

UE DATA-CEVI

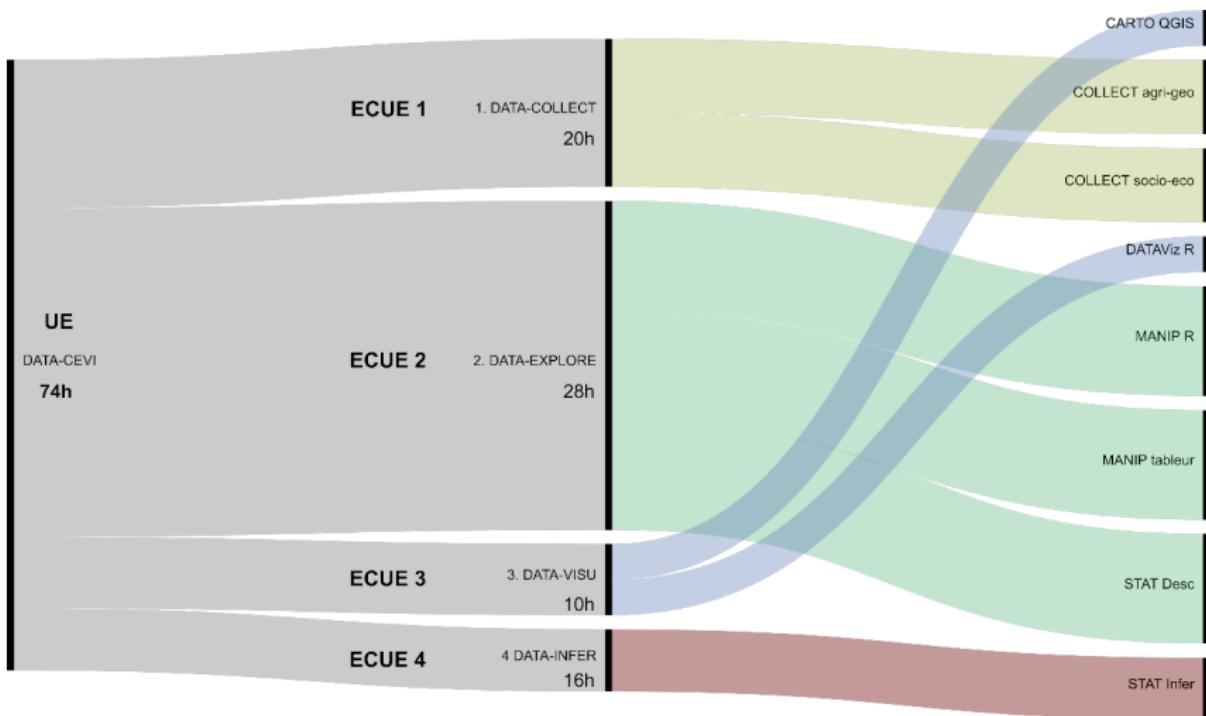
– Données agri-environnementales –
collecter, explorer, visualiser, inférer

david.sheeren@ensat.fr

Organisation de l'UE

Composition

Une UE Ressource



Composition

ECUE 1 : DATA-COLLECTE

- **Resp.** : J. Brailly
- **Intervenants** : J. Brailly, M.L. de Capella, C. Gendre, D. Sheeren
- **Volume horaire** : 6 CM (1h20) + 5 TD (2h ou 3h) + 10h autonomie
- **Objectifs d'apprentissage** : connaître différentes sources de données agri-environnementales, identifier leur nature et comprendre leurs caractéristiques, appliquer différentes techniques de recherche, collecte et extraction de données dans des formats variés (BD, webscraping, import de flux...), intégrer des données spatiales dans un SIG, appliquer des méthodes de collecte de données en SHS (enquêtes).
- **Outils associés** : tableur (xls), R, QGIS

Composition

ECUE 2 : DATA-EXPLORE

- **Resp.** : C. Laplanche
- **Intervenants** : C. Laplanche, E. Maza, J. Pirrello, M.-F. Jardinaud, intervenant extérieur
- **Volume horaire** : 2 CM (1h20) + 5 TD (2h) + 10h autonomie
- **Objectifs d'apprentissage** : manipuler des données dans un tableau et sous R (import, extraction, applications de fonctions), décrire et résumer des jeux de données en termes statistiques et à l'aide de graphiques simples adaptés
- **Outils associés** : tableau (xls), R

Composition

ECUE 3 : DATA-VISU

- **Resp.** : D. Sheeren
- **Intervenants** : C. Laplanche, Y. Holtz, C. Gendre, D. Sheeren
- **Volume horaire** : 2 CM (1h20) + 1 TD (2h) + 4h autonomie
- **Objectifs d'apprentissage** : identifier et décrire différentes représentations graphiques possibles selon la nature des données et le message à communiquer (comparer, mettre en relation, représenter une distribution...), appliquer une méthode existante de représentation graphique de données uni ou bivariées, composer une carte thématique en respectant les règles de sémiologie graphique et d'habillage.
- **Outils associés** : QGIS, R

Composition

ECUE 4 : DATA-INFER

- **Resp.** : E. Maza
- **Intervenants** : M.-F. Jardinaud, E. Maza, J. Pirrello
- **Volume horaire** : 3 CM (1h20) + 6 TD (2h)
- **Objectifs d'apprentissage** : définir et différencier les concepts liés à la théorie de l'échantillonnage, à l'estimation statistique et aux tests statistiques ; calculer les estimations ponctuelles et par intervalle de confiance de paramètres d'une population, réaliser des tests statistiques de conformité
- **Outils associés** : R

Relations

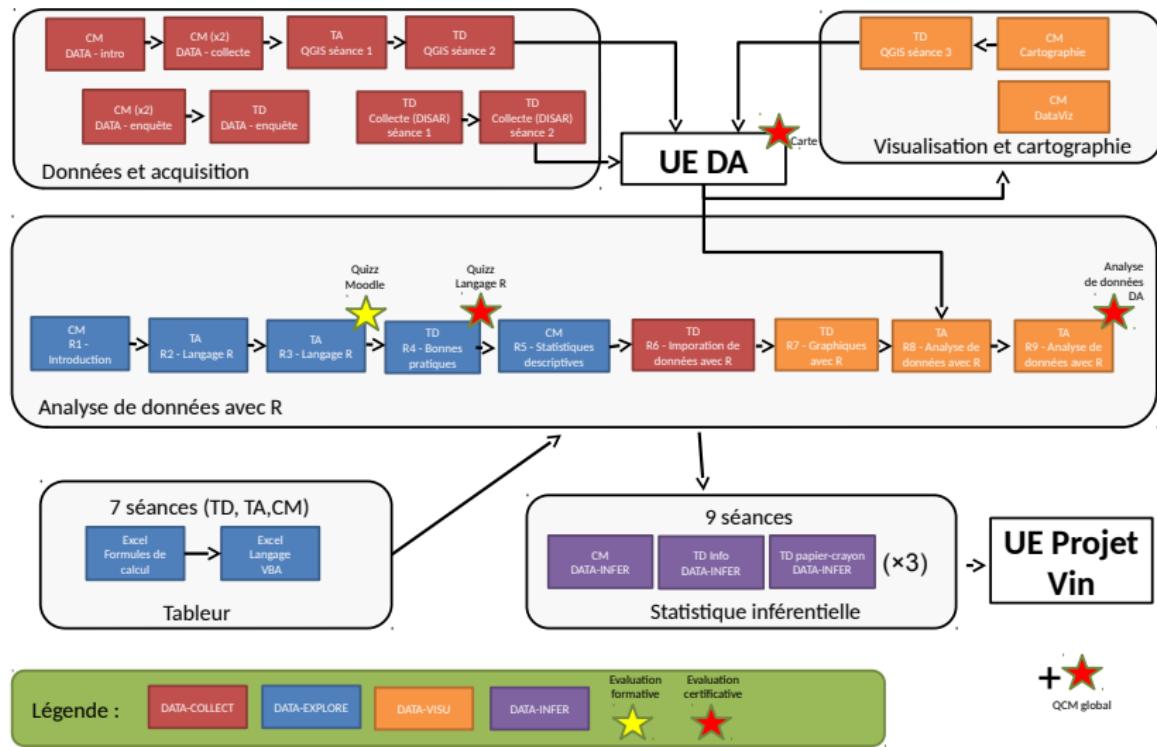
Liens avec les UE projets du S5 et S6

- Projet **Diagnostiquer, conseiller un agro-écosystème** (DA & CA) :
 - ▶ recueil de données issues des BD institutionnelles (AGRESTE, INSEE)
 - ▶ collecte de données climatiques, pédologiques...
 - ▶ représenter l'exploitation agricole dans son contexte (PRA, enviro)
- Projet **Production de vin et validation par analyses** (VIN) :
 - ▶ utilisation des outils de statistiques descriptives et tests
 - ▶ représentations graphiques
- Projet tutoré **partenarial** (PPART) :
 - ▶ collecte et représentation de données diverses

Composition

Une UE Ressource

DATA-CEVI (Année 2022-2023)



Evaluation

- Evaluation individuelle **notée par les enseignants** :
 - ▶ questionnaire à choix multiple (**QCM**) intégrant des questions relatives aux 4 ECUE (coef. 0.6)
 - ▶ contrôle continu en DATA-EXPLORE (coef. 0.05)
- Evaluations individuelles **notées par les pairs et supervisées par les enseignants** :
 - ▶ réalisation de 2 **analyses de données + graphiques simples sous R** en lien avec l'UE DA (coef. 0.2)
 - ▶ réalisation de 1 **carte sous QGIS** du parcellaire agricole et de l'assoulement en lien avec l'UE DA (coef. 0.15)

Module 1 : DATA-COLLECTE

1. Les données

1. Les données

Qu'est-ce qu'une donnée ?

Définition (Larousse)

- "Ce qui est connu ou admis comme tel, sur lequel on peut fonder un raisonnement, qui sert de point de départ pour une recherche" (ex. théorie en biologie)
- "Quantité connue dans l'énoncé d'un problème et qui sert à trouver la solution" (ex. énoncé mathématique)
- "Résultats d'observations ou d'expériences faites délibérément ou à l'occasion d'autres tâches et soumis aux méthodes statistiques" (ex. relevés écologiques)
- "Ensemble des indications enregistrées en machine pour permettre l'analyse ou la recherche automatique des informations" (ex. base de données)

1. Les données

Qu'est-ce qu'une donnée ?

Donnée \neq information

- Une ***donnée*** est le **reflet symbolique de nombres, grandeurs ou faits** (Le Duigou 2010). C'est le résultat d'un processus de perception de la réalité.
 - ▶ Ex. 25 (lecture du nombre sur un thermomètre)
- Une ***information*** résulte d'une **action d'interprétation d'une donnée**. C'est une donnée placée dans un contexte (Le Duigou 2010). Elle est issue d'un traitement de la donnée faisant appel à un savoir.
 - ▶ Ex. 25 °C (lien entre nombre et son unité; on donne du sens à 25)

1. Les données

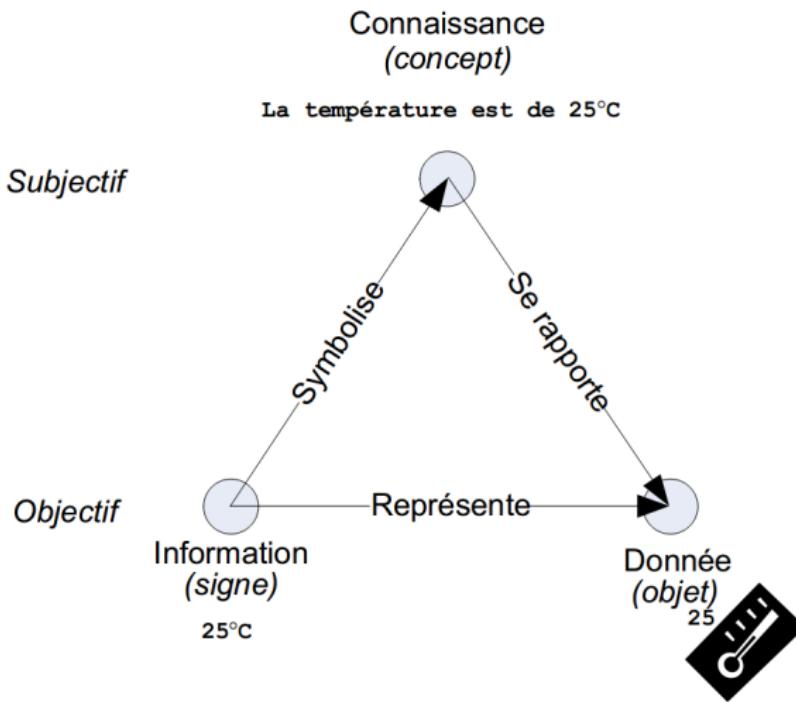
Qu'est-ce qu'une donnée ?

Information \neq connaissance

- Une *connaissance* est une **information qui prend un sens** dans un contexte particulier (Le Duigou 2010). C'est la signification que l'on donne à une information.
 - ▶ Ex. Température de 25 °C (information rattachée à son concept)

1. Les données

Qu'est-ce qu'une donnée ?



(Source : M. Héon 2010)

1. Les données

Typologies de données

① Données **qualitatives** (non mesurables, lien d'appartenance)

① **Nominales** : qui distingue sans hiérarchie (modalité)

★ Calcul d'effectifs ou de fréquences relatives

② **Ordinalés** : qui distingue de façon ordonnée (rang)

★ Calcul d'effectifs, de fréquences relatives, de rang moyen

Exemples :

- ▶ nom de famille, niveau d'étude, genre, lieu de naissance, classe d'âge, statut socio-économique...

1. Les données

Typologies de données

① Données **qualitatives** (non mesurables, lien d'appartenance)

① **Nominales** : qui distingue sans hiérarchie (modalité)

★ Calcul d'effectifs ou de fréquences relatives

② **Ordinales** : qui distingue de façon ordonnée (rang)

★ Calcul d'effectifs, de fréquences relatives, de rang moyen

Elles peuvent être :

- ▶ **Exhaustives** : autant de modalités que d'éléments
- ▶ **Discrètes** : moins de modalités que d'éléments (→ catégories)

1. Les données

Typologies de données

② Données **quantitatives** (cardinales)

① **Absolues** ou *échelle de rapport*

- ★ le zéro de l'échelle est naturel et non arbitraire ($0 =$ absence du phénomène considéré)
- ★ il peut s'agir d'un *stock* (quantité à un instant t) ou d'un *flux* (quantité sur une période de temps)
- ★ la comparaison d'intervalles (écart) et de rapports est possible

Exemple : distance, temps, température en Kelvin, nombre d'individus . . .

- ★ le nombre d'habitants d'une ville
- ★ la distance qui sépare 2 lieux

1. Les données

Typologies de données

② Données **quantitatives** (cardinales)

② Relatives ou échelle d'intervalle

- ★ le zéro de l'échelle est arbitraire (repère conventionnel); il correspond à une valeur (caractère exprimé)
- ★ seule la comparaison d'intervalles (écart) est possible, pas la calcul de rapports

Exemples : température en °Celsius, QI...

- ★ Si $T^\circ = 20^\circ\text{C}$ à Paris et $T^\circ = 40^\circ\text{C}$ à Toulouse, il ne fait pas 2x plus chaud à Toulouse... ($^\circ\text{K} = ^\circ\text{C} + 273,15$)
 $20^\circ\text{C} = \sim 293^\circ\text{K}$ et $40^\circ\text{C} = \sim 313^\circ\text{K} \rightarrow 313^\circ\text{K} \neq (2 \times 293^\circ\text{K})$
- ★ Si QI de A = 50 et QI de B = 100, le QI de B > QI de A mais il n'est pas 2x plus intelligent

1. Les données

Typologies de données

③ Données **semi-quantitatives**

- ▶ Données dans une unité de mesure **non additive**
 - ★ Cas des %, taux, densité, ratio, rang...
 - ★ Application de moyennes pondérées

Exemple :

- ★ Si pays A mesure 2 km² avec 6 habitants, densité = 3 hab/km²
- ★ Si pays B mesure 1 km² avec 8 habitants, densité = 8 hab/km²
- ★ La densité de AB (4,6 hab/km²; moyenne pondérée) ≠ densité de A+B (11 hab/km²)

1. Les données

Typologies de données

④ Données **textuelles**

- ▶ Données sous forme de mots et phrases non codées arbitrairement (absence de classe)
 - ★ On y applique des traitements lexicaux spécifiques

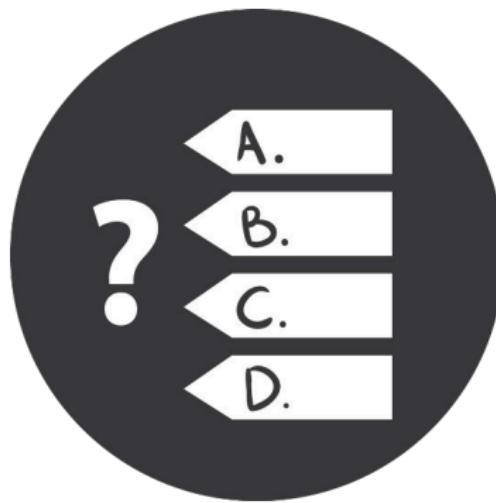
Exemple :

- ★ "Les étudiants de l'ENSAT écoutent avec attention leur enseignant"

1. Les données

Typologies de données

Quizz !



1. Les données

Typologies de données

Q1. Parmi les variables suivantes, laquelle est une variable *qualitative nominale* ?

- ① numéro d'étudiant
- ② densité de population
- ③ niveau de richesse

1. Les données

Typologies de données

Q1. Parmi les variables suivantes, laquelle est une variable *qualitative nominale* ?

- ① [X] numéro d'étudiant
- ② densité de population =*semi-quantitative* (ratio)
- ③ niveau de richesse =*qualitative ordinaire* (catégorie)

1. Les données

Typologies de données

Q2. Parmi les variables suivantes, laquelle est une variable *quantitative relative (échelle d'intervalle)* ?

- ① âge
- ② date de naissance
- ③ durée

1. Les données

Typologies de données

Q2. Parmi les variables suivantes, laquelle est une variable *quantitative relative* (*échelle d'intervalle*) ?

- ① âge =*quantitative absolue* (*échelle de rapport*)
- ② [X] date de naissance *an 0 arbitraire* (\neq *du début de l'univers*)
- ③ durée =*quantitative absolue* (*échelle de rapport*)

1. Les données

Typologies de données

Q3. De quelle nature est la variable pH ?

- ① quantitative
- ② semi-quantitative
- ③ qualitative ordonnée

1. Les données

Typologies de données

Q3. De quelle nature est la variable pH ?

- ① quantitative
- ② [X] semi-quantitative *le pH n'est pas additif !*
- ③ qualitative ordonnée

1. Les données

Typologies de données

Q4. De quelle nature est la variable de concentration en g/l ?

- 1 quantitative
- 2 semi-quantitative
- 3 textuelle

1. Les données

Typologies de données

Q4. De quelle nature est la variable de concentration en g/l ?

