication Nature. La publication n’a d’ailleurs pas le nom de data paper mais celui de « Data Descriptor ». Selon le journal, ce type de publication doit aider à la réutilisation des données et permettre la reconnaissance de ceux qui publient. Ces publications sont automatiquement indexée dans PubMed, MEDLINE, Google Scholar et Clarivate’s Web et automatiquement déposés dans PubMed Central qui est une archive ouverte. Le coût de publication s’élève à 1305 euros.

Les data descriptor combinent du contenu narratif traditionnel et une description structurée des données, permettant de fournir un nouveau cadre de partage et de réutilisation des données. Les principes du data descriptor s’alignent sur les principes FAIR.

Selon le journal la structure d’une telle publication s’articule de la manière suivante :

* Titre (110 caractères maximum)
* Abstract (170 mots maximum, sans références) comprenant un tableau récapitulatif avec des colonnes portant sur le type de conception du jeu de données, le type de mesures, de technologies, de facteurs ainsi que sur les caractéristiques de l’échantillon.
* Eléments de contexte (700 mots maximum)
* Méthodes
* Données
* Validation technique
* Notes d’utilisation
* Références
* (Remerciements)

Afin de faire un choix entre différents journaux, le site présente quelques chiffres clés :

* Facteur d’impact (rapport entre le nombre de citations et le nombre d’articles publiés) sur 2 ans : 5.305
* Facteur d’impact sur 5 ans : 5.862
* Indice d’immédiateté : 0.843
* Note d’influence de l’article : 2.597
* Eigenfactor® score: 0.00855

### Contenu du data paper

Concernant la structure, la partie la plus développée dans ce data paper est la partie sur les méthodes. Le data paper comprend au total 7 figures et 2 tableaux.

L’abstract se compose d’une partie rédigée décrivant le contenu et la structure du jeu de données ainsi que le tableau récapitulatif décrit précédemment.

La partie « background and summary » est en fait une sorte d’introduction où les auteurs présentent le contexte, définissent les mots clés et présentent l’intérêt des légumineuses ainsi que quelques chiffres clés.  
Cette partie démontre également l’intérêt d’un tel jeu de données et le décrit brièvement, notamment à l’aide de figures comprenant une carte et des tableaux récapitulatifs. Les possibilités de réutilisation du jeu de données sont abordées par la suite ainsi que les éventuelles possibilités de le compléter.

La troisième partie du data descriptor concerne les méthodes. Les données publiées étant issues de la littérature, les auteurs indiquent le type de journaux dans lesquels ils ont effectués leurs recherches ainsi que la démarche effectuée pour arriver à leurs résultats.   
Cette partie permet également de décrire l’organisation de la base de données ainsi que d’expliquer certaines valeurs pouvant susciter des interrogations telle que la valeur « NULL ». De plus, ils abordent les sources d’erreurs possibles liées à leur démarche et en donnent une estimation chiffrée.  
Enfin, les fichiers supplémentaires disponibles sont décrits et localisés.

La partie suivante du data descriptor concerne les données en tant que telles. Ici, les auteurs ont décrits les attributs principaux de la table contenant le plus de données expérimentales extraites de la littérature. Les attributs des autres tables sont décrits dans un fichier annexe disponible au téléchargement.  
Cette description inclue la présentation des calculs effectués, la justification des choix pris au vue des valeurs ainsi que la façon de procéder lorsque certaines valeurs manquaient.   
Cette partie comprend également des figures résumant la distribution de plusieurs variables pour un grand nombre d’espèces de légumineuses afin de mieux rendre compte de la diversité du jeu de données.

La cinquième partie de cet article est consacrée à la validation technique des données. Elle sert à exposer les mécanismes mis en œuvre afin de s’assurer de la véracité des données publiées.  
Dans le cas étudié, les auteurs évoquent par exemple que les articles sélectionnés dans la littérature ont été relus plusieurs fois, les données reportées dans les tables ont elles aussi été vérifiées à plusieurs reprises et des requêtes ont été mise en place afin de s’assurer qu’aucune aberration n’en ressortait. Afin de s’affranchir des doutes sur la valeur de certaines données, 17 auteurs ont été contactés.

La partie suivante porte sur les possibilités de réutilisation des données. Les auteurs rappellent brièvement la source de leurs données et proposent plusieurs exemples de réutilisation détaillés avec la marche à suivre. Cette partie permet également d’aborder l’ajout de nouvelles données à la base de données. Pour cela les auteurs indiquent plusieurs possibilités de données à intégrer à leur base en détaillant les solutions techniques. D’autre part un exemple de requête SQL est fourni ce qui permet de faciliter l’utilisation d’une telle base de données par un individu quelconque.

La dernière partie obligatoire du data paper porte sur les références. Ici, les auteurs en donnent 188 (sachant que leur base de données a été créée à partir de 173 articles).  
A la fin de l’article se trouve une partie «  informations supplémentaires » qui permet notamment de donner la licence sous laquelle les données sont publiées.

## Data on spatio-temporal representation of mineral N fertilization and manure N application as well as ammonia volatilization in French regions for the crop year 2005/06, Génermont et al., 2018

### Généralités sur le journal

Ce data paper est issu du journal Data in Brief proposé par le groupe ELSEVIER permettant de faciliter le partage et la réutilisation des jeux de données, quelque soit le domaine scientifique auquel ils appartiennent. Pour ce faire, ce type d’article permet une description détaillée des données et une recherche facilitée qui contribue à augmenter le nombre de citations et de nouvelles collaborations.  
Ce journal est en open access (l’article peut être librement lu, téléchargé, copié et distribué) et est examiné par les pairs. Le coût d’une telle publication s’élève à 500 USD (dollar américain) soit environ 450 euros hors taxes. Toutefois, ELSEVIER a mis en place des collaborations avec des organismes de financement ce qui permet aux auteurs de se voir rembourser les frais de publication à condition de se conformer aux politiques de ces organismes.  
Data in brief utilise un contrat de licence exclusif qui permet d’attribuer des droits d’édition et de distribution à l’éditeur tout en accordant la conservation des droits d’auteurs.

Les chiffres clés présentés par le site sont les suivants :

* Taux de citation : 0.7
* SNIP (Source Normalized Impact per Paper) : 0.287
* SJR (SCImago Journal Rank) : 0.341
* Taux d’acceptation : 77%

Pour publier un article dans Data in Brief, les données doivent avoir été produites et appartenir à l’auteur ou à son institution.

Le structure d’un data paper publié dans Data in Brief s’articule de la manière suivante :

* Titre, auteurs et labo
* Abstract : purement descriptif, il donne la localisation des données et se doit de citer l’article scientifique lié aux données le cas échéant.
* Tableau des spécifications :

|  |  |
| --- | --- |
| Subject area | *E.g., physics, chemistry, biology, economics, psychology* |
| More specific subject area | *Describe narrower subject area.* |
| Type of data | *Table, image (x-ray, microscopy, etc.), text file, graph, figure* |
| How data was acquired | *Microscope, survey, SEM, NMR, mass spectrometry, etc.; if an instrument was used, please give the model and make.* |
| Data format | *Raw, filtered, analyzed, etc.* |
| Experimental factors | *Brief description of any pretreatment of samples/data* |
| Experimental features | *Very brief experimental description* |
| Data source location | *City, country,institution and/or latitude and longitude (and GPS coordinates) for collected samples/data.* |
| Data accessibility | *State if data is with this article or in public repository; if public repository, please explicitly name repository and data identification number, and provide a direct URL to data. We recommend* [***Mendeley Data***](https://data.mendeley.com/)*if you do not have a trusted repository.* |
| Related research article | *If your data article is submitted as a companion paper to a research article, please cite your associated research article here; you may reference this as “in press” [1].*  *If this is a direct submission to* Data in Brief*, please cite the most relevant research article here from your reference list.* |

Tableau 1 : Indications sur le tableau des spécifications, (ELSEVIER, 2018)

* Valeur des données : 3 à 5 points expliquant en quoi les données ont de la valeur pour la communauté scientifique, comment elles peuvent être réutilisées, avec quoi elles peuvent être comparées.
* Données : description brève du jeu de données inclus dans l’article (env. 250 mots).
* Plan expérimental, matériels et méthodes : description complète de la manière dont ont été obtenues les données.
* (Remerciements)
* Références

### Contenu

Tout comme l’exemple précédent, la partie principale de ce data paper est celle portant sur le matériel et méthodes. Il comprend au total 2 tableaux. D’autres issus du même journal ne contiennent pas toujours de figures/tableaux mais certains sont plus fournis avec par exemple 2 tableaux et 3 figures (Oyeyemi et al., 2018).

L’abstract répond aux exigences du journal et permet également dans cet exemple de citer la licence utilisée.

Mis à part la ligne concernant les facteurs expérimentaux, le tableau des spécifications est rempli selon les indications du journal.

Les points concernant la valeur des données décrivent les différentes réutilisations possibles et en expliquent les intérêts. Une phrase supplémentaire indique que les auteurs sont ouverts à toute éventuelle collaboration.

La partie sur les données comprend une phrase de contexte avant de décrire les données. Les données supplémentaires sont également décrites. Cette partie est essentiellement composée de 2 tableaux descriptifs du jeu de données.

Le matériel et méthode fait référence à l’article associé au data paper où les données et calculs y sont déjà largement décrits. Les lignes suivantes servent à citer la source des données brutes ainsi qu’à expliquer les démarches suivies concernant le traitement des données.

Vient ensuite les remerciements, les liens vers les données supplémentaires ainsi que les références qui sont au nombre de 10 dans cet exemple. Ce chiffre est largement inférieur à celui du document précédent ce qui peut être expliqué par la provenance des données.

## Fifteen-year record of soil temperature at the Bear Brook Watershed in Maine, Patel et al., 2018

Cet article est également issu de Scientific data (Nature) mais se différencie du premier exemple car les données ne sont pas organisée en base de données ni issues de recherches dans la littérature. Ces données sont issues d’une expérimentation à long terme sur l’évolution des températures dans les sols. En termes de figures et de tableaux, ce data record en contient 4 de chaque. Les parties sont similaires au premier exemple, toutefois quelques points différent.  
En effet, la partie « background and summary » ne traite pas des possibilités de réutilisation du jeu de données.  
La partie sur les méthodes décrit d’abord le matériel utilisé, expose les méthodes mises en œuvre afin d’éliminer les valeurs aberrantes, les méthodes de contrôle de cohérence, le traitement des données et les méthodes de traitement des valeurs manquantes.

La partie suivante sur le jeu de données est assez brève et sert à détailler le contenu des fichiers disponibles. Cette partie comporte une figure représentant les valeurs moyennes des températures journalières au cours des années 2001 à 2016 pour les 4 compartiments étudiés (à savoir l’air, la surface organique du sol, la surface minérale du sol à 10 et 20cm de profondeur).

Afin de s’assurer de l’exactitude des données publiées, la validation technique comprend la calibration et la vérification des appareils, la vérification de la corrélation des capteurs entre eux ainsi que la vérification qu’il n’y ait pas de biais du au nombre de réplications. De plus, les maximum et minimum de température journaliers sont été comparées avec ceux obtenus dans une station voisines afin d’éviter toute valeur erronée.

La partie sur la réutilisation des données propose des exemples de réutilisation comme c’est le cas dans le premier article étudié. Toutefois, les solutions techniques ne sont pas détaillées. Cette partie permet également d’énoncer les limites du jeu de données.

L’article comporte 39 références et le type de licence est également donné dans la partie « informations supplémentaires ».

Le tableau ci-dessous (*Tableau 2*) propose une comparaison générale des deux revues étudiées dans ces études de cas (à savoir Data in Brief et Scientif data).