- ① quantitative
- ② [X] semi-quantitative *la concentration n'est pas additive !*
- ③ textuelle

1. Les données

Typologies de données

Q5. Peut-on passer d'une variable *quantitative absolue* à une variable *qualitative ordinaire* ?

- 1 oui
- 2 non

1. Les données

Typologies de données

Q5. Peut-on passer d'une variable *quantitative absolue* à une variable *qualitative ordinaire* ?

- ① [X] oui *en appliquant une méthode de discrétisation*
- ② non

1. Les données

Typologies de données

Q6. De quelle nature est la variable *genre* avec 1 pour *homme*, 2 pour *femme*, 3 pour *autre* ?

- ① quantitative absolue
- ② qualitative ordinaire
- ③ qualitative nominale

1. Les données

Typologies de données

Q6. De quelle nature est la variable *genre* avec 1 pour *homme*, 2 pour *femme*, 3 pour *autre* ?

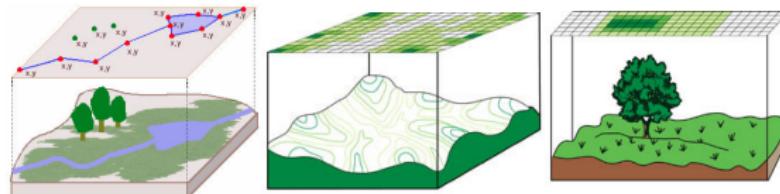
- ① quantitative absolue
- ② qualitative ordinaire
- ③ [X] qualitative nominale *0 peut être réservé à absence de réponse*

1. Les données

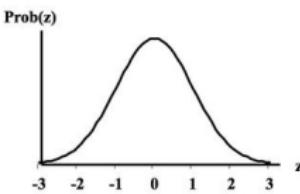
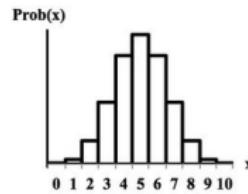
Typologies de données : autres distinctions

- Données **discrètes** vs **continues**

- ▶ *Discrète* (dénombrées) : ne peut prendre que certaines valeurs d'un intervalle de nombres (ex. entiers)
- ▶ *Continue* (mesurées, approximées) : peut prendre toutes les valeurs possibles d'un intervalle de nombres réels



(Source : ESRI)



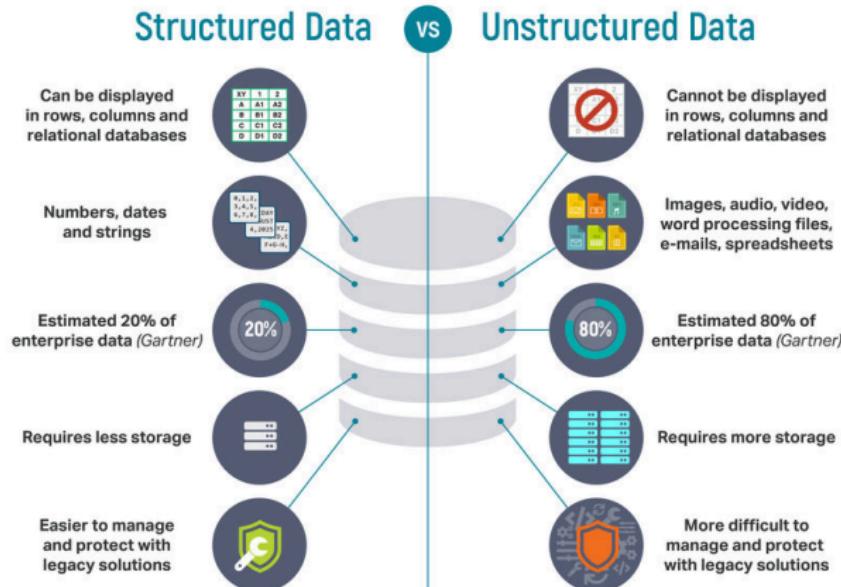
1. Les données

Typologies de données : autres distinctions

- Données **structurées** vs **non structurées** vs **semi-structurées**
 - ▶ *Structurée* : donnée organisée selon un modèle défini de façon à être traitée automatiquement (structure + contenu)
Ex. : données dans BD ou un tableur, lignes de code
 - ▶ *Semi-structurée* : donnée dont seule une partie peut être exploité par une machine (structure reconnue)
Ex. : fichier dans format avec balises (html, tex, md), fichier dans format avec metadonnées (MP3...)
 - ▶ *Non structurée* : le reste ! (absence de structure spécifique; requiert un développement ad-hoc)
Ex. : fichier .txt, .pdf, .doc, .png (sans mots-clés, balises, non indexé...)

1. Les données

Typologies de données : autres distinctions



1. Les données

Typologies de données : autres distinctions

Données semi-structurées : du HTML au XML

Une bibliographie en HTML

```
<H2>Bibliographie XML</H2>
<UL>
  <LI> Jean-Christophe Bernadac et François Knab, <I>Construire une application XML</I>, Paris, Eyrolles, 1999</LI>
  <LI> Alain Michard, <I>XML, Langage et Applications</I>, Paris, Eyrolles, 1998 </LI>
  <LI> William J. Pardi, <I>XML en Action</I>, Paris, Microsoft Press, 1999, adapté de l'anglais par James Guerin</LI>
</UL>
```