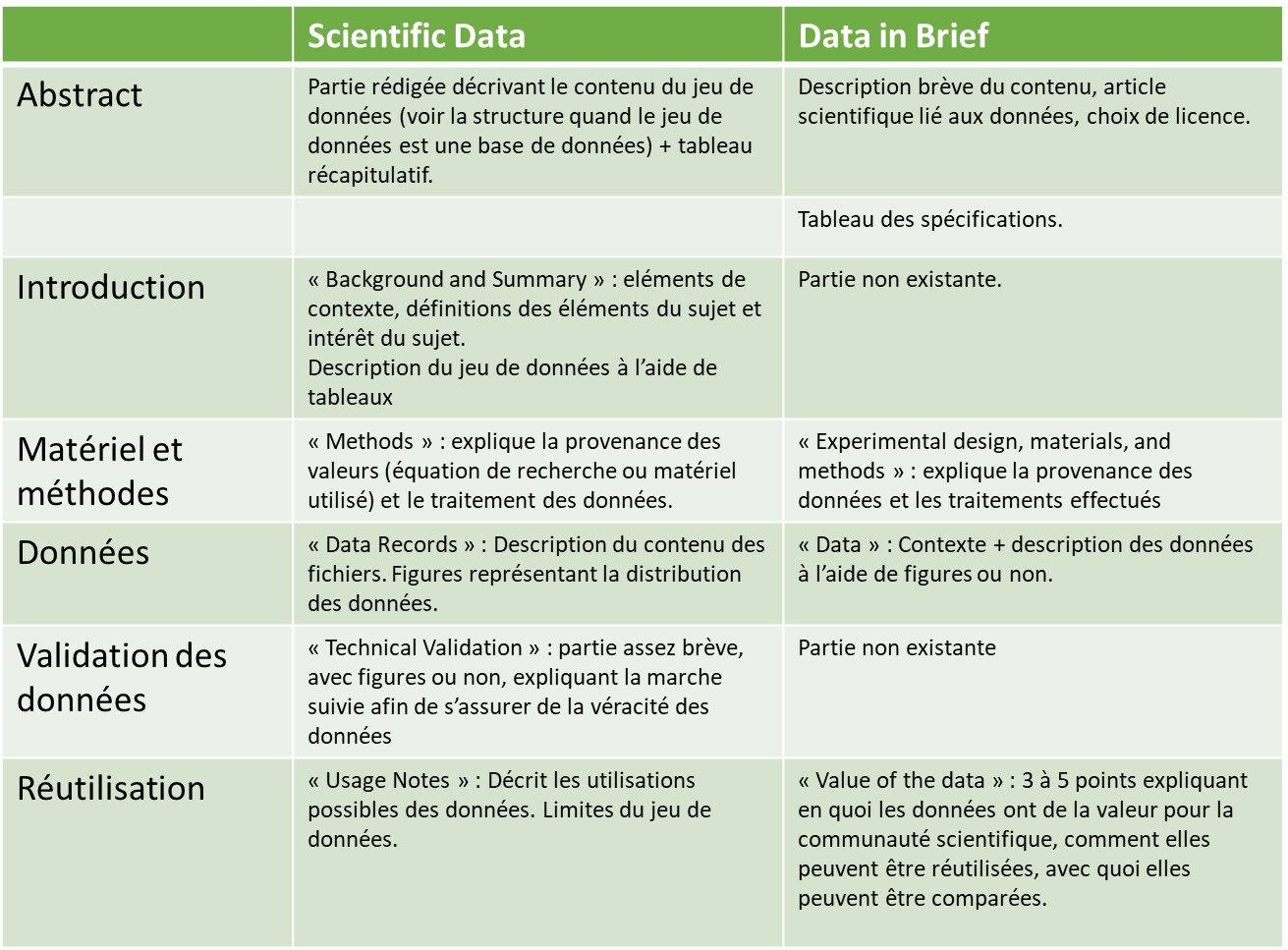


Tableau 2 : Comparaison du contenu attendu pour les journaux Scientific Data et Data in Brief

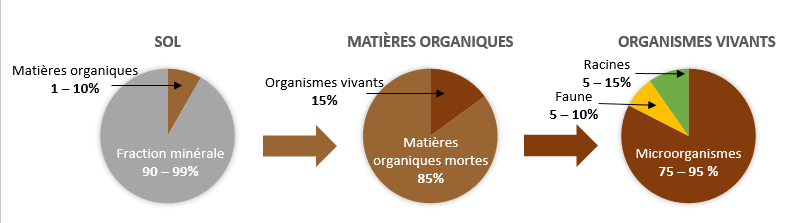
##### Conclusion

Afin de s’inscrire dans une démarche Open science et donc de produire des données FAIR, la première étape indispensable est l’établissement d’un plan de gestion des données qui décrira les grandes lignes des différentes étapes de leur cycle de vie. Dès la collecte des données, il est nécessaire de définir les données pouvant être partagées de celles devant rester confidentielles. Le nommage et l’organisation des fichiers ainsi que la documentation des données qu’ils contiennent à l’aide de métadonnées, idéalement issues de standards, doivent également être réalisés dès la collecte. Il est nécessaire que les fichiers soient dans un format le plus pérenne possible adapté aux données qu’ils contiennent. Tout au long de la collecte et du traitement des données, les fichiers doivent être stockés et sauvegardés dans un entrepôt de stockage par exemple. A la fin du projet, les données doivent être archivées de manière pérenne dans un entrepôt prévu à cet effet ou une plateforme d’archivage. Il est possible de valoriser son travail en publiant ses données dans un journal ou encore sur le web des données (annexe 3).

##### Perspectives

Concernant notre projet, les données sont déjà acquises et tout le travail de gestion repose donc sur la collecte des informations nécessaires. Les fichiers devront être nommés et organisés selon les principales règles énoncées dans les parties correspondantes. Les métadonnées seront décrites à l’aide de standards tel que celui établit par Dublin Core. Toutefois, ces standards sont assez généraux. Afin d’augmenter le potentiel de réutilisation des données, il sera demandé aux contributeurs du projet de fournir des informations sur les méthodes utilisées pour l’acquisition de leurs données ainsi que sur le matériel. Idéalement nous utiliserons du vocabulaire contrôlé ou des thésaurus provenant de l’AgroPortal. Il existe notamment un thésaurus : « anaeeThes » qui pourrait correspondre. L’organisation des données et donc par conséquent le format de fichier restent encore à définir. Les fichiers seront à minima des fichiers Excel en format csv mais les données pourront également, de façon plus ambitieuse, être déposées dans une base de données ouvertes au langage SQL. Le choix de la licence et des entrepôts de stockage et d’archivage n’est pas encore établit à ce stade, bien que DataInra semble être en terme d’entrepôt la solution la plus simple et la plus adéquate au projet. Enfin, concernant la valorisation des données, celles-ci seront publiées dans un data paper, l’objectif étant de viser le journal Scientific Data en premier lieu. La publication sur le web des données semble être un peu ambitieuse au vue de nos compétences en la matière. Toutefois, ces données seront doublement valorisées car une partie d’entre elles serviront à calibrer les modèles de décompositions des résidus végétaux disponibles sur la plateforme Virtual Soil. Une collaboration avec la communauté STICS est également envisagée en termes de modélisation. Cependant, la communauté n’étant pas Open Science, cette collaboration passera au second plan.

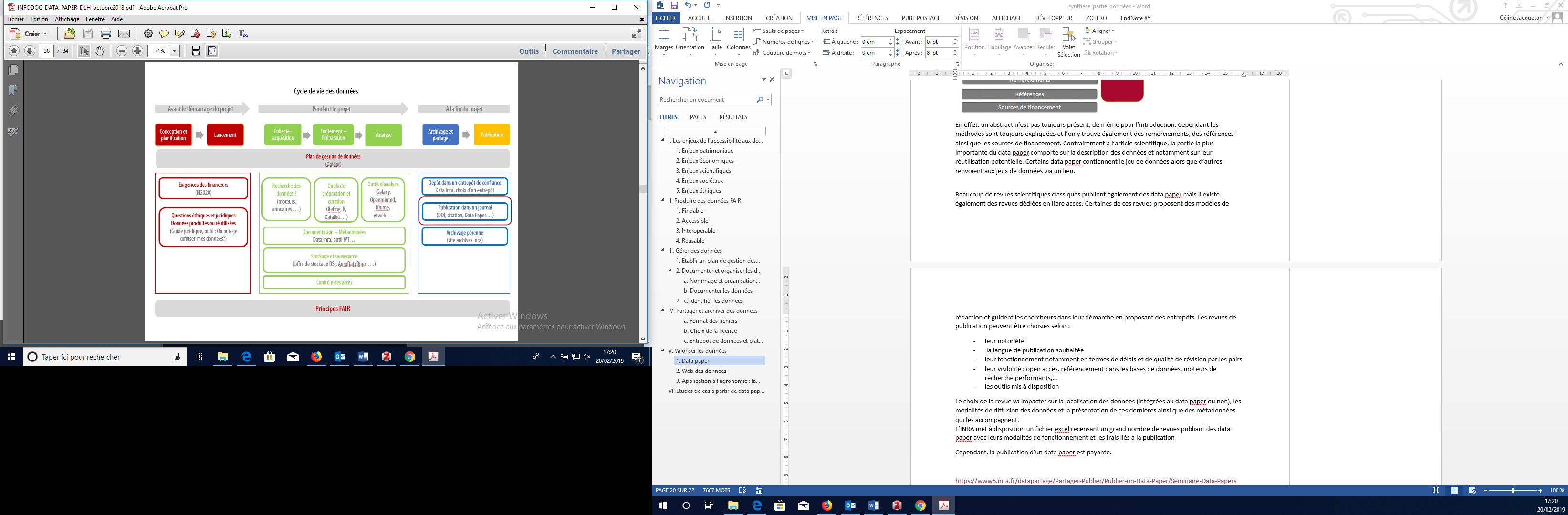
##### Annexes



Annexe 1 : Proportion approximative des constituants des sols, d’après (Marie Sauvadet, 2016)

|  |  |
| --- | --- |
| Elément du PGD | Questions à examiner |
| 1. **Résumé des données** | • Indiquer l’objectif de la collecte/génération de données.  • Expliquer la relation par rapport aux objectifs du projet.  • Préciser les types et formats des données générées/collectées.  • Préciser si des données existantes sont réutilisées (le cas échéant).  • Préciser l’origine des données.  • Indiquer le volume prévu des données (s’il est connu).  • Mettre en évidence le potentiel de réutilisation des données : à qui seront-elles utiles ? |
| 1. **Données FAIR**   2.1 Rendre les données trouvables, y compris la fourniture des métadonnées | • Souligner la trouvabilité des données (fourniture des métadonnées).  •Décrire la procédure d’identification des données et faire référence à un système standard d’identification. Utilisez-vous des identifiants pérennes et uniques tels que les DOI ?  • Définir les conventions de nommage utilisées.  • Présenter la démarche de recherche par mot-clé.  • Décrire l’approche utilisée pour un versionnage explicite.  • Préciser les standards de création de métadonnées (le cas échéant). Si aucun standard n’existe dans votre discipline, décrire comment les métadonnées seront créées ainsi que leur type. |
| 2.2 Rendre les données librement accessibles | • Préciser quelles données seront librement disponibles. Si certaines données ne sont pas mises à disposition, le justifier.  • Préciser comment les données seront rendues disponibles.  • Préciser quelles techniques ou outils logiciels sont nécessaires pour accéder aux données. La documentation sur le logiciel nécessaire pour accéder aux données est-elle fournie ? Est-il possible d’intégrer le logiciel adéquat (p. ex. en code open source) ?  • Préciser où les données et les métadonnées associées, la documentation et le code sont déposés.  • Préciser comment l’accès sera fourni au cas où des restrictions s’appliquent. |
| 2.3 Rendre les données interopérables | • Évaluer l’interopérabilité de vos données. Préciser quels vocabulaires décrivent les données et métadonnées, quels standards ou méthodologies seront appliqués pour faciliter l’interopérabilité.  • Préciser si vous utiliserez du vocabulaire standard pour tous les types de données présents dans votre jeu de données pour permettre une interopérabilité interdisciplinaire. Dans le cas contraire, fournirez-vous un alignement avec les ontologies les plus fréquemment utilisées ? |
| 2.4 Accroître la réutilisation des données (au moyen de licences explicites) | • Préciser quelle licence sera attribuée à vos données afin de permettre la réutilisation la plus large possible.  • Préciser quand les données seront accessibles pour leur réutilisation. Le cas échéant, préciser pour quelle raison et pendant combien de temps un embargo sur les données est nécessaire.  • Préciser si les données produites et/ou utilisées dans le projet sont exploitables par des parties tierces, en particulier après la fin du projet. Si la réutilisation de certaines données est restreinte, expliquer pourquoi.  • Décrire les processus d’assurance qualité des données.  • Préciser la durée pendant laquelle les données resteront réutilisables. |
| 1. **Allocation des ressources** | • Estimer les coûts pour rendre vos données conformes aux principes FAIR. Décrire comment vous comptez financer ces coûts.  • Identifier clairement les responsabilités de gestion des données dans votre projet.  • Décrire les coûts et la valeur potentielle de la conservation à long terme |
| 1. **Sécurité des données** | • Examiner la récupération des données ainsi que le stockage sécurisé et le transfert de données sensibles. |
| 1. **Aspects éthiques** | • Pour être couvert dans le cadre de l’évaluation éthique, la section éthique de la Description de l’action et les livrables éthiques. Inclure des références et les aspects techniques liés s’ils ne sont pas couverts par la précédente évaluation. |
| 1. **Autres** | • Mentionner les autres procédures nationales/sectorielles/départementales/de financeurs pour la gestion des données que vous utilisez (le cas échéant). |

Annexe 2 : Tableau récapitulatif des questions à examiner dans le PGD – selon (COMMISSION EUROPÉENNE - Direction Générale de la Recherche et de l’Innovation, 2016)

  
Annexe 3 : Schéma récapitulatif du cycle de vie des données FAIR – selon (Pôle Formation IST, 2018)

##### Bibliographie

| CREATIVE COMMONS FRANCE, [sans date]. 6 LICENCES gratuites. In : [en ligne]. [Consulté le 15 février 2019]. Disponible à l’adresse : http://creativecommons.fr/licences/.

ADEME, 2018. Les sols en danger. In : *ADEME* [en ligne]. 7 avril 2018. [Consulté le 13 juin 2019]. Disponible à l’adresse : https://www.ademe.fr/en/particuliers-eco-citoyens/dossiers-comprendre/dossier/sols-tresor-a-proteger/sols-danger.

AFNOR, 2004. *XP U 44-162*. 11 juin 2004. S.l. : s.n.

AGRONUTRITION, 2014. Biodiversité des sols : tout savoir sur son intérêt pour l’agriculture. In : *Biofertilisants.fr* [en ligne]. 1 avril 2014. [Consulté le 13 juin 2019]. Disponible à l’adresse : http://www.biofertilisants.fr/comprendre-les-biofertilisants/biodiversite-des-sols-tout-savoir-sur-son-interet-pour-agriculture/.

ALAVOINE, Gonzague et LASHERMES, Gwenaëlle, 2018. *TP « Biodégradabilité des résidus de culture dans le sol » - Détermination de la teneur en biomasse microbienne du sol par fumigation-extraction*. 2018. S.l. : s.n.

ALLISON, Franklin E., 1973. *Soil organic matter and its role in crop production*. Amsterdam, New York : Elsevier Scientific Pub. Co. Developments in soil science, 3. ISBN 978-0-444-41017-7. S598 .A5997

AMIN, Bilal Ahmad Zafar, 2012. *Rôle des enzymes lignocellulolytiques dans le processus de biodégradation de résidus végétaux dans les sols: Influence de la qualité des résidus sur l’efficacité des enzymes et leur dynamique*. Reims : Université de Reims Champagne-Ardenne.

ANDRO, Mathieu, MORCRETTE, Nathalie et GANDON, Nathalie, 2018. Système expert d’aide à la decision pour diffuser les donnees de la recherche. In : [en ligne]. 16 octobre 2018. [Consulté le 19 février 2019]. Disponible à l’adresse : http://www.bibliotheque-numerique.fr/DonneesDiffusables.php.

ARCHIVES D’ETAT, 2011. *Formats de fichiers adaptés à l’archivage éléctronique à moyen et long terme* [en ligne]. 19 octobre 2011. S.l. : s.n. Disponible à l’adresse : https://vsa-aas.ch/wp-content/uploads/2015/06/AEG\_formats\_adaptee\_a\_l\_archivage.pdf.

ARMS, C., FLEISCHHAUER, C. et MURRAY, K., 2013. Sustainability of digital formats planning for Library of Congress collections : external dependencies. In : [en ligne]. 2013. [Consulté le 14 février 2019]. Disponible à l’adresse : https://www.loc.gov/preservation/digital/formats/fdd/dataset\_fdd.shtml.