Son équivalent (partiel) en XML

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<BIBLIO SUBJECT="XML">

  <BOOK ISBN="9782212090819" LANG="fr" SUBJECT="applications">
    <AUTHOR>
      <FIRSTNAME>Jean-Christophe</FIRSTNAME>
      <LASTNAME>Bernadac</LASTNAME>
    </AUTHOR>
    <AUTHOR>
      <FIRSTNAME>François</FIRSTNAME>
      <LASTNAME>Knab</LASTNAME>
    </AUTHOR>
    <TITLE>Construire une application XML</TITLE>
    <PUBLISHER>
      <NAME>Eyrolles</NAME>
      <PLACE>Paris</PLACE>
    </PUBLISHER>
    <DATEPUB>1999</DATEPUB>
  </BOOK>
```

Des balises génériques de mise en forme
(<H2>, , <L1>...)

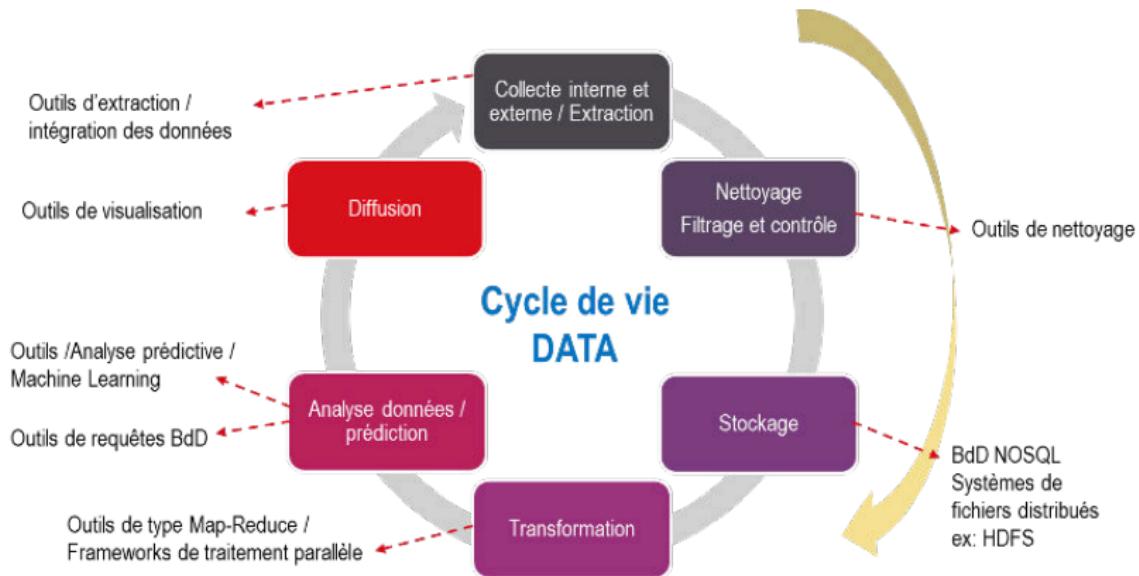
Des balises « sémantiques »
utilisateur (<BOOK>,
<AUTHOR>...)

Indépendance du contenu et
de la forme

(Source : <http://xml.chez.com/initiation/>)

1. Les données

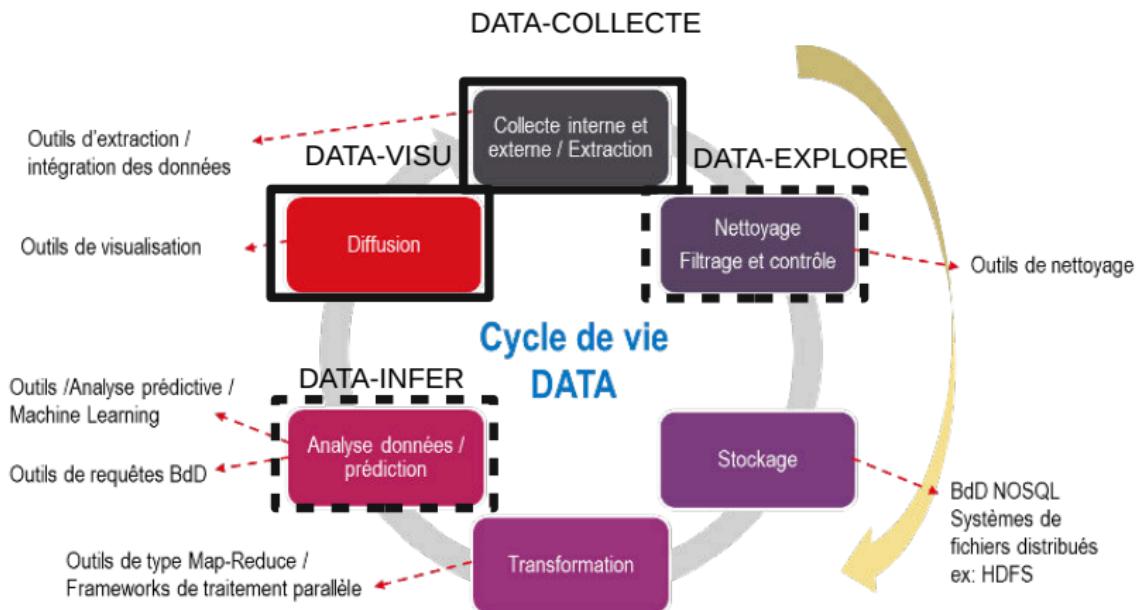
Cycle de vie



(Source : IT-expert Magazine)

1. Les données

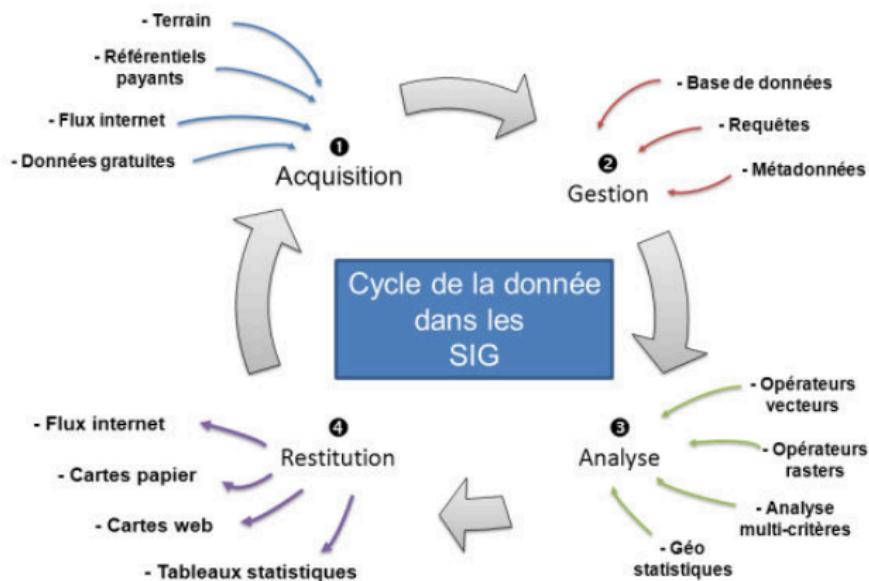
Cycle de vie



(Source : IT-expert Magazine)

1. Les données

Cycle de vie



(Source : Montpellier SupAgro)

2. Le monde des données aujourd'hui

2. Le monde des données aujourd'hui

Ouverture (*open data*)

Donnée ouverte (d'après l'Open Knowledge Foundation)

- Donnée numérique publiée sous **licence ouverte** qui garantit un **accès libre** (de préférence sur internet) et autorise une **réutilisation** et un **partage** sans restrictions techniques, juridiques ou financières tels que copyright, brevet, mécanisme de contrôle...
- Provenance de **n'importe quelle source** (entreprises, associations, fondations, ONG...) mais beaucoup sont issues du secteur public.

2. Le monde des données aujourd'hui

Ouverture (*open data*)

10 critères à respecter (d'après Sunlight Foundation)

- Etre **primaire** (brute, à la source, sans transformation)
- **Complète** (diffusion de l'intégralité de ce qui est collecté, avec métadonnées)
- **Actualisée / fraîche** (maj aussi vite que possible ; maintien de leur valeur)
- **Accessible** (de façon physique) ou électronique
- **Exploitable** (i.e. électroniquement lisible par une machine; web sémantique)
- **Non discriminatoire** (i.e. pour tous)
- **Non propriétaire** (format ouvert si possible standardisé)
- **Libre de droit** (sous licence libre - droits d'auteur libérés)
- **Pérenne** (avec archivage des versions antérieures)
- **Sans coût d'utilisation** (coût de mise à disposition limité, pas forcément gratuit)

2. Le monde des données aujourd'hui

Ouverture (*open data*)

Ouverture des **données publiques** (= politique publique) :

- Mouvement initié aux USA en 2007 lors de la campagne de B. Obama
 - ▶ Inspiré de la démarche *open source* (partage de codes)
 - ▶ Les données publiques = un bien commun d'intérêt général
 - ▶ 2009 : *Open Government Initiative* (premier projet après élection)
 - ▶ 3 principes : transparence, participation, collaboration
 - ▶ Mise en ligne du portail <https://www.data.gov>

2. Le monde des données aujourd'hui

Ouverture (*open data*)

Ouverture des **données publiques** :

- En France

- ▶ 1978 : Loi sur la liberté d'accès aux documents administratifs
- ▶ 2003 : directive européenne incitative (faciliter la réutilisation info.)
- ▶ 2008 : directive INSPIRE pour mettre en oeuvre les politiques enviro.
 - ★ concerne données géographiques numériques publiques
 - ★ obligation des autorités publiques de les rendre accessibles sur internet
 - ★ principes : collecte unique, cataloguer, partager, rendre interopérable
- ▶ 2011 : mission interministérielle *Etalab* dédiée à l'ouverture
 - ★ Mise en ligne du portail <https://www.data.gouv.fr>
- ▶ 2016 : loi pour une république numérique (ouverture par défaut des données publiques)

2. Le monde des données aujourd'hui

Ouverture (*open data*)

Ouverture des **données publiques** :

- En France (suite)
 - ▶ Mise à disposition sur internet de toute donnée produite ou détenue par l'administration dans le cadre de sa mission de service public
 - ▶ Sont exclues : informations nominatives, personnelles, protégées
 - ▶ Libre accès (s'impose à l'état), parfois payant, réutilisable gratuitement
 - ▶ Sous *licence ouverte*
 - ▶ Diffusion sous forme de fichiers (CSV, JSON, XML...) ou flux (via des API) pour un usage automatisé

2. Le monde des données aujourd'hui

Ouverture (*open data*)

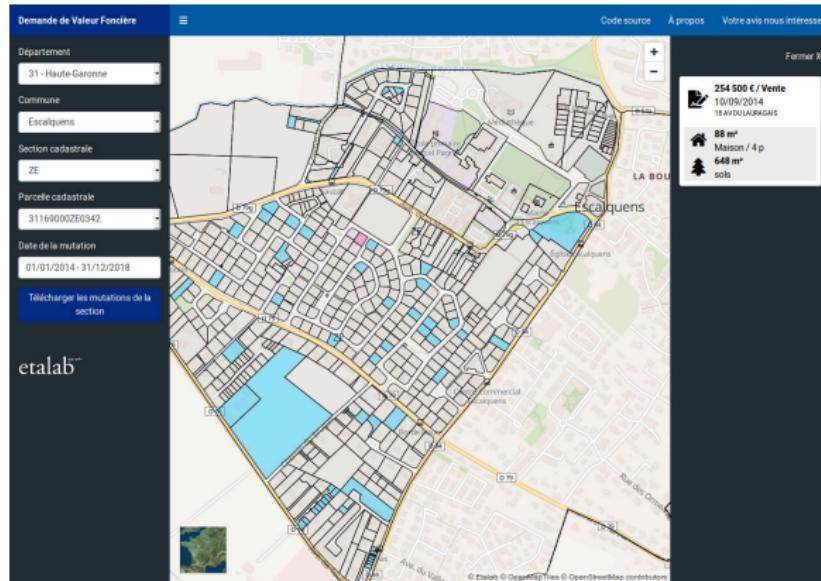
Ouvrir pourquoi ? (Gouigoux 2016)

- Transparency
 - ▶ Informer davantage sur ce que fait le gouvernement (activités, décisions)
 - ▶ "Renforcer davantage la démocratie et son efficacité" (B. Obama)
- Valorisation économique
 - ▶ Favoriser l'innovation, inventer de nouveaux services et usages
 - ▶ "Prolonger l'action de l'état" (Etalab)
- Valorisation interne (par l'état)
 - ▶ Administrations bénéficiaires (circulation d'informations)
 - ▶ Partage inter-services, qualité améliorée, intégration

2. Le monde des données aujourd'hui

Ouverture (*open data*) : exemple d'application

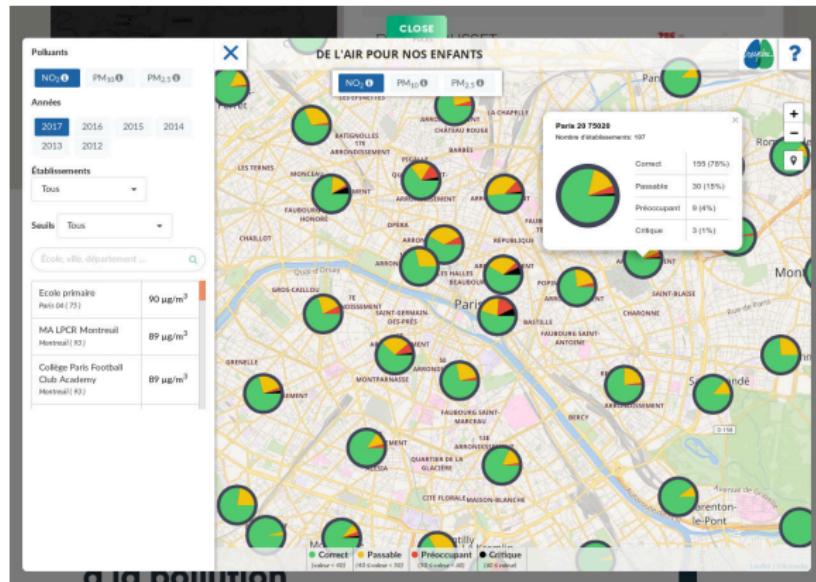
- Explorateur des valeurs foncières : <https://app.dvf.etalab.gouv.fr/>
 - ▶ Données DVF sur les mutations à titre onéreux (DG finances publiques)
 - ▶ But : rendre plus transparent le marché foncier



2. Le monde des données aujourd'hui

Ouverture (*open data*) : exemple d'application

- Explorateur de la pollution de l'air dans les écoles : *lien*
 - ▶ 100 millions de données Airparif
 - ▶ But : informer et sensibiliser le citoyen (transparence)

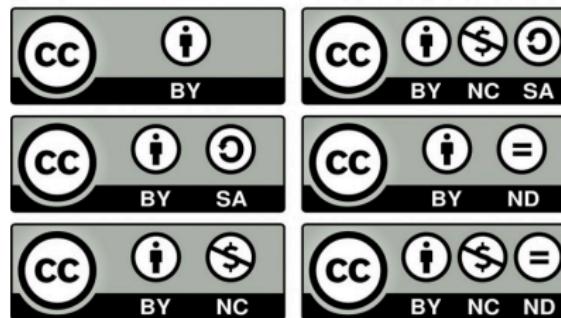


2. Le monde des données aujourd'hui

Ouverture (*open data*)

Les **licences** ouvertes :

- **Creative Commons** (CC) : des niveaux de licences standardisées, à la carte, permettant ou non d'*attribuer* (citer), *copier*, *modifier*, *partager*, *diffuser*, *distribuer*, *exploiter*.



2. Le monde des données aujourd'hui

Ouverture (*open data*)

The infographic illustrates the Creative Commons license spectrum, ranging from the most permissive (CC0) to the most restrictive (CC BY-NC-ND). A vertical color bar on the left transitions from green at the top to red at the bottom, with a grey band in the middle. To the right of the bar, each license is represented by a circular icon containing symbols for attribution (person), non-commercial use (dollar sign crossed out), and no derivatives (circle with a slash). Below each icon, the license name and a brief description are provided.

LICENCE	SYMBOLS	DÉCRITON
DOMAINE PUBLIC CC 0	Ø	Vous êtes autorisé à partager et modifier l'œuvre quelque soit votre usage, y compris commercial.
ATTRIBUTION CC BY	Ø	Vous êtes autorisé à partager et modifier l'œuvre quelque soit votre usage, y compris commercial, à condition de créditer l'auteur.
ATTRIBUTION - PARTAGE DANS LES MEMES CONDITIONS CC BY-SA	Ø Ø	Vous êtes autorisé à partager et modifier l'œuvre quelque soit votre usage, y compris commercial, à condition de créditer l'auteur et de partager les modifications sous la même licence.
ATTRIBUTION - PAS DE MODIFICATIONS CC BY-ND	Ø =	Vous êtes autorisé à partager et modifier l'œuvre quelque soit votre usage, y compris commercial, à condition de créditer l'auteur et de ne pas modifier ou transformer l'œuvre.
ATTRIBUTION - PAS D'UTILISATION COMMERCIALE CC BY-NC	Ø \$	Vous êtes autorisé à partager et modifier l'œuvre, à condition de créditer l'auteur et de ne pas faire d'utilisation commerciale.
ATTRIBUTION - PAS D'UTILISATION COMMERCIALE - PARTAGE DANS LES MEMES CONDITIONS CC BY-NC-SA	Ø \$ Ø	Vous êtes autorisé à partager et modifier l'œuvre, à condition de créditer l'auteur et de utiliser la même licence si vous modifiez l'œuvre et de ne pas faire d'utilisation commerciale.
ATTRIBUTION - PAS D'UTILISATION COMMERCIALE - PAS DE MODIFICATIONS CC BY-NC-ND	Ø \$ =	Vous êtes autorisé à partager l'œuvre, à condition de créditer l'auteur et de ne pas modifier ou transformer l'œuvre et de ne pas faire d'utilisation commerciale.

+ OUVERTURES + RESTRICTIVES

Cette image est sous licence CC BY SA
<https://creativecommons.org/licenses/by-sa/3.0/>

www.auboutdufil.com
Traduit et modifié de la source : <http://fitter.com/blog/how-to-attrbute-creative-commons-photos/>

2. Le monde des données aujourd'hui

Ouverture (*open data*)

- Licence Ouverte / Open Licence (Etat français)

- ▶ créée par *Etalab* pour encadrer l'ouverture des *données publiques* en France.
- ▶ permet de *copier, modifier, partager, distribuer, exploiter*
- ▶ compatible avec les licences CC-BY (attribution/paternité obligatoire)



LICENCE OUVERTE
OPEN LICENCE

(Source : [Etalab](#))

2. Le monde des données aujourd'hui

Ouverture (*open data*)

Le **mouvement d'ouverture** se généralise
et ne se limite pas aux données !



(Source : Wikipedia & Medium)

2. Le monde des données aujourd'hui

Déluge (*big data*)

2. Le monde des données aujourd’hui

Déluge (*big data*)

Ce qu'il se passe en 1 minute sur Internet (année 2020)...



2. Le monde des données aujourd'hui

Déluge (*big data*)

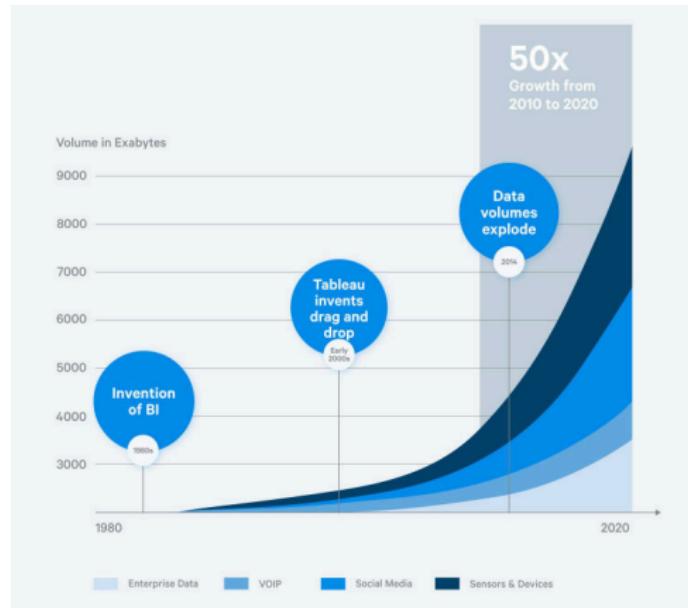


1To = 1000 Go ; 1Po = 1000 To ; 1 Eo = 1000 Po ; 1 Zo = 1000 Eo...

2. Le monde des données aujourd'hui

Déluge (*big data*)

Toujours plus *volumineux*, plus *vite*, plus *varié*. . .
et non structuré à 80 % ! (emails, vidéos, audios, réseaux sociaux. . .)



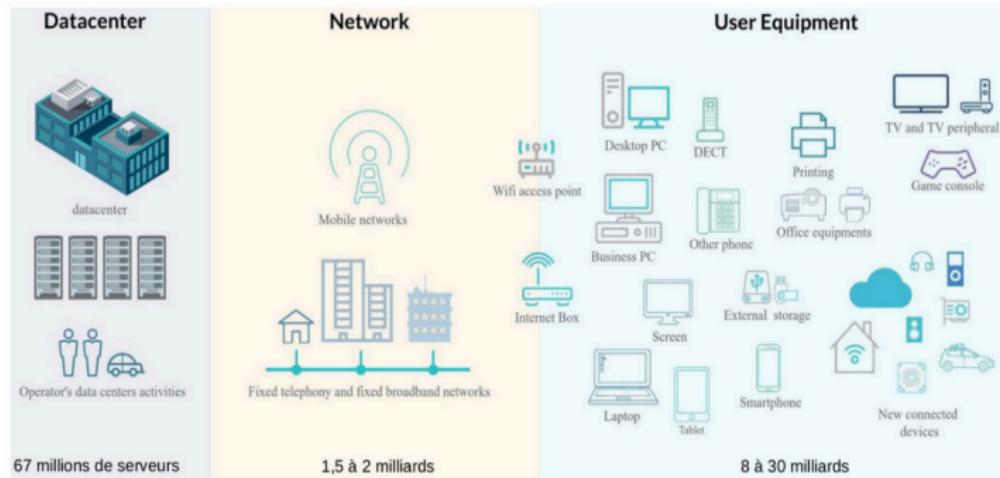
(Source : Periscope Data)

2. Le monde des données aujourd'hui

Déluge (*big data*)

Au passage, quelques chiffres à garder en tête... (d'après CNRS 2018) :

- Secteur du numérique : de 6 à 10 % de la consommation mondiale d'électricité



Malmodin, J.; Lundén, D., The Energy and Carbon Footprint of the Global ICT and E&M Sectors 2010–2015, CC BY 4.0, Sustainability, 2018
Frédéric Bordage, GreenIT.fr, Empreinte environnementale du numérique mondial, CC By-NC-ND, 2019 ; Graphique : D. Benaben ; <https://www.visme.co/>

2. Le monde des données aujourd'hui

Déluge (*big data*)

- Emissions de gaz à effet de serre générées par le numérique : 4%

En 2019, la hiérarchie des sources d'impacts est la suivante, par ordre décroissant d'importance :

1. Fabrication des équipements utilisateurs ;
2. Consommation électrique des équipements utilisateurs ;
3. Consommation électrique du réseau ;
4. Consommation électrique des centres informatiques ;
5. Fabrication des équipements réseau ;
6. Fabrication des équipements hébergés par les centres informatiques (serveurs, etc.).

F. Bordage, Empreinte environnementale du numérique mondiale, GreenIT (2019)

2. Le monde des données aujourd'hui

Déluge (*big data*)

Mais aussi... (d'après ADEME 2014) :

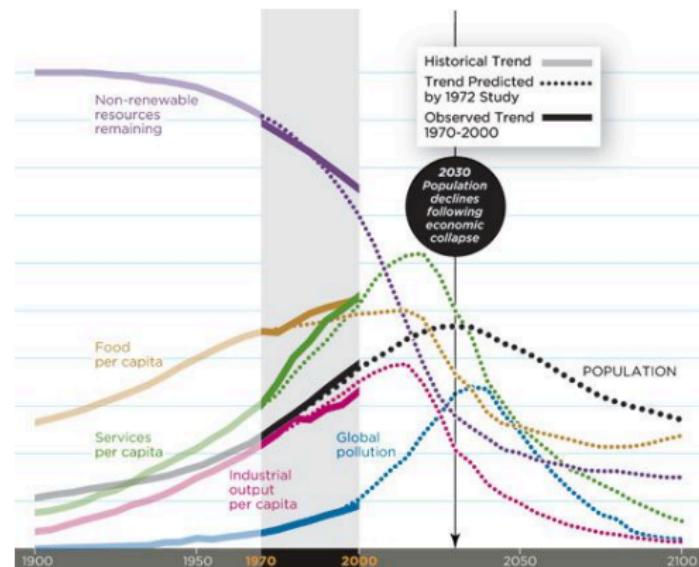
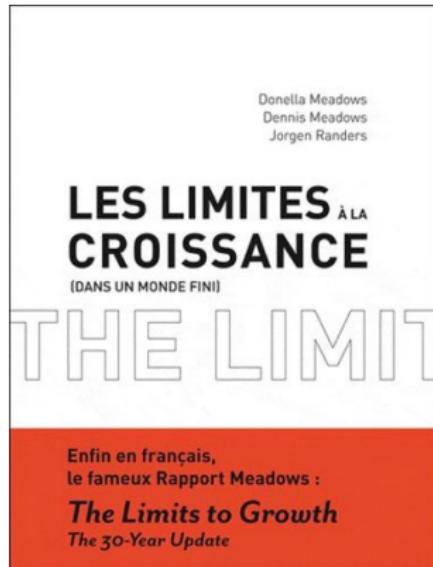
- 1 mail de 1 Mo = ~20 gr équivalent CO₂ émis
 - ▶ En 1h, 12 milliards de mails mondiaux = 24 000 A/R Paris-New York
- 1 requête Google = 2,8 gr équivalent CO₂ émis
- 1 requête avec consultation résultats = 10 gr équivalent CO₂ émis
- préférable d'imprimer que de lire sur écran si :
 - ▶ durée lecture > 2min/page et impression N&B
 - ▶ durée lecture > 8min/page et impression Couleur

A consulter : La face cachée du numérique (ADEME)

2. Le monde des données aujourd'hui

Déluge (*big data*)

Et après... ?

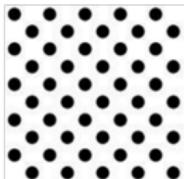
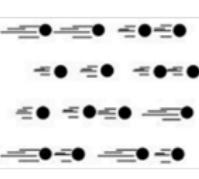
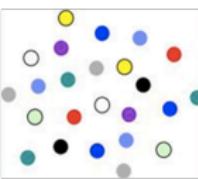
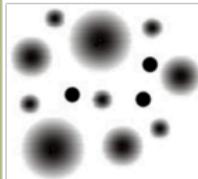


(Source : *Les limites à la croissance*)

2. Le monde des données aujourd'hui

Déluge (*big data*) : caractéristiques

Big in... (les 5V)

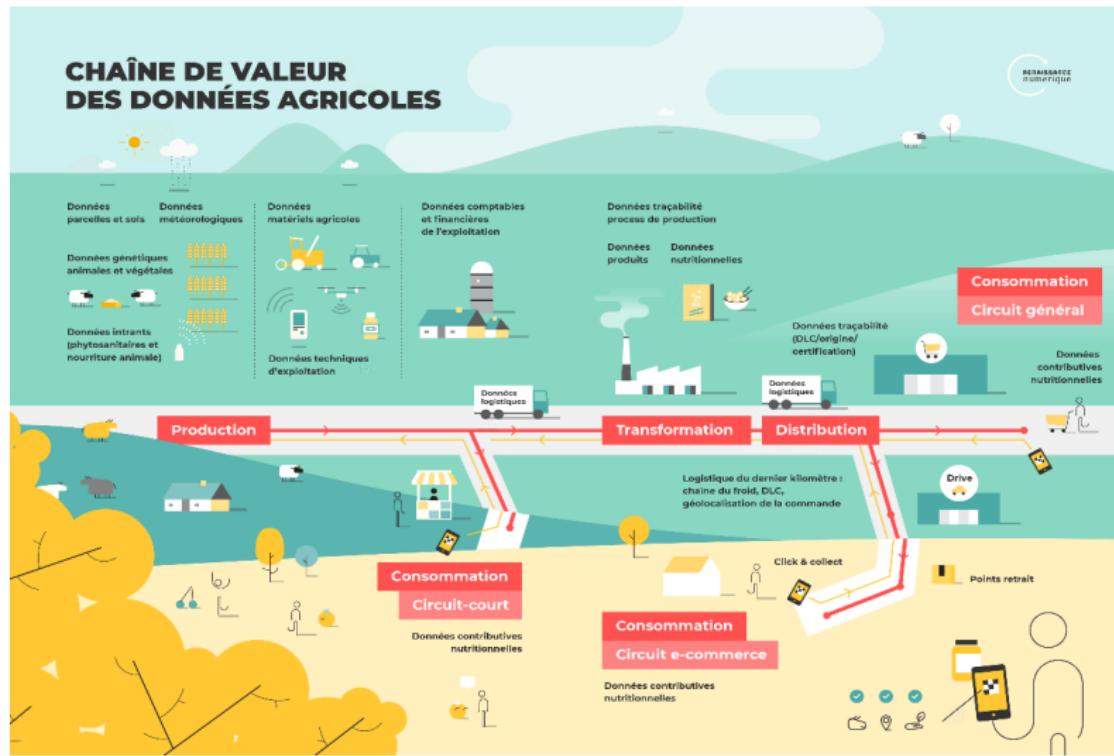
Volume	Velocity	Variety	Veracity	Value
				
Data at Rest Terabytes to Exabytes of existing data to process	Data in Motion Streaming data, requiring milliseconds to seconds to respond	Data in Many Forms Structured, unstructured, text, multimedia,...	Data in Doubt Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations	Data into Money Business models can be associated to the data

Adapted by a post of Michael Walker on 28 November 2012

(Source : [Information Catalyst](#))

2. Le monde des données aujourd'hui

Déluge (*big data*) : en agriculture



2. Le monde des données aujourd'hui

Déluge (*big data*) : en agriculture

Des données qui se multiplient à tous les niveaux
(“*de la fourche à la fourchette*”) :

- **Production :**

- ▶ données agronomiques (semences, génétique animale, intrants...)
- ▶ données météorologiques (capteurs, services web)
- ▶ données parcellaires, sol, environnement (limites, cadastre, pédologie, plan de fertilisation, physiologie des cultures par drône...),
- ▶ données liées au matériel (robot de traite, épandeur connecté...)
- ▶ données de l'exploitation (comptabilité, subventions PAC, registres d'élevage...)

2. Le monde des données aujourd'hui

Déluge (*big data*) : en agriculture

Des données qui se multiplient à tous les niveaux
(“*de la fourche à la fourchette*”) :

- **Transformation :**

- ▶ données produits (ingrédients, valeur nutritionnelle...)
 - ▶ données sur processus, traçabilité (industrielle, sanitaire)

- **Distribution :**

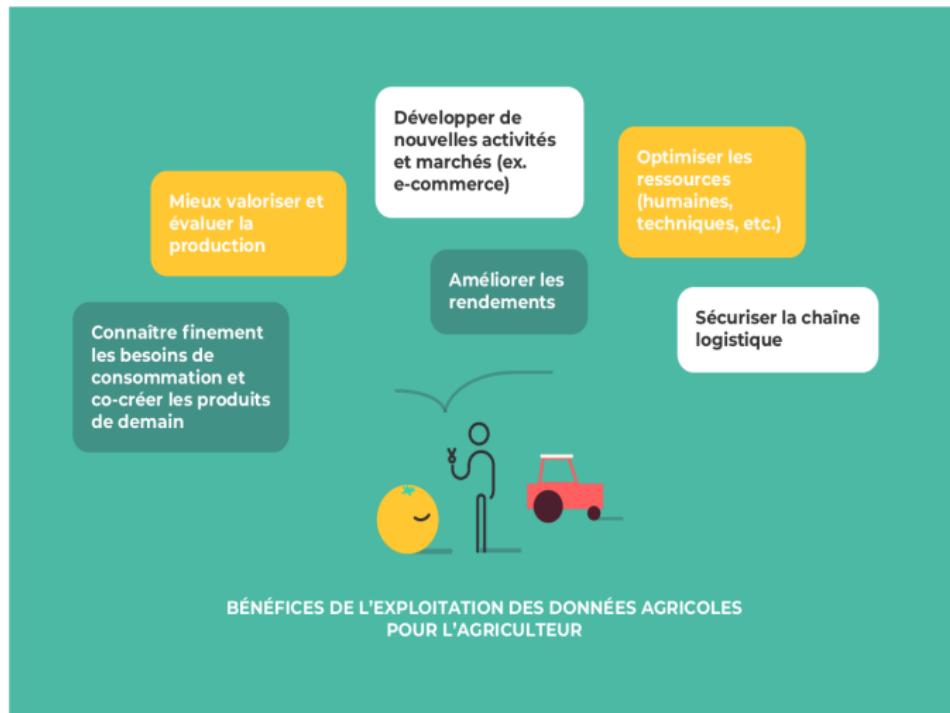
- ▶ volume, destination, chaîne de distribution, mode de réception

- **Consommation :**

- ▶ profils des consommateurs, attentes...

2. Le monde des données aujourd'hui

Déluge (*big data*) : en agriculture



(Source : Renaissance numérique, 2018)

2. Le monde des données aujourd'hui

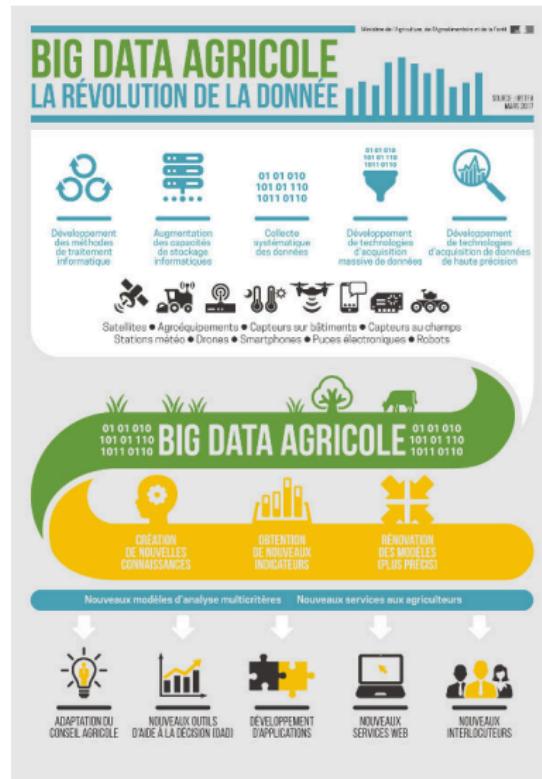
Déluge (*big data*) : en agriculture



(Source : Renaissance numérique, 2018)

2. Le monde des données aujourd'hui

Déluge (*big data*) : en agriculture



2. Le monde des données aujourd'hui

Déluge (*big data*)

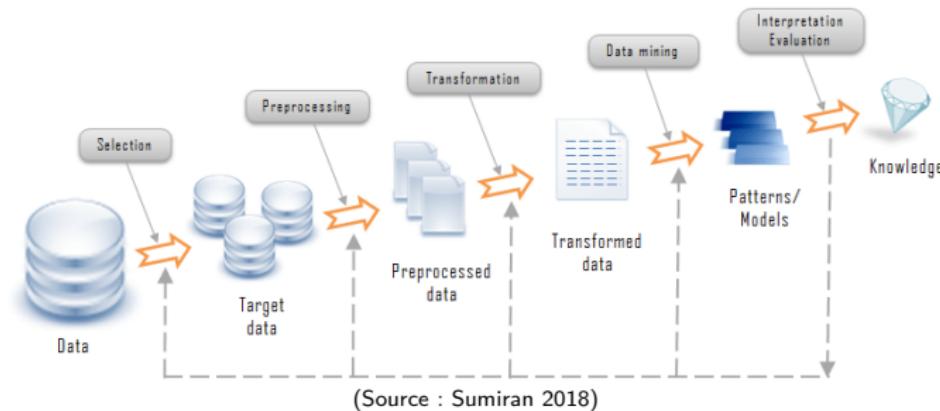


Video : [comment les nouvelles technologies transforment l'agriculture ?](#)

2. Le monde des données aujourd'hui

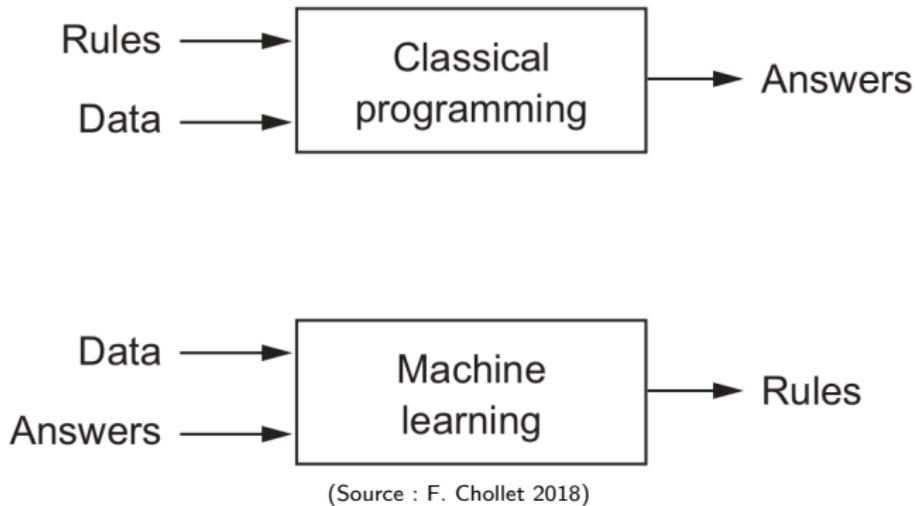
Fouille (*data mining*)

- Démarche d'exploration et d'**extraction automatique** de connaissances (*patterns*) à partir de masses de données
 - ▶ Connaissances implicites, non triviales, inconnues et sans doute utiles
 - ▶ Repose sur des techniques statistiques et informatiques issues de l'IA (apprentissage machine)



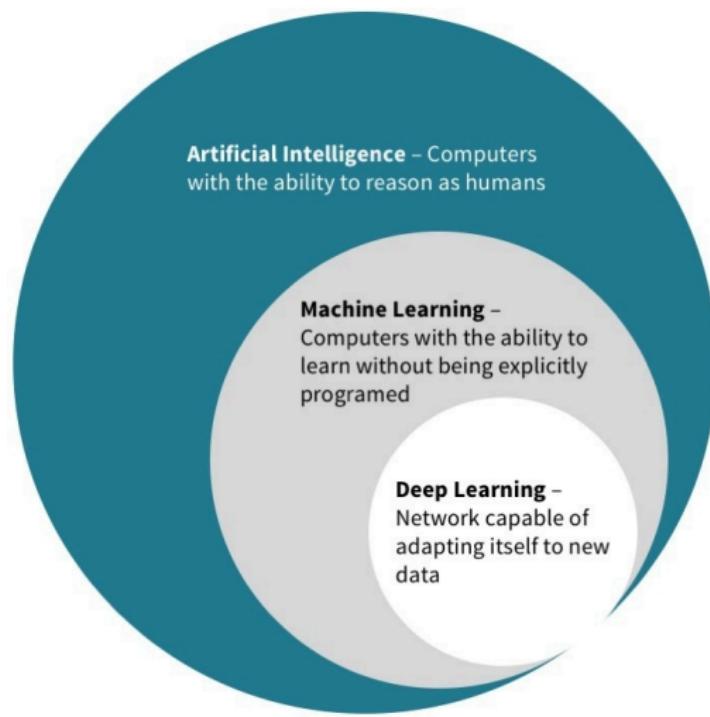
2. Le monde des données aujourd'hui

Fouille (*data mining*)



2. Le monde des données aujourd'hui

Fouille (*data mining*)



2. Le monde des données aujourd'hui

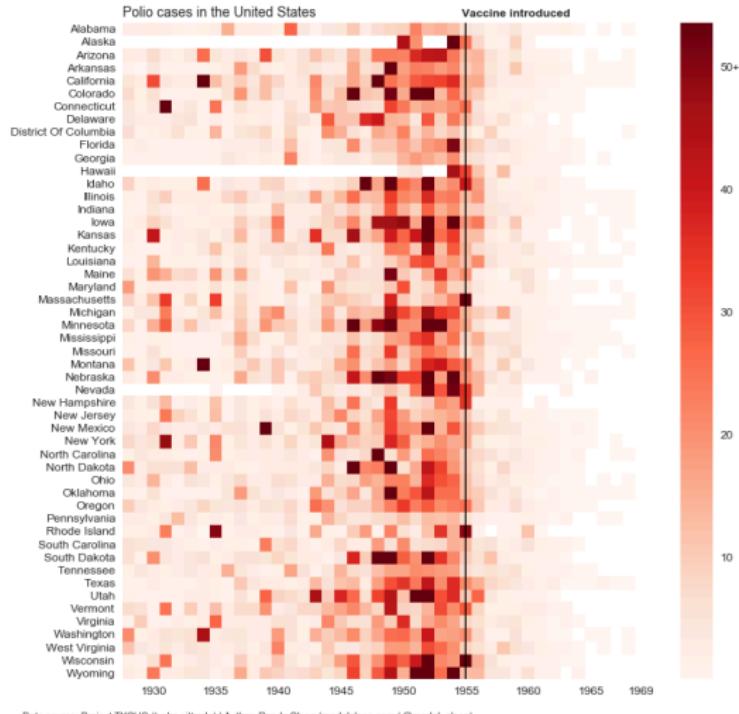
Visualisation (*dataviz*)

Rôles de la visualisation :

- Faciliter l'**exploration**, la **compréhension** et l'**analyse** de la masse des données
 - ▶ Rendre l'information complexe lisible
 - ▶ Simplifier tout en restant juste et précis
- **Communiquer** un message de façon graphique et esthétique
 - ▶ Raconter une histoire de façon attractive (cf. infographie)

2. Le monde des données aujourd'hui

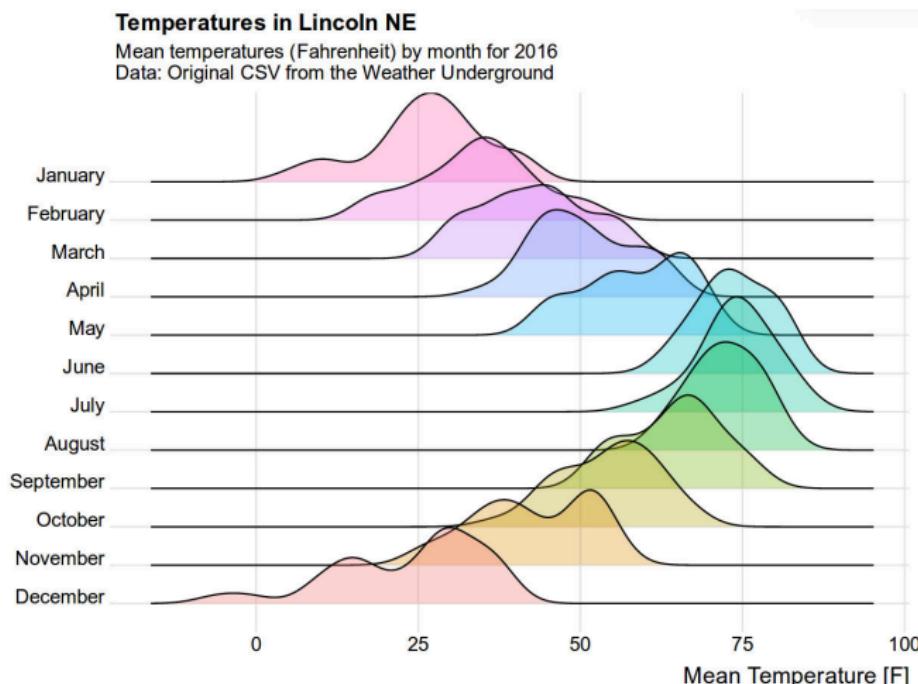
Visualisation (*dataviz*)



Heatmap - (Source : Randy Olson, blog)

2. Le monde des données aujourd'hui

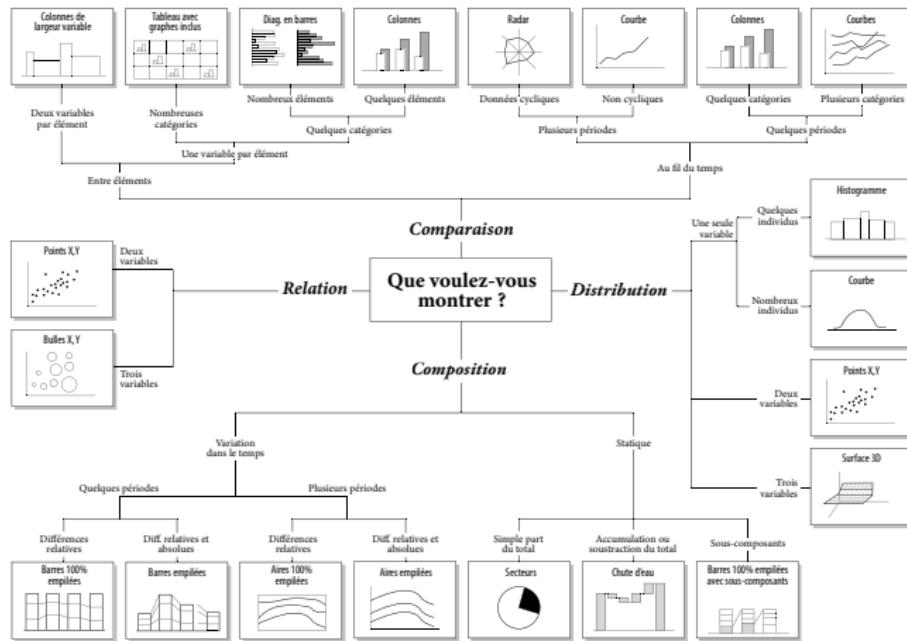
Visualisation (*dataviz*)



2. Le monde des données aujourd'hui

Visualisation (*dataviz*)

Arbre de décision des graphiques



2. Le monde des données aujourd'hui

Visualisation (*dataviz*)



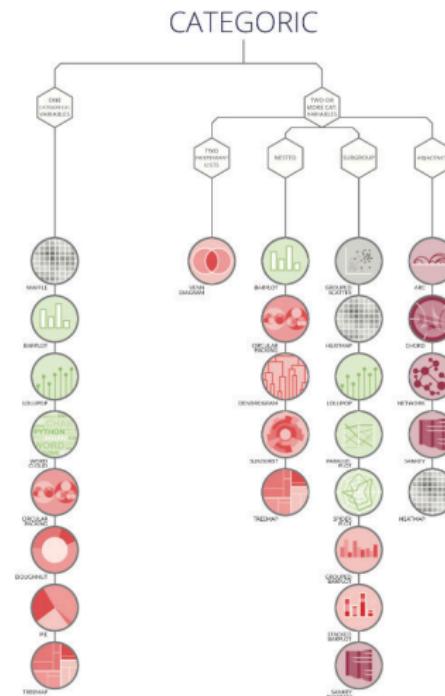
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

- ① Identify what type of data you have.
- ② Go to the corresponding decision tree and follow it down to a set of possible charts.
- ③ Choose the chart from the set that will suit your data and your needs best.

Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

data-to-viz.com

(Source : [from Data to Viz, Y. Holtz](#))



2. Le monde des données aujourd'hui

Visualisation (*dataviz*)

Data Visualization with ggplot2

Cheat Sheet



Basics

ggplot2 is based on the **grammar of graphics**, the idea that you can build every graph from the same few components: a **data set**, a set of **geom**-visual marks that represent data points, and a **coordinate system**.



To display data values, many variables in the data set to aesthetic properties of the geom like **size**, **color**, and **x** and **y** locations.



Build a graph with **qplot()** or **ggplot()**

`qplot(x=x, y=y, color = c, data = mpg, geom = "point")`
Creates a complete plot with given data, geom, and mappings. Supplies many useful defaults.

`ggplot(data = mpg, aes = cty, y = hwy)`
Begin a plot that you finish by adding layers to. No defaults, but provides more control than qplot().

`ggplot(mpg, aes(hwy, cty)) +
 geom_point(aes(color = cyl)) +
 geom_smooth(method = "lm") +
 geom_text(aes(label = as.character(mpg))) +
 scale_color_gradient() +
 theme_bw()`
Add a new layer to a plot with a **geom_**"**0**" or **stat_**"**0**" function. Each provides a geom, a set of aesthetic mappings, and a default stat and position adjustment.

last_plot()

Returns the last plot

`ggave("plot.png", width = 5, height = 5)`

Saves last plot as 5 x 5 file named "plot.png" in working directory. Matches file type to file extension.

Geoms - Use a geom to represent data points, use the geom's aesthetic properties to represent variables. Each function returns a layer.

One Variable

Continuous

```
a <- ggplot(mpg, aes(cty, hwy))  
a + geom_area(stat = "bin")  
a + geom_area(aes(y = density, stat = "bin")  
a + geom_density(kernel = "gaussian")  
a + geom_dotplot()  
a + geom_freqpoly()  
a + geom_histogram(binwidth = 5)  
a + geom_hex()  
a + geom_rug(sides = "BL")  
a + geom_smooth(model = lm)  
a + geom_step(direction = "Tu")
```

Discrete

```
b <- ggplot(mpg, aes(lt))  
b + geom_bar()  
b + geom_text(aes(label = lt))
```

Graphical Primitives

```
c <- ggplot(mpg, aes(long, lat))  
c + geom_polygon(aes(group = group))  
c + geom_rect(aes(xmin = long, xmax = long,  
                  ymin = lat, ymax = lat))
```

```
d <- ggplot(economics, aes(date, unemployment))  
d + geom_path(lineend = "butt",  
             linejoin = "round", lineinmit = 1)  
d + geom_rect(aes(xmin = date - 90, xmax = date + 90,  
                  ymin = unemployment - 900,  
                  ymax = unemployment + 900))  
d + geom_ribbon(aes(ymin = unemployment - 900,  
                     ymax = unemployment + 900))
```

```
e <- ggplot(seals, aes(x = long, y = lat))
```

```
e + geom_segment(aes(  
                 xend = long + delta_long,  
                 yend = lat + delta_lat))  
e + geom_rect(aes(xmin = long, ymin = lat -  
                  delta_lat, xmax = long, ymax = lat))
```

```
f <- ggplot(seals, aes(x = long, y = lat))  
f + geom_point(aes(x = long, y = lat))  
f + geom_text(aes(x = long, y = lat))  
f + geom_hex()
```

Continuous X, Continuous Y

```
f + geom_blank()  
f + geom_jitter()  
f + geom_point()  
f + geom_quantile()  
f + geom_rug(sides = "BL")  
f + geom_smooth()  
f + geom_text(aes(label = cty))  
A + geom_bar()
```

Discrete X, Continuous Y

```
g <- ggplot(mpg, aes(class, hwy))  
g + geom_bar(stat = "identity")  
g + geom_boxplot()  
g + geom_errorbar()  
g + geom_dotplot(binaxis = "y",  
                 stackdir = "center")  
g + geom_hex()  
g + geom_pointrangle()  
g + geom_violin(scale = "area")
```

Discrete X, Discrete Y