ARSENEAU, Félix, 2015. Comment classer vos documents et vos dossiers. In : *Trouve qui peut! - S’organiser dans un monde d’information.* [en ligne]. 7 mai 2015. [Consulté le 15 février 2019]. Disponible à l’adresse : https://felixarseneau.wordpress.com/2015/05/07/comment-classer-vos-documents-et-vos-dossiers/.

AUMONT, Gilles, 2017. Datapartage - Séminaire Data Papers - Contexte en enjeux pour l’Inra. In : [en ligne]. 16 juin 2017. [Consulté le 12 février 2019]. Disponible à l’adresse : https://www6.inra.fr/datapartage/Partager-Publier/Publier-un-Data-Paper/Seminaire-Data-Papers.

BANQUE MONDIALE, 2016. Terres agricoles (% du territoire) | Data. In : [en ligne]. 2016. [Consulté le 13 juin 2019]. Disponible à l’adresse : https://donnees.banquemondiale.org/indicateur/AG.LND.AGRI.ZS.

BERTRAND, Isabelle, 2013. *HDR : La qualité des litières végétales : impact sur leurs modalités de décomposition dans les sols et els dynamiques carbone et azote*. Université de Reims Champagne-Ardenne.

BIBLIOTHÈQUE NATIONALE DE FRANCE, 2017. BnF - Principaux standards du web Sémantique : les URI, RDF et SPARQL. In : [en ligne]. 21 août 2017. [Consulté le 21 février 2019]. Disponible à l’adresse : http://www.bnf.fr/fr/professionnels/web\_semantique\_boite\_outils/a.web\_semantique\_standards.html.

BOIZOT, Nathalie et CHARPENTIER, Jean-Paul, 2006. Méthode rapide d’évaluation du contenu en composés phénoliques des organes d’un arbre forestier. In : . 2006. p. 4.

BOLAN, Nanthi S, ADRIANO, Domy C et CURTIN, Denis, 2003. Soil acidification and liming interactions with nutrientand heavy metal transformationand bioavailability. In : *Advances in Agronomy* [en ligne]. S.l. : Elsevier. p. 215‑272. [Consulté le 21 décembre 2018]. ISBN 978-0-12-000796-7. Disponible à l’adresse : http://linkinghub.elsevier.com/retrieve/pii/S0065211302780061.

BUGG, Timothy DH, AHMAD, Mark, HARDIMAN, Elizabeth M et SINGH, Rahul, 2011. The emerging role for bacteria in lignin degradation and bio-product formation. In : *Current Opinion in Biotechnology*. juin 2011. Vol. 22, n° 3, p. 394‑400. DOI 10.1016/j.copbio.2010.10.009.

CARTER, M. R. et STEWART, Bobby A., 1995. *Structure and Organic Matter Storage in Agricultural Soils*. S.l. : CRC Press. ISBN 978-1-56670-033-7.

CGIAR SYSTEM ORGANIZATION, 2018. Welcome to the"4 per 1000" Initiative | 4p1000. In : [en ligne]. 2018. [Consulté le 22 mai 2019]. Disponible à l’adresse : https://www.4p1000.org/.

CHENU, Claire, 2013. Webinaire AFES (2) - 2013 - Claire Chenu -La stabilisation des matières organiques dans les sols. In : *Vimeo* [en ligne]. 7 février 2013. [Consulté le 24 octobre 2018]. Disponible à l’adresse : https://vimeo.com/59411688.

CINES, 2016. Le modèle de référence : l’OAIS. In : [en ligne]. 30 décembre 2016. [Consulté le 20 février 2019]. Disponible à l’adresse : https://www.cines.fr/archivage/un-concept-des-problematiques/le-modele-de-reference-loais/.

CINES, 2017. Les formats de fichier. In : [en ligne]. 4 juillet 2017. [Consulté le 14 février 2019]. Disponible à l’adresse : https://www.cines.fr/archivage/des-expertises/les-formats-de-fichier/.

CIRAD, 2018. 2 - Pourquoi publier un data paper ? / Data paper / Rédiger / Aide à la publication - Coopérer en information scientifique et technique - Cirad. In : [en ligne]. 30 octobre 2018. [Consulté le 25 mars 2019]. Disponible à l’adresse : https://coop-ist.cirad.fr/aide-a-la-publication/rediger/data-paper/2-pourquoi-publier-un-data-paper.

COMMISSION EUROPÉENNE - DIRECTION GÉNÉRALE DE LA RECHERCHE ET DE L’INNOVATION, 2016. *Programme H2020 - Lignes directrices pour la gestion des données FAIR dans Horizon 2020* [en ligne]. 26 juillet 2016. S.l. : s.n. [Consulté le 13 février 2016]. Disponible à l’adresse : http://www.donneesdelarecherche.fr/IMG/pdf/lignes-directrices\_gestion-donnees-fair\_horizon2020\_version\_3.0\_tr-fr.pdf.

COTRUFO, M. Francesca, SOONG, Jennifer L., HORTON, Andrew J., CAMPBELL, Eleanor E., HADDIX, Michelle L., WALL, Diana H. et PARTON, William J., 2015. Formation of soil organic matter via biochemical and physical pathways of litter mass loss. In : *Nature Geoscience*. octobre 2015. Vol. 8, n° 10, p. 776‑779. DOI 10.1038/ngeo2520.

COTRUFO, M. Francesca, WALLENSTEIN, Matthew D., BOOT, Claudia M., DENEF, Karolien et PAUL, Eldor, 2013. The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? In : *Global Change Biology*. avril 2013. Vol. 19, n° 4, p. 988‑995. DOI 10.1111/gcb.12113.

DERRIEN, D, DIGNAC, M F, BASILE-DOELSCH, I, BAROT, Sébastien, CÉCILLON, L, CHENU, C, CHEVALLIER, Tiphaine, FRESCHET, G T, GARNIER, P, GUENET, B, HEDDE, M, KLUMPP, K, LASHERMES, G, MARON, P A, NUNAN, N, ROUMET, C et BARÉ, P, 2016. Stocker du C dans les sols : quels mécanismes, quelles pratiques agricoles, quels indicateurs ? In : *Etude et Gestion des Sols*. 2016. p. 33.

DICTIONNAIRE ENVIRONNMENT ET DÉVELOPPEMENT DURABLE, 2010. Définition biomasse végétale. In : [en ligne]. 2010. [Consulté le 13 juin 2019]. Disponible à l’adresse : https://www.dictionnaire-environnement.com/biomasse\_vegetale\_ID4929.html.

DORANUM, 2017. Métadonnées, standards, formats. In : [en ligne]. 27 novembre 2017. [Consulté le 18 février 2019]. Disponible à l’adresse : https://doranum.fr/metadonnees-standards-formats/.

DORANUM, 2018a. Formation dépôt et entrepôts. In : [en ligne]. 28 mai 2018. [Consulté le 19 février 2019]. Disponible à l’adresse : https://doranum.fr/depot-entrepots/formation/.

DORANUM, 2018b. Identifiants pérennes. In : [en ligne]. 10 juillet 2018. [Consulté le 2 octobre 2017]. Disponible à l’adresse : https://doranum.fr/identifiants-perennes-pid/.

DUNOD, 2008. Comprendre le web sémantique. In : *CommentCaMarche* [en ligne]. 2008. [Consulté le 21 février 2019]. Disponible à l’adresse : https://www.commentcamarche.net/faq/14412-comprendre-le-web-semantique.