```
h <- ggplot(diamonds, aes(cut, color))  
h + geom_jitter()  
h + geom_rect()
```

```
i <- ggplot(seals, aes(delta_long * 2 + delta_lat * 2))  
i <- ggplot(seals, aes(delta_lat))  
i + geom_contour(aes(z = z))  
i + geom_hex()
```

Two Variables

Continuous Bivariate Distribution

```
j <- ggplot(movies, aes(year, rating))  
j + geom_bnd2d(binwidth = c(5, 0.5))  
j + geom_hex()  
j + geom_hex2d()
```

Continuous Function

```
j <- ggplot(economics, aes(date, unemployed))  
j + geom_area()  
j + geom_line()  
j + geom_step(direction = "Tu")
```

Visualizing error

```
df <- data.frame(rho = c("A", "B"), fit = 4.5, se = 1.2)  
k <- ggplot(df, aes(rho, fit, ymin = fit - se, ymax = fit + se))
```

```
k + geom_crossbar(fatten = 2)  
k + geom_errorbar()  
k + geom_hexrange()  
k + geom_pointrange()  
k + geom_violin()
```

Mapa

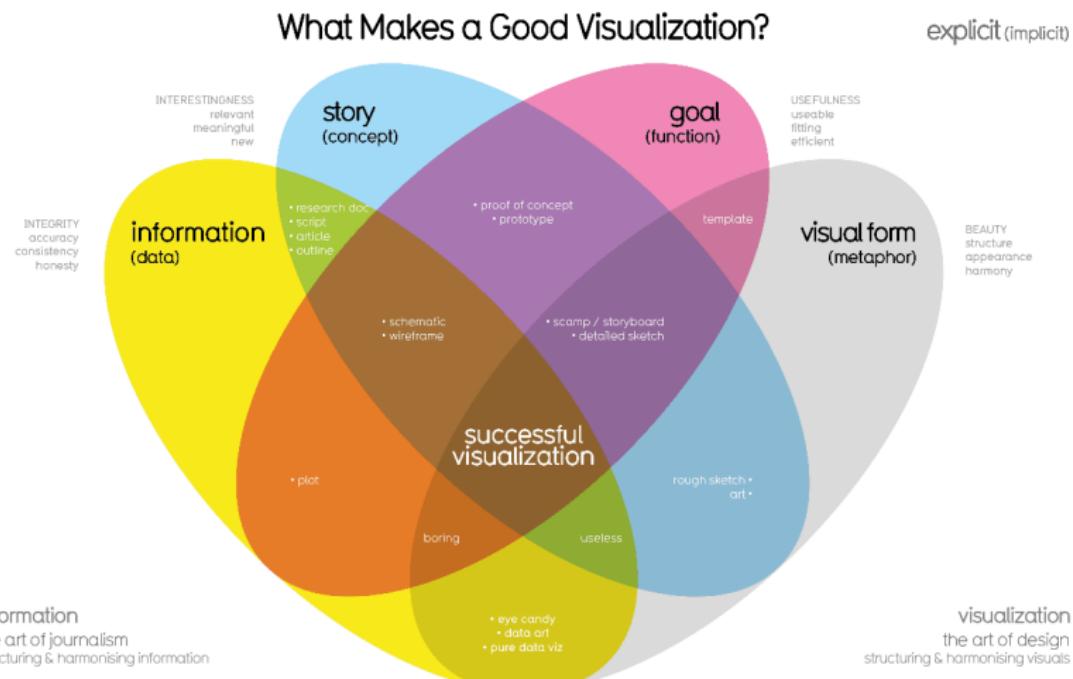
```
data <- dta.frame(murder = USArrests$Murder,  
                  state = USArrests$State,  
                  map = USArrests$Murder)  
map <- map_usa  
l <- ggplot(la, aes(la, mu = murder)) +  
  geom_map(aes(map_id = state, map = map)) +  
  expand_limits(x = map$x, y = map$y)
```

Three Variables