ELSEVIER, 2018. *Data-in-brief\_template*. 2018. S.l. : s.n.

EUROPEAN COMMISSION - DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION, 2016. *H2020 Programme Guidelines on FAIR Data Managementin Horizon 2020* [en ligne]. 26 juin 2016. S.l. : s.n. [Consulté le 13 février 2019]. Disponible à l’adresse : http://ec.europa.eu/research/participants/data/ref/h2020/grants\_manual/hi/oa\_pilot/h2020-hi-oa-data-mgt\_en.pdf.

FIERER, Noah, BRADFORD, Mark A. et JACKSON, Robert B., 2007. TOWARD AN ECOLOGICAL CLASSIFICATION OF SOIL BACTERIA. In : *Ecology*. juin 2007. Vol. 88, n° 6, p. 1354‑1364. DOI 10.1890/05-1839.

FILATRE, A. et HENSENS, H., 2017. *Réaliser un plan de gestion de données avec DMP OPIDoR*. 7 juillet 2017. S.l. : s.n.

FILY, M. F., 2015. Connaître et utiliser les licences Creative Commons, en 6 points. In : *CIRAD* [en ligne]. 14 septembre 2015. [Consulté le 15 février 2019]. Disponible à l’adresse : https://coop-ist.cirad.fr/aide-a-la-publication/publier-et-diffuser/connaitre-et-utiliser-les-licences-creative-commons/3-les-licences-creative-commons-cc-generalites.

FRESCHET, Grégoire T., AERTS, Rien et CORNELISSEN, Johannes H. C., 2012. A plant economics spectrum of litter decomposability. In : *Functional Ecology*. 2012. Vol. 26, n° 1, p. 56‑65. DOI 10.1111/j.1365-2435.2011.01913.x.

FREY, S.D., ELLIOTT, E.T., PAUSTIAN, K. et PETERSON, G.A., 2000. Fungal translocation as a mechanism for soil nitrogen inputs to surface residue decomposition in a no-tillage agroecosystem. In : *Soil Biology and Biochemistry*. mai 2000. Vol. 32, n° 5, p. 689‑698. DOI 10.1016/S0038-0717(99)00205-9.

GALL, Maud Le, MONTAGNE, Lucile, JAGUELIN-PEYRAUD, Yolande, PASQUIER, Anne et GAUDRE, Didier, 2012. Prédiction de la teneur en fibres totales et insolubles de matières premières courantes dans l’alimentation du porc à partir de leur composition chimique. In : . 15 novembre 2012. p. 9.

GATOUX, S., 2011. Laboratoire d’Analyse des Sols d’Arras - SOL-0604 : Phosphore méthode de Olsen. In : [en ligne]. 2 avril 2011. [Consulté le 12 juin 2019]. Disponible à l’adresse : https://www6.lille.inra.fr/las/Methodes-d-analyse/Sols/06.-Phoshore-assimilable/SOL-0604-Phosphore-methode-de-Olsen.

GAUTIER POUPEAU, 2010. Vous avez dit « format pérenne » ? | Les petites cases. In : [en ligne]. 5 mai 2010. [Consulté le 14 février 2019]. Disponible à l’adresse : https://www.lespetitescases.net/vous-avez-dit-format-perenne.

GERHARDT, 2015. *Analyse de l’azote - La méthode de Jean Dumas* [en ligne]. mars 2015. S.l. : s.n. Disponible à l’adresse : https://www.gerhardt.de/fileadmin/Redaktion/downloads/Stickstoffanalyse\_-\_Die\_Methode\_von\_Jean\_Dumas\_gekuerzt\_f\_Homepage-fre-FR.pdf.

GIDENNE, Thierry, 2010. Nutrition, alimentation et santé du lapin. In : *Maîtrise sanitaire dans un élevage de lapins* [en ligne]. Toulouse. 1 juin 2010. Disponible à l’adresse : https://fr.slideshare.net/rabahrabah/04-asfc-juin02010gidennenutritionmaitrisesanitaire.

HO, Adrian, LONARDO, D. Paolo Di et BODELIER, Paul L. E., 2017. Revisiting life strategy concepts in environmental microbial ecology. In : *FEMS Microbiology Ecology*. 22 janvier 2017. p. fix006. DOI 10.1093/femsec/fix006.

INES, CNRS, SIAF et TGE-ADONIS, 2011. *Guide méthodologique - Pour le choix de formats numériques pérennes dans un contexte de données orales et visuelles* [en ligne]. 26 avril 2011. S.l. : s.n. Disponible à l’adresse : https://www.huma-num.fr/sites/default/files/guide-formats-numeriques.pdf.

INIST et CNRS, 2018. Tutoriels - Le libre accès aux résultats de la recherche dans le cadre d’Horizon 2020. In : *Inist* [en ligne]. 27 juillet 2018. [Consulté le 15 février 2019]. Disponible à l’adresse : https://www.inist.fr/tutoriels/le-libre-acces-aux-resultats-de-la-recherche-dans-le-cadre-dhorizon-2020/.

INRA, 2016a. Charte des infrastructures de recherche de l’Inra. In : [en ligne]. 20 octobre 2016. [Consulté le 25 mars 2019]. Disponible à l’adresse : http://institut.inra.fr%2FReperes%2FDocuments%2FChartes%2FCharte-des-infrastructures-de-recherche-de-l-Inra.

INRA, 2016b. L’Inra renforce sa politique de libre accès à ses résultats scientifiques. In : [en ligne]. 17 octobre 2016. [Consulté le 25 mars 2019]. Disponible à l’adresse : http://presse.inra.fr%2FCommuniques-de-presse%2FInra-charte-libre-acces.

KHEDER, Fadi, 2007. *Production et purification d’acide férulique estérases. Application à la synthèse d’esters phénoliques*. Nancy : Institut National Polytechnique de Lorraine.

KROON, Paul A., WILLIAMSON, Gary, FISH, Neville M., ARCHER, David B. et BELSHAW, Nigel J., 2000. A modular esterase from Penicillium funiculosum which releases ferulic acid from plant cell walls and binds crystalline cellulose contains a carbohydrate binding module. In : *European Journal of Biochemistry*. 1 décembre 2000. Vol. 267, n° 23, p. 6740‑6752. DOI 10.1046/j.1432-1033.2000.01742.x.

LAPERDRIX, Marie, 2016. *Plan de classement et nommage des fichiers numériques* [en ligne]. 8 décembre 2016. S.l. : s.n. [Consulté le 15 février 2019]. Disponible à l’adresse : http://sieil37.fr/images/Actualites/Sieil/Demat-2016-12-08/cr\_atelier\_nommage\_20170104\_VD-web.pdf.

LAPIERRE, Catherine, 2013. L’acide férulique, acteur discret mais universel de la construction des parois lignifiées. In : [en ligne]. 11 juin 2013. [Consulté le 24 octobre 2018]. Disponible à l’adresse : http://www.cepia.inra.fr%2FLe-departement-Les-recherches%2FNos-resultats%2Facide-ferulique%2F%28key%29%2F26.

LASHERMES, G, HOUOT, S, NICOLARDOT, B, PARNAUDEAU, V, MARY, B, MORVAN, T, CHAUSSOD, R, METZGER, L, THURIÈS, L, VILLETTE, C, TRICAUD, A et GUILLOTIN, M L, 2007. Apport de matières organiques exogènes en agriculture : indicateur de potentialité de stockage de carbone dans les sols et définition de classes de disponibilité d’azote. In : . 2007. p. 6.

LASHERMES, Gwenaëlle, 2017a. La matière organique des sols et les cycles du C, N, P, S. In : . Cours M1. Reims. 10 octobre 2017.

LASHERMES, Gwenaëlle, 2017b. Le processus de décomposition des matières organiques et le couplage des cycles du C et N. In : . Cours M1. Reims. 12 octobre 2017.

MACHINET, G. E., BERTRAND, I., CHABBERT, B. et RECOUS, S., 2009. Decomposition in soil and chemical changes of maize roots with genetic variations affecting cell wall quality. In : *European Journal of Soil Science*. 13 mars 2009. Vol. 60, n° 2, p. 176‑185. DOI 10.1111/j.1365-2389.2008.01109.x.

MACHINET, Gaylord Erwan, 2009. Utilisation de la variabilité génétique du maïs pour évaluer le rôle de la qualité chimique des racines sur le processus de décomposition dans les sols. In : . 2009. p. 207.

MARIE CLAUDE QUIDOZ, 2018. Les principes FAIR appliqués aux sauvegardes sur le long terme. In : *séminaire interopérabilité et pérennisation des données de la recherche* [en ligne]. Paris. 27 novembre 2018. Disponible à l’adresse : https://gt-donnees2018.sciencesconf.org/data/06\_Atelier\_Donnees\_Sauvegarde\_Quidoz.pdf.

MARIE SAUVADET, 2016. *Impacts de la qualité et quantité des résidus de culture sur la structure et le fonctionnement des communautés du sol. Application aux systèmes de grandes cultures du Nord-Est de la France.* Reims : Univesité de Reims Champagne-Ardenne.

MARSCHNER, Petra, SOLAIMAN, Zakaria et RENGEL, Zed, 2005. Growth, phosphorus uptake, and rhizosphere microbial-community composition of a phosphorus-efficient wheat cultivar in soils differing in pH. In : *Journal of Plant Nutrition and Soil Science*. 2005. Vol. 168, n° 3, p. 343‑351. DOI 10.1002/jpln.200424101.

MASTIHUBA, Vladimı́r, KREMNICKÝ, Lubomı́r, MASTIHUBOVÁ, Mária, WILLETT, J. L et CÔTÉ, Gregory L, 2002. A spectrophotometric assay for feruloyl esterases. In : *Analytical Biochemistry*. 1 octobre 2002. Vol. 309, n° 1, p. 96‑101. DOI 10.1016/S0003-2697(02)00241-5.

M.-C JACQUEMOT-PERBAL et CIOLEK-FIGIEL, A., 2016. Gestion et diffusion des données de la recherche. In : *Formation URFIST* [en ligne]. Lyon. 22 mars 2016. [Consulté le 13 février 2019]. Disponible à l’adresse : https://urfist.univ-lyon1.fr/files/2016/06/Urfist\_Lyon\_Donnees\_de\_la\_recherche\_gestion\_diffusion.pdf.

MINISTÈRE DE L’ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L’INNOVATION, 2014. Le libre accès aux publications et aux données de recherche. In : *Horizon 2020* [en ligne]. 2014. [Consulté le 25 mars 2019]. Disponible à l’adresse : http://www.horizon2020.gouv.fr/cid82025/le-libre-acces-aux-publications-aux-donnees-recherche.html.

MINISTÈRE DE L’ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L’INNOVATION, 2018. *Plan natoinal pour la science ouverte* [en ligne]. 4 juillet 2018. S.l. : s.n. [Consulté le 25 mars 2019]. Disponible à l’adresse : https://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/67/2/PLAN\_NATIONAL\_SCIENCE\_OUVERTE\_978672.pdf.

MÖLLER, Jürgen, 2014. *Comparaison des méthodes de dosage des fibres sur produits agricoles* [en ligne]. 1 février 2014. S.l. : s.n. Disponible à l’adresse : https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwidq87c9uPiAhW2BWMBHS7cC-AQFjAAegQIABAC&url=https%3A%2F%2Fwww.fossanalytics.com%2F-%2Fmedia%2Ffiles%2Fdocuments%2Fpapers%2Ffeed-and-forage-segment%2Ffibre-methods-compared\_gb.pdf&usg=AOvVaw3Aj3ffAZh2BE-UNmP-pT9N.

MOORHEAD, Daryl L. et SINSABAUGH, Robert L., 2006. A THEORETICAL MODEL OF LITTER DECAY AND MICROBIAL INTERACTION. In : *Ecological Monographs*. mai 2006. Vol. 76, n° 2, p. 151‑174. DOI 10.1890/0012-9615(2006)076[0151:ATMOLD]2.0.CO;2.

MOORHEAD, Daryl, LASHERMES, Gwenaëlle, RECOUS, Sylvie et BERTRAND, Isabelle, 2014. Interacting Microbe and Litter Quality Controls on Litter Decomposition: A Modeling Analysis. In : SINGER, Andrew C. (éd.), *PLoS ONE*. 29 septembre 2014. Vol. 9, n° 9, p. e108769. DOI 10.1371/journal.pone.0108769.

NATIONAL SCIENCE FONDATION et UNIVERSITY OF TOLEDO, 2013. I•MOLD -Leaf Anatomy. In : [en ligne]. 2013. [Consulté le 24 octobre 2018]. Disponible à l’adresse : http://imold.utoledo.edu/anatomy.html.

NUNAN, Naoise, LELOUP, Julie, RUAMPS, Léo S., POUTEAU, Valérie et CHENU, Claire, 2017. Effects of habitat constraints on soil microbial community function. In : *Scientific Reports* [en ligne]. décembre 2017. Vol. 7, n° 1. [Consulté le 19 décembre 2018]. DOI 10.1038/s41598-017-04485-z. Disponible à l’adresse : http://www.nature.com/articles/s41598-017-04485-z.

OYEYEMI, K. D., SANUADE, O. A., OLADUNJOYE, M. A., AIZEBEOKHAI, A. P., OLAOJO, A. A., FATOBA, J. O., OLOFINNADE, O. M., AYARA, W. A. et OLADAPO, O., 2018. Data on the thermal properties of soil and its moisture content. In : *Data in Brief*. 1 avril 2018. Vol. 17, p. 900‑906. DOI 10.1016/j.dib.2018.02.018.

PDF TOOLS AG, 2016. 10 choses les plus importantes sur PDF/A. In : *PDF-Tools AG* [en ligne]. 3 février 2016. [Consulté le 14 février 2019]. Disponible à l’adresse : https://www.pdf-tools.com/pdf20/fr/savoir-faire/normes-iso-pdf/10-choses-les-plus-importantes-sur-pdfa/.

PERRIN, jf, 2011. DOSAGE DE L’AZOTE TOTAL PAR METHODE DE KJELDAHL. In : . 2011. p. 4.

PÔLE DONNÉES DE LA RECHERCHE IST, 2015. Datapartage - Enjeux et critères de choix. In : [en ligne]. 3 novembre 2015. [Consulté le 14 février 2019]. Disponible à l’adresse : https://www6.inra.fr/datapartage/Gerer/Choisir-des-formats-de-fichier/Enjeux-et-criteres-de-choix.