```
m <- geom_raster(aes(fill = z), njust = 0.5,  
                  vjust = 0.5, interpolate = FALSE)  
m + geom_hex()  
m + geom_tile(aes(fill = z))
```

2. Le monde des données aujourd'hui

Visualisation (*dataviz*)



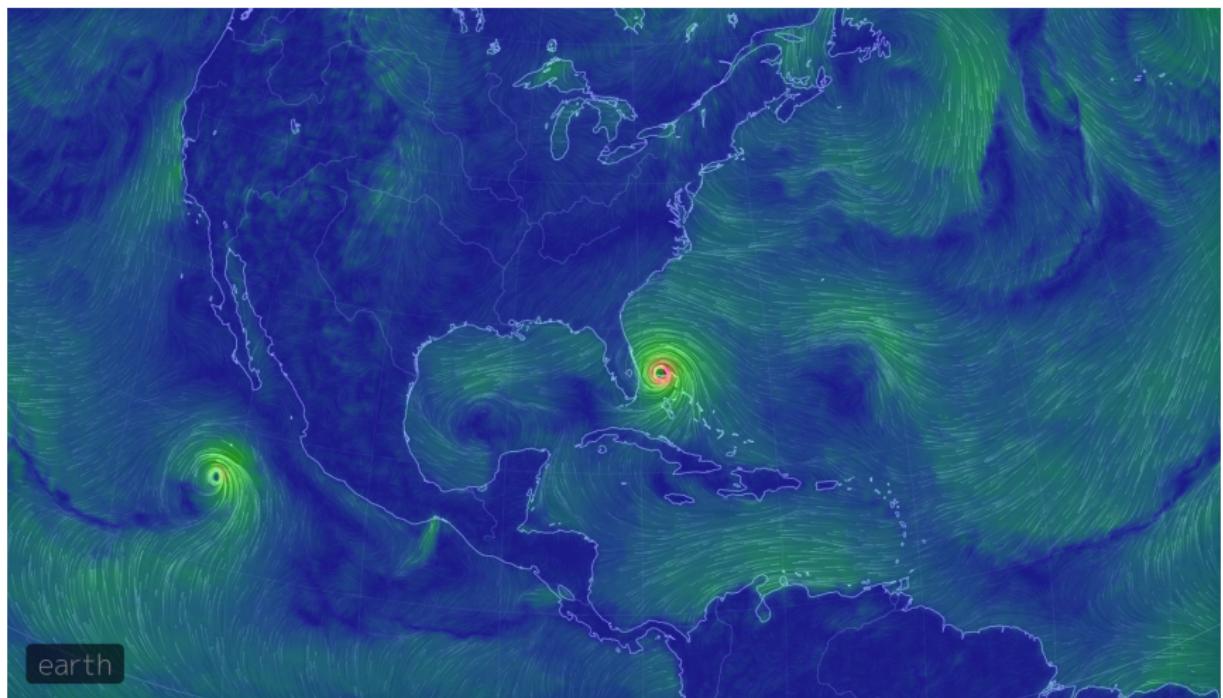
David McCandless
InformationisBeautiful.net

taken from new book
Knowledge is Beautiful

find out more
bitly/KIB_Books

2. Le monde des données aujourd'hui

Visualisation (*dataviz*)



(Source : <https://earth.nullschool.net/>)

2. Le monde des données aujourd'hui

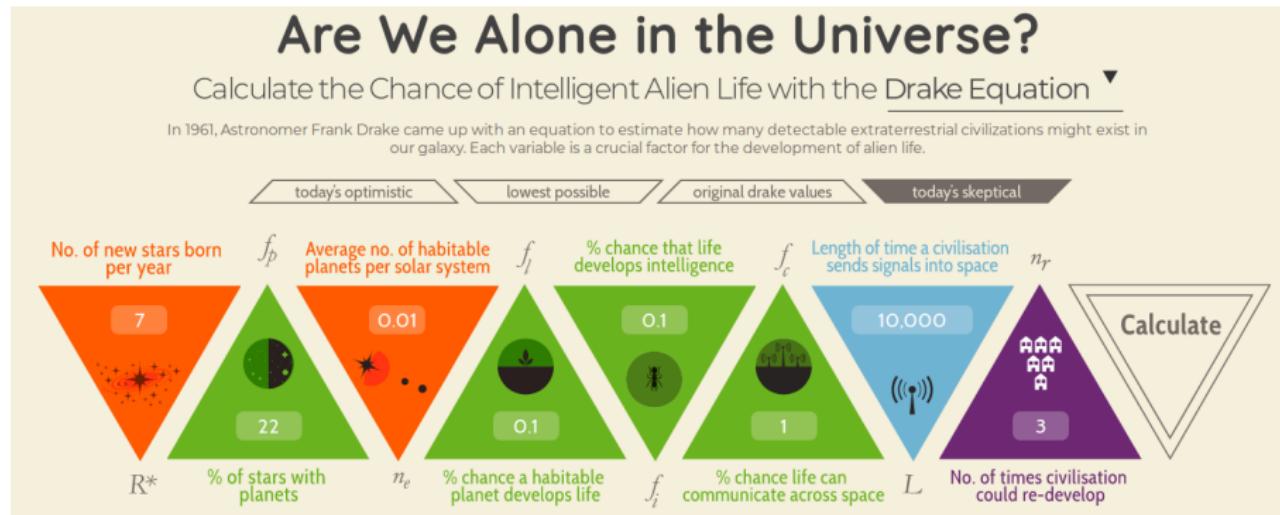
Visualisation (*dataviz*)



(Source : <https://www.data-to-viz.com>)

2. Le monde des données aujourd'hui

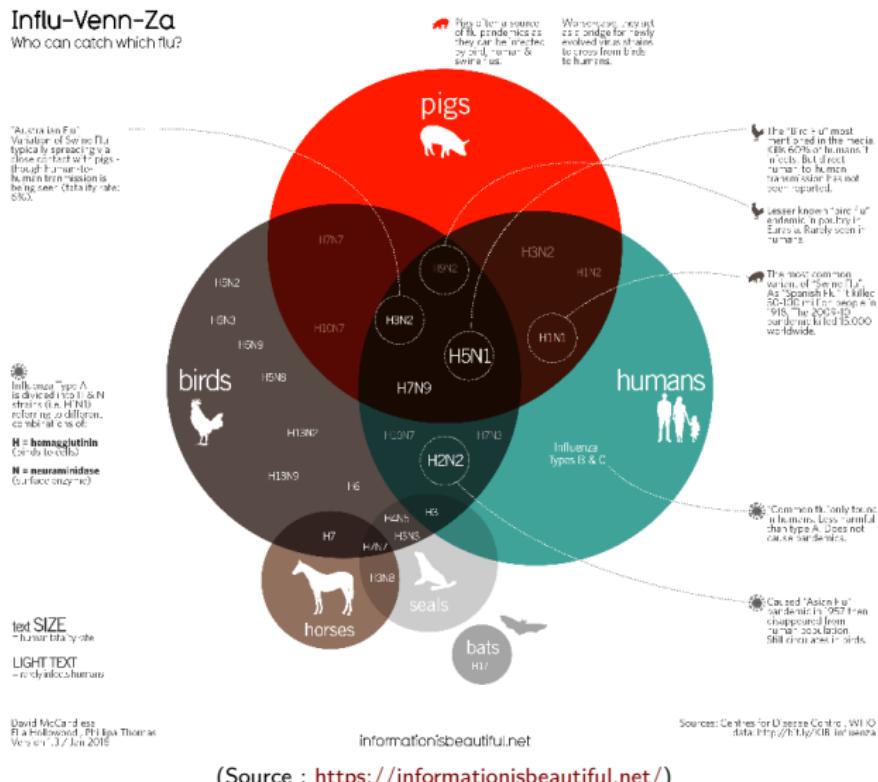
Visualisation (*dataviz*)



(Source : <https://informationisbeautiful.net/>)

2. Le monde des données aujourd'hui

Visualisation (*dataviz*)



2. Le monde des données aujourd'hui

Visualisation (*dataviz*)

What is the world's biggest cash crop?



(Source : <https://informationisbeautiful.net/>)

2. Le monde des données aujourd'hui

Visualisation (*dataviz*)

Quelques outils pour les graphiques et l'infographie (au-delà de *R*...) :

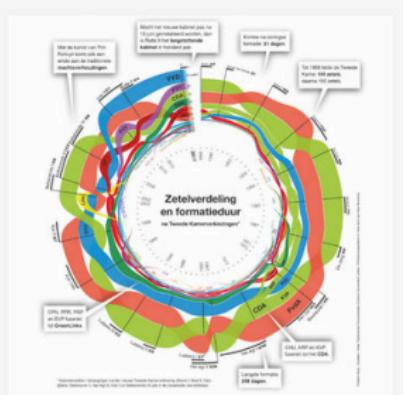
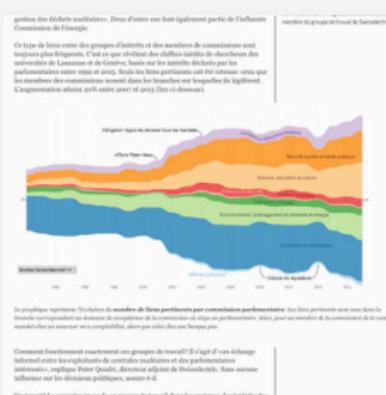
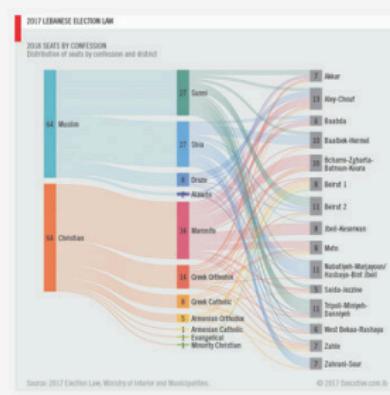


(Source : <https://d3js.org/>)

2. Le monde des données aujourd’hui

Visualisation (*dataviz*)

Quelques outils pour les graphiques et l'infographie (au-delà de R...) :



(Source : <https://rawgraphs.io/>)

2. Le monde des données aujourd'hui

Visualisation (*dataviz*)

Quelques outils pour les graphiques et l'infographie (au-delà de *R*. . .) :

easelly



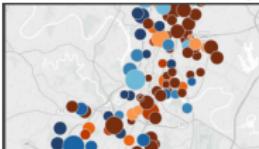
(Source : <https://www.easel.ly/>)

2. Le monde des données aujourd'hui

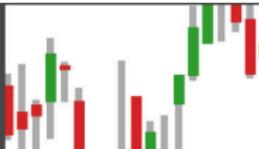
Visualisation (*dataviz*)

Quelques outils pour les graphiques et l'infographie (au-delà de *R*...) :

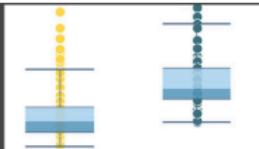
 + a b | e a u°



Suivi de l'impact des programmes de rétention des enseignants d'Austin
Les initiatives destinées à retenir les enseignants dans les établissements scolaires d'Austin sont-elles efficaces ?



Déterminez la volatilité des titres
Suivi du cours quotidien des actions sur un graphique en chandelier classique



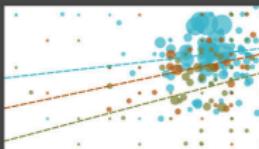
Dans quelles villes américaines les prix de l'immobilier sont-ils les plus élevés ?
Dans quelles villes des États-Unis le prix des biens vendus est-il le plus élevé ?



Facteurs de croissance à long terme pour l'économie américaine
Quels sont les leviers économiques sur le long terme ?



Transformez une analyse mathématique en visualisation interactive
Visualisations mathématiques attrayantes et interactives



Mesurez la satisfaction de la clientèle grâce aux données d'enquête
Comment visualiser les données d'une enquête de satisfaction plus efficacement ?

(Source : <https://www.tableau.com/fr-fr>)

2. Le monde des données aujourd'hui

Visualisation (*dataviz*)

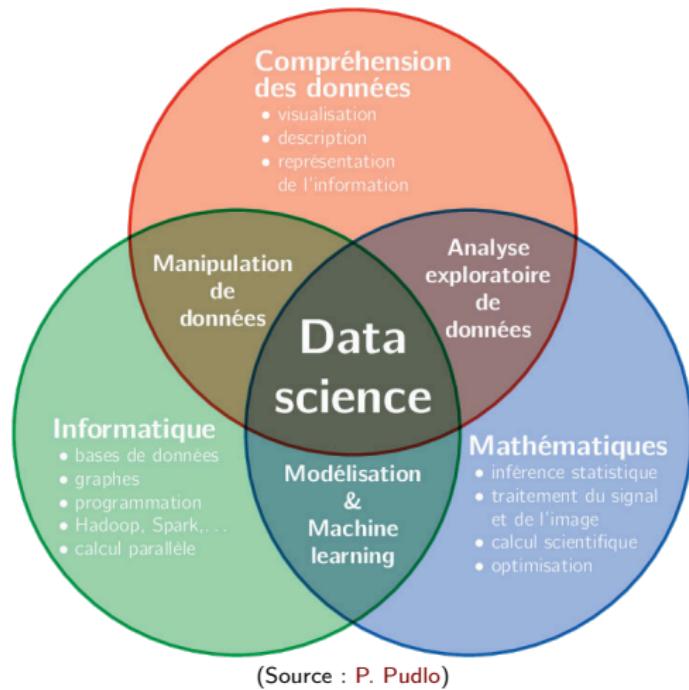
Quelques outils pour les graphiques et l'infographie (au-delà de *R*...) :



(Source : <https://infogram.com/>)

2. Le monde des données aujourd'hui

Vers un nouveau paradigme ? (*data science*)



2. Le monde des données aujourd'hui

Vers un nouveau paradigme ? (*data science*)

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21th century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.



MATH & STATISTICS <ul style="list-style-type: none">★ Machine learning★ Statistical modeling★ Experiment design★ Bayesian inference★ Supervised learning: decision trees, random forests, logistic regression★ Unsupervised learning: clustering, dimensionality reduction★ Optimization: gradient descent and variants	PROGRAMMING & DATABASE <ul style="list-style-type: none">★ Computer science fundamentals★ Scripting language e.g. Python★ Statistical computing package e.g. R★ Databases SQL and NoSQL★ Relational algebra★ Parallel databases and parallel query processing★ MapReduce concepts★ Hadoop and Hive/Pig★ Custom reducers★ Experience with xaaS like AWS
DOMAIN KNOWLEDGE & SOFT SKILLS <ul style="list-style-type: none">★ Passionate about the business★ Curious about data★ Influence without authority★ Hacker mindset★ Problem solver★ Strategic, proactive, creative, innovative and collaborative	COMMUNICATION & VISUALIZATION <ul style="list-style-type: none">★ Able to engage with senior management★ Story telling skills★ Translate data-driven insights into decisions and actions★ Visual art design★ R packages like ggplot or lattice★ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

3. Les sources de données

3. Les sources de données

Quelques **questions à se poser** avant de choisir :

- La source est-elle **fiable** ?
 - ▶ données institutionnelles ? données participatives ?
- La source est-elle **pérenne** ?
 - ▶ durée de vie ?
- Les données sont-elles **documentées** (*métadonnées*) ?
 - ▶ compréhension possible du contenu, processus de production...
- Les données répondent-elles au **besoin** (*fitness for use*) ?
 - ▶ type, résolution thématique, échelle, définition, exactitude, complétude...

3. Les sources de données

- Les données sont-elles **à jour** ?
 - ▶ quel rythme ? comment ?
- Les données sont-elles **ouvertes** ?
 - ▶ quels droits d'utilisation ? licence ? conditions d'accès ?
- Les données sont-elle **téléchargeables** ?
 - ▶ fichiers disponibles ? accès via un flux ?
- Quels **formats** disponibles ?
 - ▶ données structurées ?
- Les données sont-elles **référencées** (*catalogue*) ?
 - ▶ récupération automatique des métadonnées

3. Les sources de données

Quelques portails et producteurs incontournables

- Portail [data.gouv.fr](#) :

- ▶ Recense toutes les données publiques ouvertes françaises
- ▶ Thèmes : agriculture, économie et emploi, éducation, environnement et énergie, santé... .

The screenshot shows the data.gouv.fr homepage with a green header. The header includes the French flag logo, the text "data.gouv.fr", "Plateforme ouverte des données publiques françaises", and links for "Connexion / Inscription". Below the header, there's a navigation bar with links for "Données", "Réutilisations", "Organisations", "Tableau de bord", and "Documentation". On the left, there's a sidebar with a search bar and a list of categories: Agriculture et Alimentation, Culture, Communications, Comptes, Économie et Emploi, Éducation, Recherche, Formation, International, Europe, Environnement, Énergie, Logement, Santé et Social, Société, Droit, Institutions, and Territoires, Transports, Tourisme. The main content area has a green background with white text. It features the title "Agriculture et Alimentation" in large letters, followed by a subtext "Des données sur ce que vous mangez, du terroir à l'assiette.". It lists resources: "En savoir plus sur l'agriculture :" followed by three bullet points: "la page du ministère de l'agriculture et de l'alimentation sur data.gouv.fr", "le site du ministère de l'agriculture et de l'alimentation", and "l'article agriculture sur Wikipédia". Below this, it says "Thèmes associés :" and lists several tags: ALIMENTATION, FORETS, OGM, PAC, PECHÉ, AGROALIMENTAIRE, ALIMENTS, and AGRICULTURE.

3. Les sources de données

Quelques portails et producteurs incontournables

- Portail tout sur l'environnement :

- ▶ Toutes les informations publiques environnementales
- ▶ Recense des rapports, études, tableaux de données, cartes
- ▶ Thèmes : air, eau, biodiversité, littoral, sol...

The screenshot shows the homepage of the 'tout sur l'environnement' portal. At the top, there's a banner with the text 'tout sur l'environnement' and a small logo. Below the banner, the page title is 'Le portail de l'information publique environnementale'. A navigation bar includes links for 'Qui sommes-nous ?', 'Contributrice', 'Convention d'Aarhus', 'L'environnement en France', 'Accès aux données', and 'Actualités'. There are also social media icons for Facebook, Twitter, and LinkedIn. The main content area has a blue header 'LES RESSOURCES' with the text '115.542 ressources (rapports, études, contenus web, vidéos, tableaux de données, couches d'information géographique, etc.) sont à votre disposition sur le portail.' Below this, there's a search bar labeled 'RECHERCHER'. To the right is a map of France with regions outlined. On the left, there are three columns: 'TOUT SUR L'ENVIRONNEMENT' (with a sub-section about the Convention d'Aarhus), 'ACTUS ET DOSSIERS' (mentioning ADEME's call for climate research), and 'MILIEUX ET ENVIRONNEMENT'.

3. Les sources de données

Quelques portails et producteurs incontournables

• Portail NatureFrance :

- ▶ Site d'information sur les données pour la biodiversité
- ▶ Annuaire des portails nationaux et régionaux
- ▶ Lien vers GBIF, INPN, Carmen...

The screenshot shows the homepage of the NatureFrance portal. At the top, there is a navigation bar with links for Accueil, INP, OGB, Forum des données pour la biodiversité 2017, Gouvernance, Ressources, Réglementation, and Portails. Below the navigation bar, there is a search bar. The main content area features a banner for "NatureFrance, le site d'information sur les données pour la biodiversité". This banner includes a logo for "DATA LAB" and a link to "Communiqué de presse du 20 juillet 2017". To the right of the banner, there is a section titled "Observatoire national de la biodiversité" which describes the OGB as a federated system for biodiversity data in France, connected to various national inventories and indicators. Below this, there is a section titled "Système d'information sur la Nature et les Paysages" which describes the INPN as a federated system for nature and landscape data. The bottom of the page is divided into several boxes featuring logos and names of various organizations and platforms, such as Global Biodiversity Information Facility (GBIF), Inventaire national du patrimoine naturel (INPN), Plateforme de métadonnées de l'INPN, GINCO, Plateforme régionale générée par INPN, Plateforme de dépôt légal, and CARBEN.

3. Les sources de données

Quelques portails et producteurs incontournables

- Portail **Données et études statistiques** :

- ▶ Données statistiques du Ministère de la transition écologique et solidaire
- ▶ Thèmes : logement, construction, transport, énergie, environnement

The screenshot shows the 'Données et études statistiques' portal interface. At the top, there is a logo for 'Ministère de la Transition écologique et solidaire' and a search bar with the placeholder 'Pour le changement climatique, l'énergie, l'environnement, le logement, et les transports'. Below the search bar, the URL 'Accueil → Recherche' is visible.

RAPPEL DE LA RECHERCHE

Type de contenu : **Données**

X TOUT SUPPRIMER

MOTS-CLÉS

AFFINER LES RÉSULTATS

THÈMES

Transports

Logement

Résultats de recherche : 1430

Pertinence : Décroissant ▲ Date : Décroissant ▲

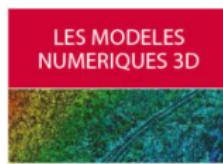
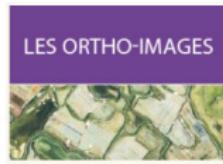
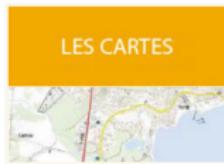
Données associées Les émissions de co2 liées à l'énergie de 1990 à 2017	XLSX (539.32 Ko)	CHANGEMENT CLIMATIQUE	05/09/2019
Données associées "Les émissions de CO2 liées à l'énergie de 1990 à 2017"	XLSX (539.32 Ko)	ENERGIE	04/09/2019
Immatriculations mensuelles de voitures particulières neuves par motorisation en France métropolitaine - 08/2019	XLS (18.44 Ko)	TRANSPORTS	04/09/2019
Immatriculations de voitures particulières neuves, données mensuelles complémentaires - 08/19	CSV (8.02 Ko)	TRANSPORTS	03/09/2019

3. Les sources de données

Quelques portails et producteurs incontournables

- BD et Geoportail de l'IGN :

- ▶ Référentiel national pour les cartes topographiques, orthophotographies, base de données vectorielles, MNT...
- ▶ Accès restreint (fichier) ou partiellement libre (flux - geoservices) selon licence
- ▶ Dans le cadre de la *licence enseignement*, accès gratuit à toutes les ressources ! Contacter votre enseignant selon besoin.



3. Les sources de données

Quelques portails et producteurs incontournables



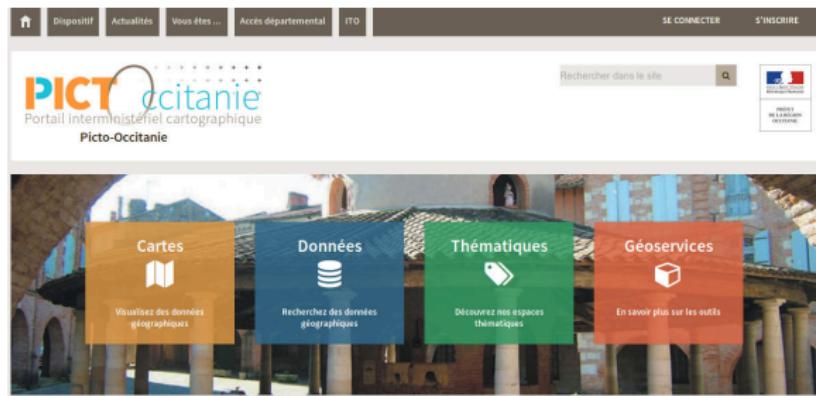
(Geoportal)

3. Les sources de données

Quelques portails et producteurs incontournables

- Portail Picto-Occitanie :

- ▶ Portail interministériel cartographique pour la région Occitanie
- ▶ Recensement des données et statistiques au sein d'un catalogue
- ▶ Accès libre à 80% (flux, fichier)
- ▶ Voir également l'outil Picto-Stat



3. Les sources de données