PÔLE DONNÉES DE LA RECHERCHE IST, 2016. Datapartage - Pourquoi. In : [en ligne]. 7 février 2016. [Consulté le 15 février 2019]. Disponible à l’adresse : https://www6.inra.fr/datapartage/Gerer/Nommer-et-organiser-ses-fichiers-de-donnees/Pourquoi.

PÔLE DONNÉES DE LA RECHERCHE IST, 2018a. Datapartage - Archiver de manière pérenne. In : *INRA* [en ligne]. 10 janvier 2018. [Consulté le 20 février 2019]. Disponible à l’adresse : https://www6.inra.fr/datapartage/Gerer/Archiver-de-maniere-perenne/Archivage-perenne.

PÔLE DONNÉES DE LA RECHERCHE IST, 2018b. Datapartage - Choisir une licence. In : *INRA* [en ligne]. 12 juin 2018. [Consulté le 15 février 2019]. Disponible à l’adresse : https://www6.inra.fr/datapartage/Partager-Publier/Choisir-une-licence.

PÔLE DONNÉES DE LA RECHERCHE IST, 2018c. Datapartage - Comment. In : [en ligne]. 26 mars 2018. [Consulté le 15 février 2019]. Disponible à l’adresse : https://www6.inra.fr/datapartage/Gerer/Nommer-et-organiser-ses-fichiers-de-donnees/Comment.

PÔLE DONNÉES DE LA RECHERCHE IST, 2018d. Datapartage - Le web sémantique. In : [en ligne]. 9 août 2018. [Consulté le 21 février 2019]. Disponible à l’adresse : https://www6.inra.fr/datapartage/Produire-des-donnees-FAIR/Comment-FAIR-en-pratique/Le-web-semantique.

PÔLE DONNÉES DE LA RECHERCHE IST, 2018e. Datapartage - Les enjeux pour l’Inra. In : [en ligne]. 27 février 2018. [Consulté le 12 février 2019]. Disponible à l’adresse : https://www6.inra.fr/datapartage/Partager-Publier/Les-enjeux-pour-l-Inra.

PÔLE DONNÉES DE LA RECHERCHE IST, 2018f. Datapartage - Publier un Data Paper. In : *INRA* [en ligne]. 13 août 2018. [Consulté le 20 février 2019]. Disponible à l’adresse : https://www6.inra.fr/datapartage/Partager-Publier/Publier-un-Data-Paper.

PÔLE FORMATION IST, 2018. Publier un Data Paper pour valoriser ses données. In : *Formation Data Paper*. S.l. 26 octobre 2018.

PUREN, Marie, 2017. Les métadonnées dans un DMP. In : . 6 juillet 2017. p. 73.

RECOUS, Sylvie, 1995. Soil inorganic N availability : effect on maize residue decomposition. In : . 1995.

RECOUS, Sylvie, LASHERMES, Gwenaëlle et BERTRAND, Isabelle, 2017. Couplages et contrôles des cycles du carbone et de l’azote par les communautés microbiennes dans les sols cultivés. In : . 2017. p. 28.

RECOUS, Sylvie et LE ROUX, Xavier, 2008. *Modifications des pratiques agricoles et impacts environnementaux : vers un meilleur couplage de la dynamique des Communautés microbiennes du Sol, des Matières Organiques du Sol et des Flux de carbone et d’azote dasn les sols (COSMOS-FLUX)* [en ligne]. S.l. Disponible à l’adresse : http://www.gessol.fr/sites/default/files/GESSOL2\_Rapport\_Final\_Cosmosflux.pdf.

ROBERGE, Alexandre, 2015. Les données scientifiques en voie d’extinction. In : *Thot Cursus* [en ligne]. 8 février 2015. [Consulté le 12 février 2019]. Disponible à l’adresse : /articles/33591.

ROUCHON, Olivier, 2010. Qualité et archivage pérenne. In : *Journée d’étude BnF/AFNOR CG46* [en ligne]. S.l. 7 juin 2010. Disponible à l’adresse : http://www.bnf.fr/documents/archivage\_cines.pdf.

RUBIN, Edward M., 2008. Genomics of cellulosic biofuels. In : *Nature*. août 2008. Vol. 454, n° 7206, p. 841‑845. DOI 10.1038/nature07190.

SAGGAR, Surinder, STEWART, J.W.B et BETTANY, J.R., 1981. Measurement of microbial sulfur in soil. In : . décembre 1981.

TABATABAI, M. A. et BREMNER, J. M., 1970. An Alkaline Oxidation Method for Determination of Total Sulfur in Soils. In : . janvier 1970. p. 4.

TARDY, Florence, 2015. Caractérisation fonctionnelle des stratégies de compétition pour les ressources aériennes et souterraines au sein d’une gamme d’espèces végétales: application à la régulation biologique des adventices par les plantes de services dans l’agrosystème bananier antillais. In : . 14 décembre 2015. p. 181.

THURIÈS, Laurent, 2009. Techniques de caractérisation des matières organiques exogènes\_Dynamique de transformation des apports organiques\_origine tropicale et cas des fertilisants organiques. In : *Formation Matière Organique*. Montpellier. 2009.

TRINSOUTROT, Isabelle, 1999. *Influence de la qualité biochimique et de la teneur en azote de résidus de colza sur les biotransformations du carbone et de l’azote au cours de leur décomposition dans le sol*. Lyon : Université Claude Bernard-Lyon I.

UNIL, 2016. Règles de nommage des fichiers et dossiers. In : [en ligne]. 11 mai 2016. [Consulté le 15 février 2019]. Disponible à l’adresse : https://uniris.unil.ch/researchdata/sujet/organiser/regles-nommage/.

VAN DER WAL, Annemieke, DE BOER, Wietse, SMANT, Wiecher et VAN VEEN, Johannes A., 2007. Initial decay of woody fragments in soil is influenced by size, vertical position, nitrogen availability and soil origin. In : *Plant and Soil*. décembre 2007. Vol. 301, n° 1‑2, p. 189‑201. DOI 10.1007/s11104-007-9437-8.

VAN DER WAL, Annemieke, GEYDAN, Thomas D., KUYPER, Thomas W. et DE BOER, Wietse, 2013. A thready affair: linking fungal diversity and community dynamics to terrestrial decomposition processes. In : *FEMS Microbiology Reviews*. juillet 2013. Vol. 37, n° 4, p. 477‑494. DOI 10.1111/1574-6976.12001.

VIGIL, M.F. et SPARK, D., 2004. Factors affecting the rate of crop residue decomposition under field conditions. Conservation tillage Fact Sheet 33-95. USDA-ARS and USDA-NRCS, Akron, Colorado. In : . 2004.

WALIGORA, Cécile, 2010. Racines et sol : un monde de communications et d’équilibres - A2C le site de l’agriculture de conservation. In : [en ligne]. 2010. [Consulté le 26 octobre 2018]. Disponible à l’adresse : http://agriculture-de-conservation.com/Racines-et-sol-un-monde-de.html.

WODDISCOVERY, 2010. Data and the Web – a great many of choices. In : *Web of Data* [en ligne]. 1 mars 2010. [Consulté le 13 février 2019]. Disponible à l’adresse : https://webofdata.wordpress.com/2010/03/01/data-and-the-web-choices/.