Quelques portails et producteurs incontournables

- Portail DATA.toulouse-metropole pour l'échelon local !

- ▶ Initiative existant pour de nombreuses métropoles
- ▶ Données locales sur de nombreuses thématiques (transport, météo, tourisme, espaces verts...)
- ▶ Accès libre (flux, fichier), catalogue, interface de visualisation



3. Les sources de données

1. Données cartographiques sur l'occupation des sols

3. Les sources de données

1. Données cartographiques sur l'occupation des sols

L'occupation du sol (***land cover***) désigne l'état biophysique ou la physionomie de la surface terrestre tel que l'eau, les surfaces bâties, la végétation herbacée, la végétation ligneuse...



Built-up area



Wetland vegetation



Water unspecified

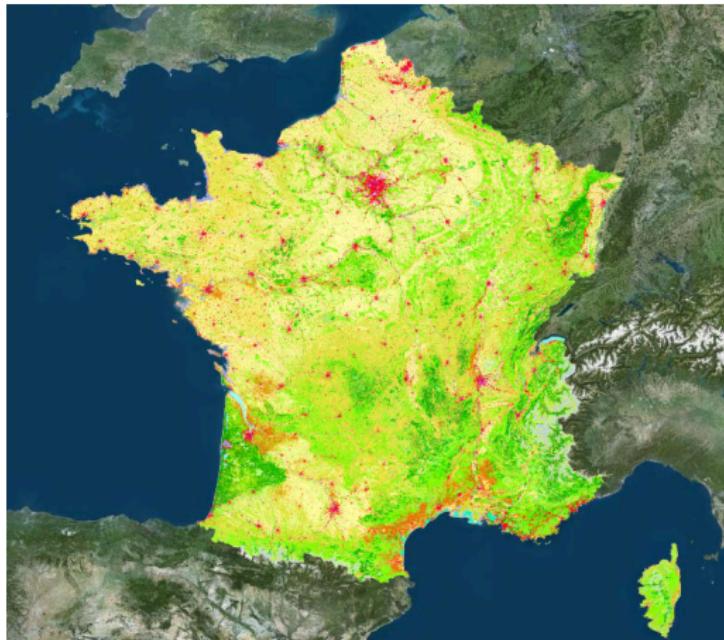
3. Les sources de données

1. Données cartographiques sur l'occupation des sols

- Base de données CORINE Land Cover (CLC) :
 - ▶ Harmonisée à l'échelle européenne, produite par l'EEA
 - ▶ Issue d'un traitement d'images satellitaires
 - ▶ Echelle d'usage : 1/100.000 (surface minimale 25 ha)
 - ▶ Résolution thématique élevée mais niveau de détail spatial faible
 - ▶ Accès libre (flux, fichier), actualisée (1990, 2000, 2006, 2012, 2018)
 - ▶ Très utilisée

3. Les sources de données

1. Données cartographiques sur l'occupation des sols



CORINE Land Cover

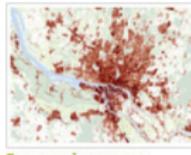
111 Tissu urbain continu
112 Tissu urbain discontinu
121 Zones industrielles et commerciales
122 Réseaux routier et ferroviaire et espaces associés
123 Zones portuaires
124 Aéropôts
131 Extraction de matériaux
132 Décharges
133 Chantiers
141 Espaces verts urbains
142 Équipements sportifs et de loisirs
211 Terres arables hors périmètres d'irrigation
212 Périmètres irrigués en permanence
213 Rizières
221 Vignobles
222 Vergers et petits fruits
223 Oliveraies
231 Prairies
241 Cultures annuelles associées aux cultures permanentes
242 Systèmes culturels et parcellaires complexes
243 Surfaces essentiellement agricoles, interrompues par des espaces naturels importants
244 Territoires agro-forestiers
311 Forêts de feuillus
312 Forêts de conifères
313 Forêts mélangées
321 Pelouses et pâturages naturels
322 Landes et broussailles
323 Végétation sclérophylle
324 Forêt et végétation arbustive en mutation
331 Plages, dunes et sable
332 Roches nues
333 Végétation clairsemée
334 Zones incendiées
335 Glaciers et neiges éternelles
411 Marais intérieurs
412 Tourbières
421 Marais maritimes
422 Marais salants
423 Zones intertidales
511 Cours et voies d'eau
512 Plans d'eau
521 Lagunes littorales
522 Estuaires
523 Mers et océans

3. Les sources de données

1. Données cartographiques sur l'occupation des sols

- Produit High Resolution Layers (HRL) :

- ▶ Harmonisé à l'échelle européenne, produite par l'EEA (programme Copernicus)
- ▶ Issu d'un traitement automatique d'images satellitaires
- ▶ Echelle d'usage : entre 1/25.000 et 1/50.000
- ▶ 8 thèmes (5+3)
- ▶ Résolution spatiale : 20 m (raster)
- ▶ Accès libre (fichier), deux millésimes (2012, 2015)
- ▶ Produit récent, en évolution



Imperviousness



Forests



Grassland



Water & Wetness



Coming Soon
Small Woody Features

3. Les sources de données

1. Données cartographiques sur l'occupation des sols

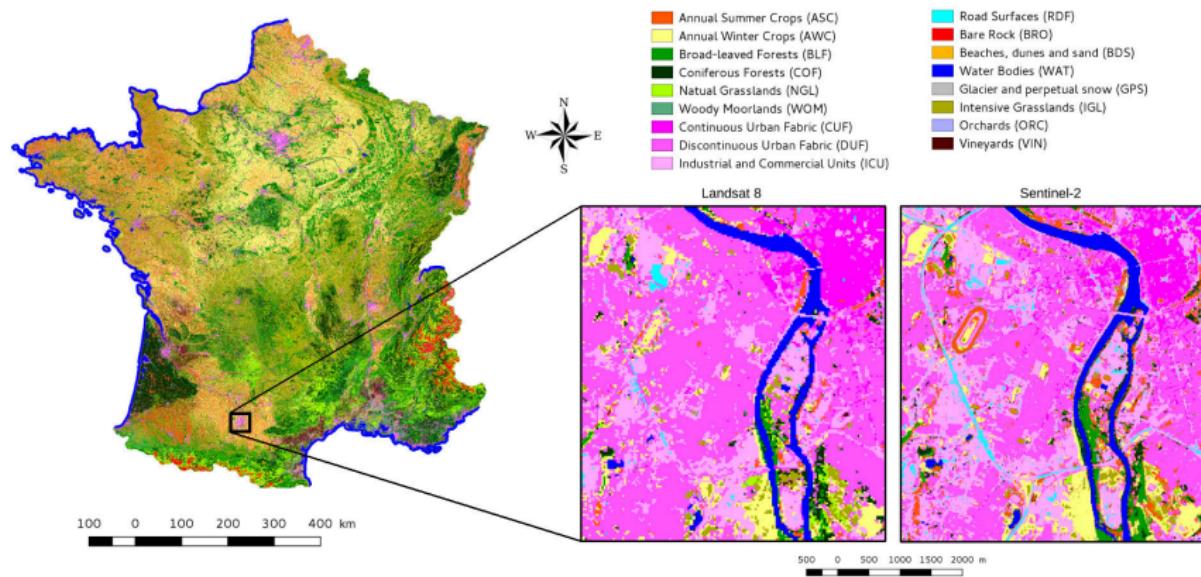
- Produit Occupation du sol (OSO) :

- ▶ Disponible à l'échelle nationale, réalisé par CESBIO / pôle THEIA
- ▶ Issu d'un traitement automatique d'images satellitaires
- ▶ Mélange entre occupation et usage
- ▶ Echelle d'usage : ~1/25.000
- ▶ Résolution thématique : 17 à 23 classes
- ▶ Résolution spatiale : 10 m (raster) et 20 m (vecteur)
- ▶ Accès libre (flux, fichier), mise à jour annuelle depuis 2016
- ▶ Produit très récent, en évolution, pas d'équivalent

3. Les sources de données

1. Données cartographiques sur l'occupation des sols

France land cover classification, from Landsat 8 to Sentinel-2.



3. Les sources de données

1. Données cartographiques sur l'occupation des sols

- Référentiel à grande échelle Occupation du sol (OCS GE) :
 - ▶ Référentiel national IGN
 - ▶ Croisement harmonisé de différentes sources et photo-interprétation
 - ▶ Echelle d'usage : 1/5.000 ou 1/10.000 (précision métrique)
 - ▶ Résolution thématique : cf. specs détaillées (couverture et usage)
 - ▶ Format vectoriel
 - ▶ Accès libre (flux, fichier), mise à jour prévue mais rythme non fixé
 - ▶ BD récente, en cours de réalisation (pas dispo partout)

3. Les sources de données

1. Données cartographiques sur l'occupation des sols

Couverture du sol										
CS1. Sans végétation				CS2. Avec végétation						
CS1.1 Surfaces anthropisées		CS1.2 Surfaces naturelles		CS2.1 Végétation ligneuse			CS2.2 Végétation non ligneuse			
CS1.1.1 Zones improductives	CS1.1.2 Zones productives	CS1.2.1 Solis nus	CS1.2.2 Surfaces d'eau	CS1.2.3 Nèvres et glaciers	CS2.1.1 Formations arborées	CS2.1.2 Formations arbustives et sous-arboiseaux	CS2.1.3 Autres formations ligneuses	CS2.2.1 Formations herbacées	CS2.2.2 Autres formations ligneuses	
Zones improductives - Sable, pierres, roches, sédiments, ...	Zones productives - (Eau continentale et marine)				(Landes basses, formations arbustives, formations ligneuses régulières, ...)		(Vignes et autres terres)			
CS1.1.1 Zones bâties	CS1.1.2 Zones non bâties (plaine, plaines, plateaux, vallées, plateaux, vallées, zones humides, zones tempérées, zones subtropicales, zones tropicales, ...)	CS1.2.1 Peuplement de feuilles	CS1.2.2 Peuplement de conifères	CS1.2.3 Zones à autres matériaux composites (boissons, ...)						

Usage du sol	US1. Production primaire	U91.1 Agriculture
	US1.2 Sylviculture	U91.3 Activités d'extraction
	US1.4 Pêche et aquaculture	U91.5 Autre
	US2.2.5 Production secondaire, tertiaire et usage résidentiel (regroupement des US2, US3 et US5 de la nomenclature nationale)	
	US4. Réseaux de transport logistiques et infrastructures	US4.1 Réseaux de transport
	US4.2 Services de logistique et de stockage	US4.3 Réseaux d'utilité publique
	US6. Autre usage	US6.1 Zones en transition
		US6.2 Zones abandonnées
		US6.3 Sans usage
		US6.4 Usage inconnu

OCS GE, IGN

3. Les sources de données

1. Données cartographiques sur l'occupation des sols

- Base OpenStreetMap (OSM) :

- ▶ Base mondiale développée de façon participative par des *contributors*
- ▶ Echelle d'usage : 1/5.000 ou 1/10.000
- ▶ Résolution thématique : très hétérogène, selon la zone
- ▶ Format vectoriel et raster (tuiles)
- ▶ Accès libre (flux, fichier), mise à jour “temps réel”
- ▶ Plus grand projet mondial de BD carto libre ; en 2018, plus de 2 milliards de tuiles OSM France envoyées pour affichage = trafic de 32 To !
- ▶ Accès également possible via data.gouv.fr

3. Les sources de données

1. Données cartographiques sur l'occupation des sols



3. Les sources de données

2. Données cartographiques sur l'usage des sols

3. Les sources de données

2. Données cartographiques sur l'usage des sols

L'utilisation du sol (**land use**) fait référence à la fonction que joue la couverture du sol pour les activités humaines. Exemples :

- surface bâtie → bâti résidentiel
- surface en eau → canal
- surface herbacée → pâturage
- végétation → culture d'été
- ...



Irrigated non-woody



Dryland agriculture



Orchards / vineyards

3. Les sources de données

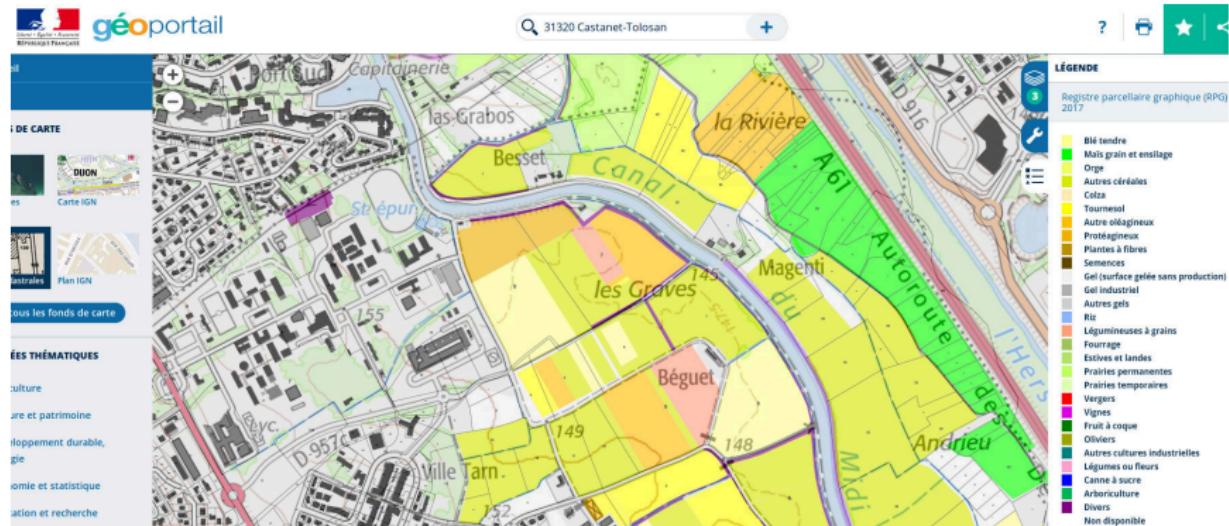
2. Données cartographiques sur l'usage des sols

- Registre Parcellaire Graphique (RPG) :

- ▶ Données du *système intégré de gestion et contrôle d'aides de la PAC*
- ▶ Parcellaire agricole (depuis 2015) et îlots culturaux (2014 et antérieurs) avec culture déclarée par l'agriculteur (écart avec la réalité possible !)
- ▶ Plusieurs niveaux d'information (1 à 4) pour préserver l'anonymat
- ▶ Niveau 1 : 28 groupes de cultures avant 2015; au-delà après 2015
- ▶ Niveau 2 : mesures agri-environnementales et climatiques (MAEC)
- ▶ Echelle d'usage : 1/10.000, format vectoriel
- ▶ Accès libre (flux, fichier), mise à jour annuelle (1 année décalage)
- ▶ Diffusion via les **Geoservices** en version *dégradée* (îlot avec culture majoritaire) et via **data.gouv.fr** (fichiers)

3. Les sources de données

2. Données cartographiques sur l'usage des sols



Registre Parcellaire Graphique (Geoportail)

3. Les sources de données

2. Données cartographiques sur l'usage des sols

- Urban Atlas :

- ▶ Harmonisé à l'échelle européenne, produite par l'EEA (programme Copernicus)
- ▶ Disponible pour les agglomérations urbaines (> 100.000 hab)
- ▶ Mélange entre occupation et usage
- ▶ Issu d'un traitement automatique d'images satellitaires + photo-interprétation
- ▶ Résolution thématique : 21 (version 2006) à 27 classes (version 2012)
- ▶ Niveau de détail : 0.25 à 1 ha
- ▶ Echelle d'usage : $\sim 1/25.000$
- ▶ Accès libre (flux et fichier), deux millésimes (2006, 2012)
- ▶ Format vectoriel

3. Les sources de données

2. Données cartographiques sur l'usage des sols



Urban Atlas

Legend Code	Urban Atlas	
	2006	2012
11000	Continuous Urban Fabric (SL>80%)	
11210	Discontinuous Dense Urban Fabric (SL..50% - 80%)	
11220	Discontinuous Medium Density Urban Fabric (SL..30% - 50%)	
11230	Discontinuous Low Density Urban Fabric (SL..10% - 30%)	
11240	Discontinuous Very Low Density Urban Fabric (SL<10%)	
11300	Isolated Structures	
12100	Industrial, commercial, public, military and private units	
12210	Fast transit roads and associated land	
12220	Other roads and associated land	
12230	Railways and associated land	
12300	Port areas	
12400	Airports	
13100	Mineral extraction and dump sites	
13300	Construction sites	
13400	Land without current use	
14100	Green urban areas	
14200	Sports and leisure facilities	
20000	Agricultural + Semi-natural areas + Wetlands	
21000	Arable land (annual crops)	
22000	Permanent crops	
23000	Pastures	
24000	Complex and mixed cultivation patterns	
25000	Orchards	
31000	Forests	
32000	Herbaceous vegetation associations	
33000	Open spaces with little or no vegetation	
40000	Wetlands	
50000	Water bodies	

3. Les sources de données

3. Données cartographiques sur les sols

3. Les sources de données

3. Données cartographiques sur les sols

- Portail **Infoterre** du BRGM :

- ▶ Référentiel national
- ▶ Grand nombre de thèmes : géologie, eaux souterraines, banque du sous-sol, cavités, alée retrait gonflement argiles...
- ▶ Notices des cartes géol disponibles
- ▶ Echelles : 1/50.000, 1/250.000 et plus
- ▶ Accès libre (flux et fichier) + interface visualisation
- ▶ Format raster et vecteur selon données

3. Les sources de données

3. Données cartographiques sur les sols



InfoTerre



Géosciences pour une Terre durable
brgm



InfoTerre, un accès à l'information scientifique et technique du BRGM

Accédez aux données du BRGM : cartes géologiques du 1/10 000 000 au 1/50 000, dossiers de la Banque de données du Sous-Sol et logs géologiques, cartes des risques naturels et industriels, données sur les eaux souterraines...

InfoTerre™ utilise exclusivement les standards de l'interopérabilité internationale édités par l'Open Geospatial Consortium (OGC). Il est cohérent avec les obligations techniques de la Directive européenne INSPIRE.

InfoTerre et son visualiseur cartographique sont des outils de découverte pour connaître et comprendre le sous-sol. InfoTerre offre aux professionnels de l'aménagement et au grand public un accès à la connaissance géologique digitale et à des données géoscientifiques interprétables?

InfoTerre vous permet de :

Composer votre carte

Actualités



05/07/2019
Science ouverte : le BRGM donne accès à la Banque du Sous-Sol (BSS)



03/07/2019
Le portail InfoTerre lance son compte Twitter



01/07/2019
Nouvelle fonctionnalité : un moteur de recherche sur la partie éditoriale d'InfoTerre

(InfoTerre)

3. Les sources de données

3. Données cartographiques sur les sols

- Le **GisSol** :

- ▶ Groupement d'intérêt scientifique qui développe un SI sur les sols de France
- ▶ Coordination de plusieurs programmes :
 - ★ **IGCS** : Inventaire, Gestion et Conservation des sols
 - ★ **RMQS** : Réseau de mesures de la qualité des sols
 - ★ **BDAT** : Base de données d'analyses des terres
 - ★ **BDETM** : Base de données d'éléments traces métalliques
- ▶ Nombreuses cartes mais données brutes pas toujours accessibles
- ▶ Mise à disposition de flux (webservices)

Contact ENSAT pour accès restreint : M. Guiresse

3. Les sources de données

3. Données cartographiques sur les sols

- L'**IGCS** (programme du GisSol):

- ▶ Fait l'inventaire spatialisé des principaux type de sols en France à trois échelles :
 - ★ Secteur de référence : < 1/50.000.
 - ★ Connaissance pédologique de France : entre 1/50.000 et 1/100.000
 - ★ Référentiel Régional Pédologique: au 1/250.000
- ▶ Alimente la BD **DoneSol** → référentiel RRP = carte + BD
- ▶ Productions recensées dans **Refersols** (recensement des études carto des sol à l'échelle nationale)
- ▶ Les cartes RRP sont numérisées (format vectoriel); aux autres échelles, formats très hétérogènes (et encore beaucoup de papier !)

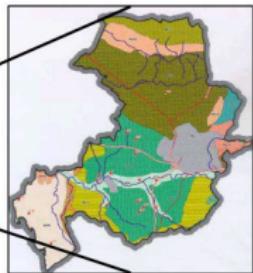
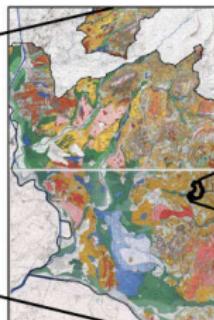
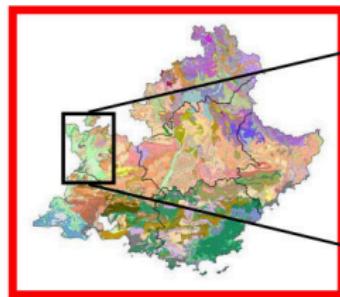
3. Les sources de données

3. Données cartographiques sur les sols

Référentiels Régionaux
Pédologiques (RRP)

Connaissance Pédologique
de la France (CPF)

Secteurs de
référence (SR)



1/250 000

Régions, Départements

1/50 000 à 1/100 000

Petites Régions

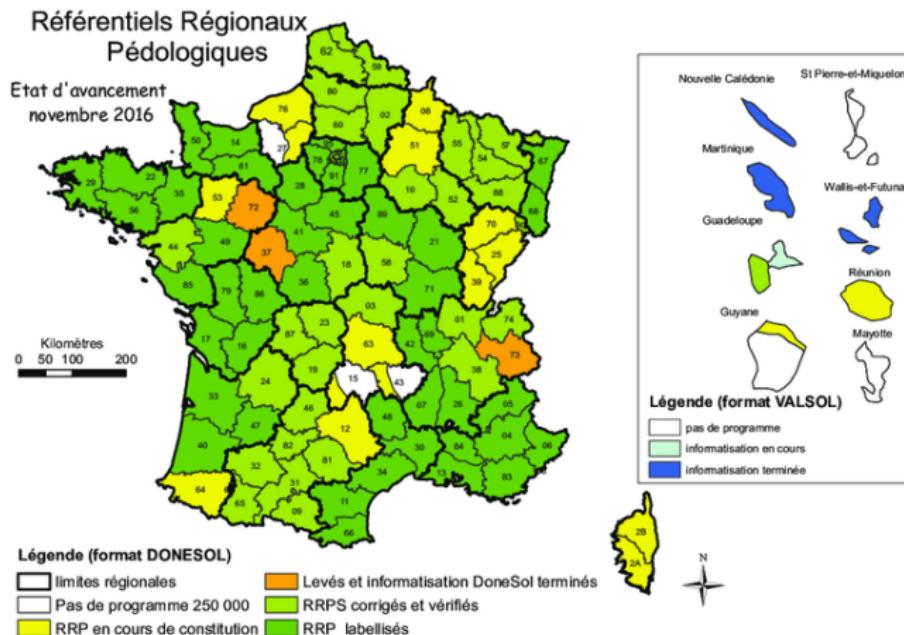
1/10 000

Parcelles

(Source : C. Jolivet, INRA)

3. Les sources de données

3. Données cartographiques sur les sols



(Source : Lemercier et al. 2017)

3. Les sources de données

3. Données cartographiques sur les sols

Accès sur demande (format vectoriel) :

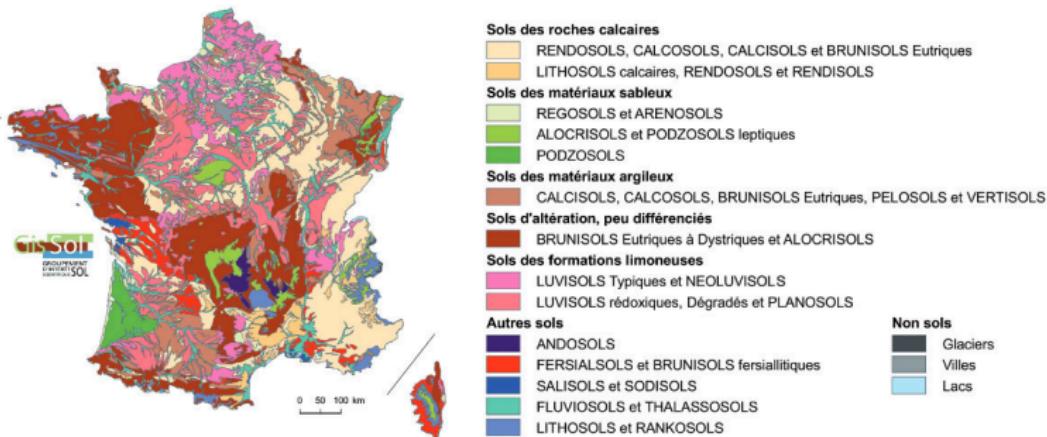
- RRP département Haute-Garonne
- RRP département Ariège
- RRP département Gers
- RRP département Lot
- RRP département Haute-Pyrénées
- RRP département Tarn-et-Garonne

3. Les sources de données

3. Données cartographiques sur les sols

- Carte nationale des sols (BDGSF) :

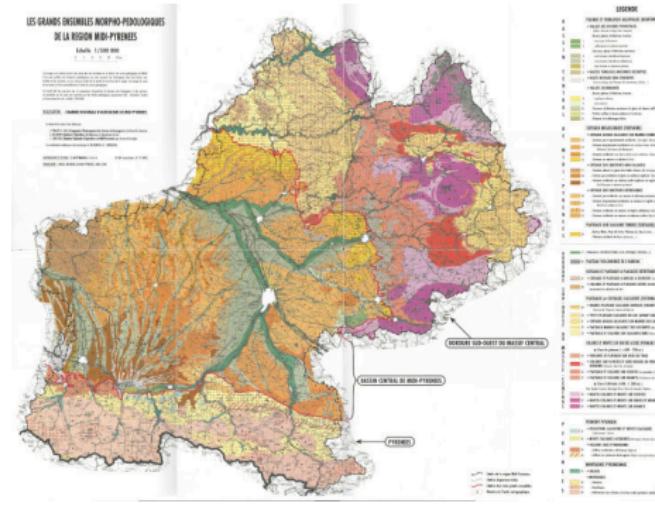
- ▶ Seule représentation complète des sols dominants au niveau national
- ▶ Echelle du 1/1.000.000 – année 1998
- ▶ Format papier (dispo BU) ou numérique (sur demande GisSol)



3. Les sources de données

3. Données cartographiques sur les sols

- Carte régionale des **pédo-paysages** de Midi-Pyrénées :
 - ▶ Echelle du 1/500.000 – année 1995
 - ▶ Format papier (dispo BU) ou numérique (sur demande à D. Sheeren)
 - ▶ Notice sur le site de la **chambre d'agriculture régionale Occitanie**



3. Les sources de données

3. Données cartographiques sur les sols

- Base mondiale SoilGrids :
 - ▶ Initiative internationale du World Soil Information ISRIC
 - ▶ Carte de propriétés physiques et chimiques des sols : texture, profondeur, éléments grossiers, pH, carbone organique...
 - ▶ Données à une résolution spatiale de 1km ou 250m
 - ▶ Adaptée à aux échelles régionales, nationales ou internationales
 - ▶ Accès libre (téléchargement)
- A mettre en relation avec the GlobalSoilMap.
 - ▶ Pour la France, voir l'article de Mulder et al. 2016

3. Les sources de données

3. Données cartographiques sur les sols

Un **bon guide** pour aller plus loin :

“Guide d'utilisation des bases de données sol pour la production de cartes thématiques” (RMT Sols & Territoires)

3. Les sources de données

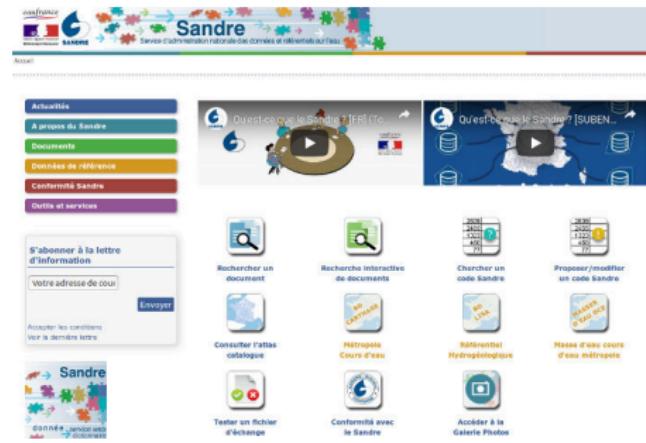
4. Données sur l'eau

3. Les sources de données

4. Données sur l'eau

- Portail Sandre :

- ▶ Référentiel national des données sur l'eau constitué par un réseau d'organismes : AFB, ONEMA, agences de l'eau, DREAL...
- ▶ Très nombreux jeux : cours d'eau, stations de mesures, taxons...
- ▶ Données alphanumériques et géographiques
- ▶ Accès libre (flux et téléchargement)



3. Les sources de données

4. Données sur l'eau

- Système d'Information EauFrance :

- ▶ Regroupe un ensemble d'informations sur l'eau, les milieux aquatiques, les usages
- ▶ Liens vers de nombreuses BD (données hydrométriques, limites des systèmes aquifères, données d'étiages...)



Liste des sites où vous trouverez des données, des cartographies et des informations dans le cadre du Système d'information sur l'eau (SIE)

3. Les sources de données

4. Données sur l'eau

- BD CarTHAge :
 - ▶ Référentiel national du réseau hydrographique français élaboré par les agences de l'eau à partir de données IGN
 - ▶ Regroupe les cours d'eau, bassins versants, points de captage...
 - ▶ Edition 2016
 - ▶ Accès libre (flux, fichiers), format vectoriel
 - ▶ A terme : remplacement prévu par la BD Topage plus exhaustif et précis

3. Les sources de données

5. Données sur les habitats

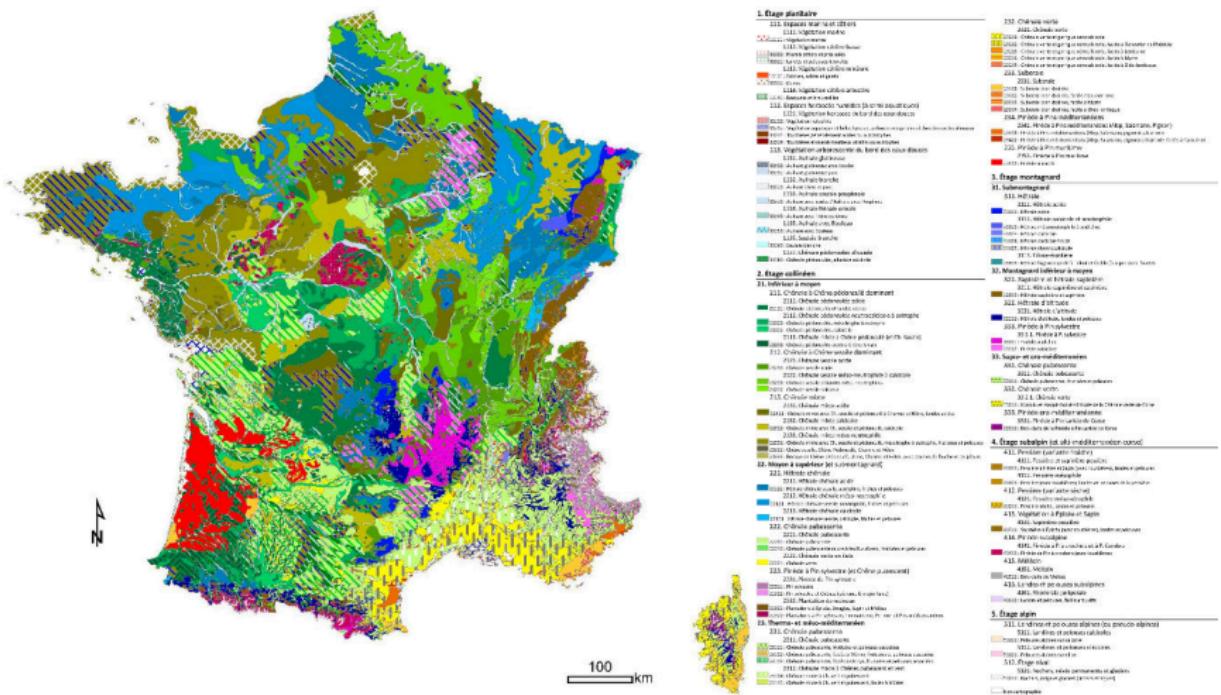
3. Les sources de données

5. Données sur les habitats

- Carte de la Végétation *potentielle* de France ou **carte de Gaussens** :
 - ▶ Elaborée par le CNRS entre 1947 et 1991 (service à Toulouse)
 - ▶ 64 feuilles représentant :
 - ★ les formations végétales dans leur dynamique (i.e. séries de végétation)
= succession de stades possibles avant d'atteindre un climax
 - ★ des informations complémentaires (cartons) moins détaillées sur les conditions édaphiques, la pluviométrie, la végétation potentielle...
 - ▶ A l'origine (version papier, dispo BU) :
 - ★ échelle du 1/200.000 pour séries et 1/1.250.000 pour cartons
 - ▶ Version numérisée (**BDGvegFR**, accès sur demande)
 - ★ échelle du 1/200.000 pour séries (scans géoréférencés)
 - ★ échelle du 1/1.000.000 de la végétation potentielle (format vectoriel harmonisé)

3. Les sources de données

5. Données sur les habitats



(Source : Leguédois et al. 2011)

3. Les sources de données

5. Données sur les habitats

- Le projet en cours CarHAB :

- ▶ Constitution d'un référentiel cartographique national des habitats terrestres naturels (horizon 2025)
- ▶ Initié par le MEEM, en cours de production avec les CBN, le MNHN...
- ▶ Approche phytosociologique sigmatiste : classification d'associations végétales traduisant de façon intégrée les facteurs environnementaux
- ▶ S'appuie notamment sur les données des conservatoires botaniques, les données IGN et la télédétection
- ▶ Echelle envisagée : 1/25.000
- ▶ Ce n'est **pas pour tout de suite !**

3. Les sources de données

5. Données sur les habitats

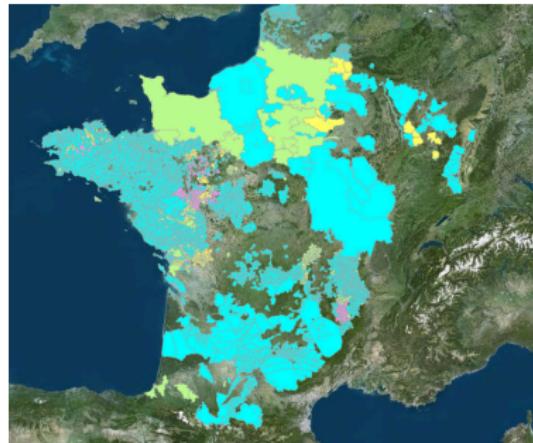
- Portail des Conservatoires Botaniques Nationaux (CBN) :
 - ▶ SI sur la flore de France (21 millions d'observations de 7687 espèces)
 - ▶ Données issues d'inventaires terrain avec nomenclature progressivement harmonisée (référentiel TAXREF)
 - ▶ Plateforme de diffusion en cours d'élaboration



3. Les sources de données

5. Données sur les habitats

- Réseau partenarial des données sur les **Zones Humides** :
 - ▶ Dresse une carte des inventaires des zones humides réalisés par les partenaires et membres du réseau
 - ▶ Accès libre (flux, téléchargement GeoJSON) + interface de visualisation
 - ▶ Voir également le site dédié **Zones Humides**



3. Les sources de données

5. Données sur les habitats

- Référentiel **HABREF** (programme INPN) :

- ▶ Référentiel des typologies nationales des habitats et de la végétation : phytosociologie, CORINE Biotope, EUNIS

The screenshot shows the homepage of the INPN HABREF website. At the top, there is a navigation bar with links for "Recherche personnalisée" (Personalized search), "English Version", and "Connectez-vous Créez un compte ?". Below the navigation bar, there is a main menu with categories: "À PROPOS", "ACTUALITÉS", "CONTEXTE", "PROGRAMMES", "DONNÉES & OUTILS", and "PARTICIPER". A secondary navigation bar below the main menu includes links for "EN SAVOIR PLUS", "ACTUALITÉS", "PRÉSENTATION", "DOCUMENTATION", "DONNÉES HABITATS", and "TÉLÉCHARGER HABREF". On the left side, there is a sidebar with a tree icon and a list of links under "Les Programmes": "Référentiel taxonomique (TAXREF)", "Base de connaissance « Statuts »", "Référentiel habitats (HABREF)" (which is expanded to show "Présentation", "Typologies d'habitats ou de végétations", "Listes d'habitats issues de directives ou de conventions", "Documentation", and "Espaces protégés"), and "Typologies phytosociologiques sigmatistes", "Typologies phytosociologiques synusiales", "Habitats marins métropolitains", "Habitats terrestres d'outre-mer", "Habitats marins d'outre-mer", and "CORINE biotopes".

3. Les sources de données

5. Données sur les habitats

- Cartes des **forêts** :

- ▶ **BDForêt** de l'IGN-IF :

- ★ Accessible par flux depuis le Geoportail
 - ★ Accès restreint au format vectoriel

- ▶ **Sylvoécorégions** de l'IGN-IF :

- ★ Intégrées au sein de grandes régions écologiques, basées sur des facteurs déterminant la production et les habitats forestiers

- ▶ **Données ONF** :

- ★ Contours des parcelles forestières (pas la composition), réserves, documents d'aménagement
 - ★ Accessible par flux (serveur **Carmen**)

3. Les sources de données

6. Données météorologiques et climatiques

3. Les sources de données

6. Données météorologiques et climatiques

- Données **Météo France** (Publithèque) :

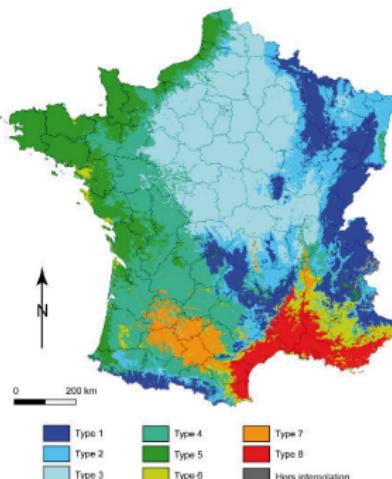
- ▶ Température, pression, précipitations, évapo-transpiration potentielle
- ▶ Différentes fréquences disponibles : quotidienne, mensuelle...
- ▶ Accès restreint (contact M. Guiresse) - quelques données déjà sur Moodle !



3. Les sources de données

6. Données météorologiques et climatiques

- Carte des **régions climatiques** de France :
 - ▶ 8 types principaux définis à partir de 14 variables relatives aux précipitations et températures basées sur une série temporelle de 30 ans



(Source : Joly et al. 2010)

3. Les sources de données

6. Données météorologiques et climatiques

• Données WorldClim :

- ▶ Ensemble de variables climatiques spatialisées à l'échelle mondiale
- ▶ T°, précipitation, vitesse du vent, radiation solaire
- ▶ Résolution spatiale : 1 km² ; synthèses mensuelles (1970-2000)
- ▶ Voir aussi <https://fr.climate-data.org/>



WorldClim Version 2

WorldClim version 2 has average monthly climate data for minimum, mean, and maximum temperature and for precipitation for 1970-2000.

You can download the variables for different spatial resolutions, from 30 seconds (~1 km²) to 10 minutes (~340 km²). Each download is a "zip" file containing 12 GeoTiff (.tif) files, one for each month of the year (January is 1; December is 12).

variable	10 minutes	5 minutes	2.5 minutes	30 seconds
minimum temperature (°C)	tmin 10m	tmin 5m	tmin 2.5m	tmin 30s
maximum temperature (°C)	tmax 10m	tmax 5m	tmax 2.5m	tmax 30s
average temperature (°C)	tavg 10m	tavg 5m	tavg 2.5m	tavg 30s
precipitation (mm)	prec 10m	prec 5m	prec 2.5m	prec 30s
solar radiation (kJ m ⁻² day ⁻¹)	sradi 10m	sradi 5m	sradi 2.5m	sradi 30s
wind speed (m s ⁻¹)	wind 10m	wind 5m	wind 2.5m	wind 30s
water vapor pressure (kPa)	vapr 10m	vapr 5m	vapr 2.5m	vapr 30s

3. Les sources de données

7. Et encore...

- Limites administratives : données IGN GeoFla sur data.gouv.fr
- Relief (MNT) : données IGN **RGE Alti** ou **BDAlti** (accès restreint) ou **SRTM** (accès libre)
- Limites des espaces protégés, ZNIEFF... : données **INPN** (libre)
- Cadastre : données IGN **BDParcellaire** ou cadastre.data.gouv.fr
- Foncier : données **Cerema** (libre)
- ...

Reportez-vous également au Catalogue sur geo.data.gouv.fr

4. Les données spatiales

4. Les données spatiales

La nature des phénomènes géographiques

4. Les données spatiales

La nature des phénomènes géographiques

Un **phénomène géographique** est un fait perceptible du monde réel qui peut être localisé, décrit ou mesuré, directement ou indirectement, par des données ou des informations (Donnay 1999, adapté).

4. Les données spatiales

La nature des phénomènes géographiques

- les **phénomènes discrets** (*object-based*) :

- ▶ discret par nature (existence discontinue dans l'espace)
- ▶ localisé, délimitation non arbitraire
- ▶ considéré homogène dans ses constituants

Ex. un bâtiment, un arbre isolé, un lac...



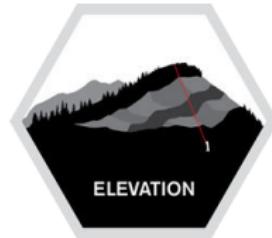
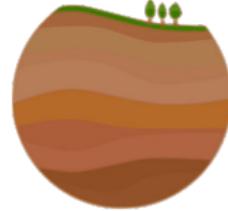
4. Les données spatiales

La nature des phénomènes géographiques

- les **phénomènes continus** (*field-based*)

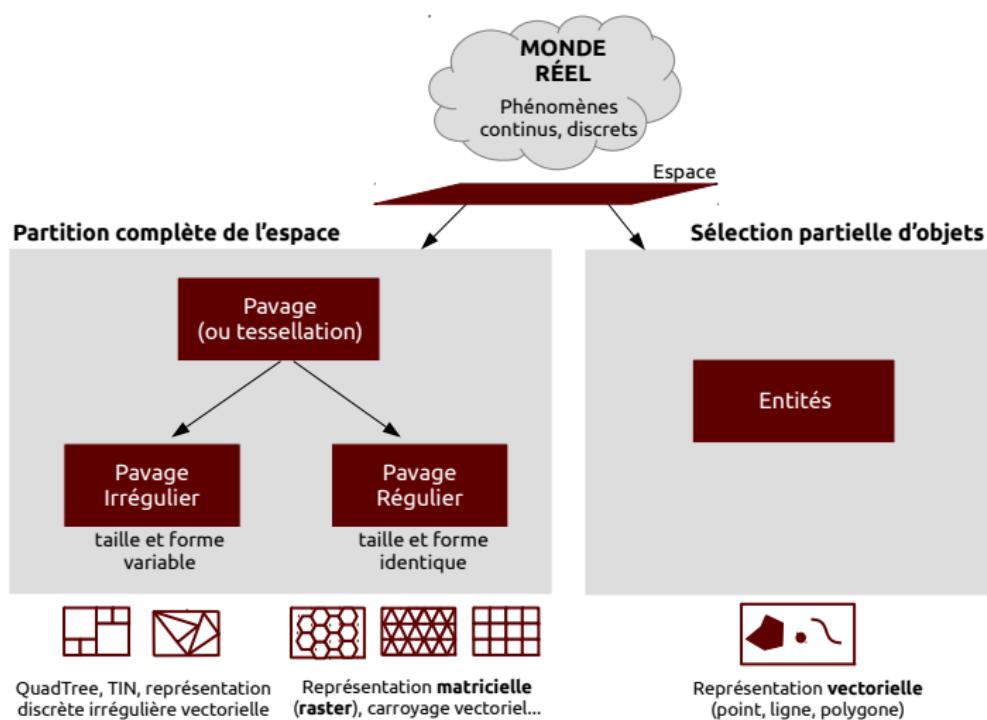
- ▶ continu par nature (existence d'une valeur en tout point de l'espace)
- ▶ variations graduelles ou abruptes
- ▶ délimitation possible par discréétisation

Ex. la pente, l'altitude, le type de sol (pédo, géol)...



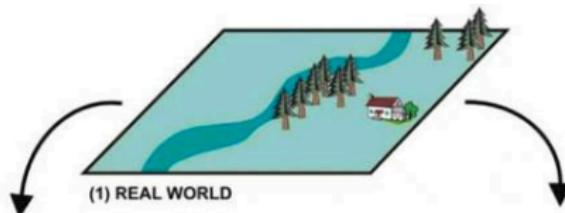
4. Les données spatiales

La représentation des phénomènes géographiques

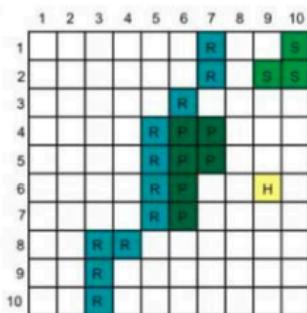


4. Les données spatiales

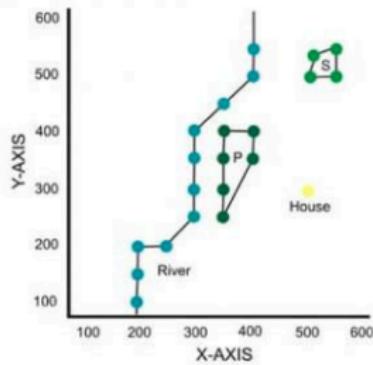
La représentation des phénomènes géographiques



(1) REAL WORLD



(2) RASTER REPRESENTATION



(3) VECTOR REPRESENTATION

4. Les données spatiales

La représentation des phénomènes géographiques

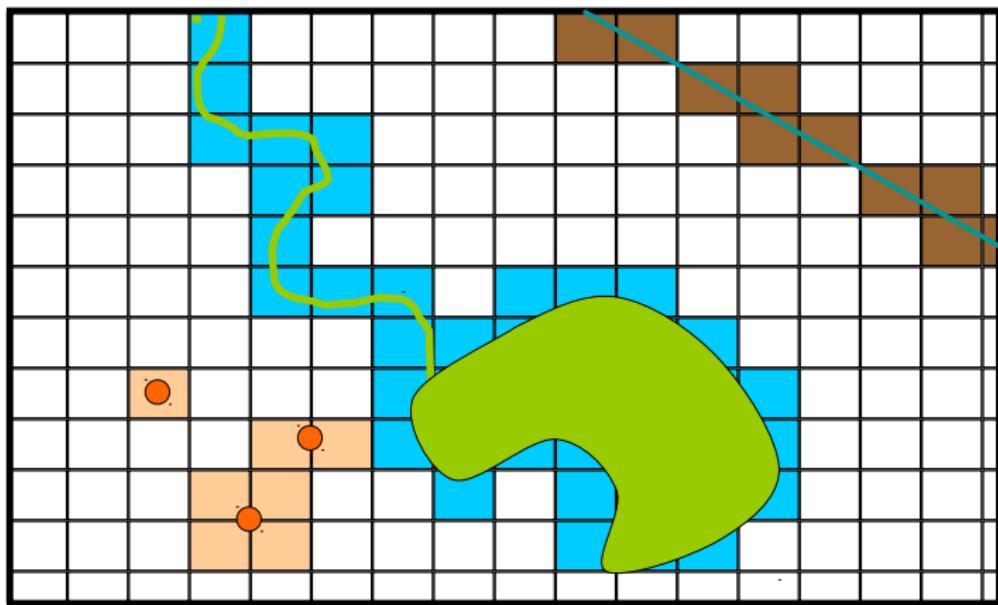
Représentation **vectorielle** (ou à *objet*)



4. Les données spatiales

La représentation des phénomènes géographiques

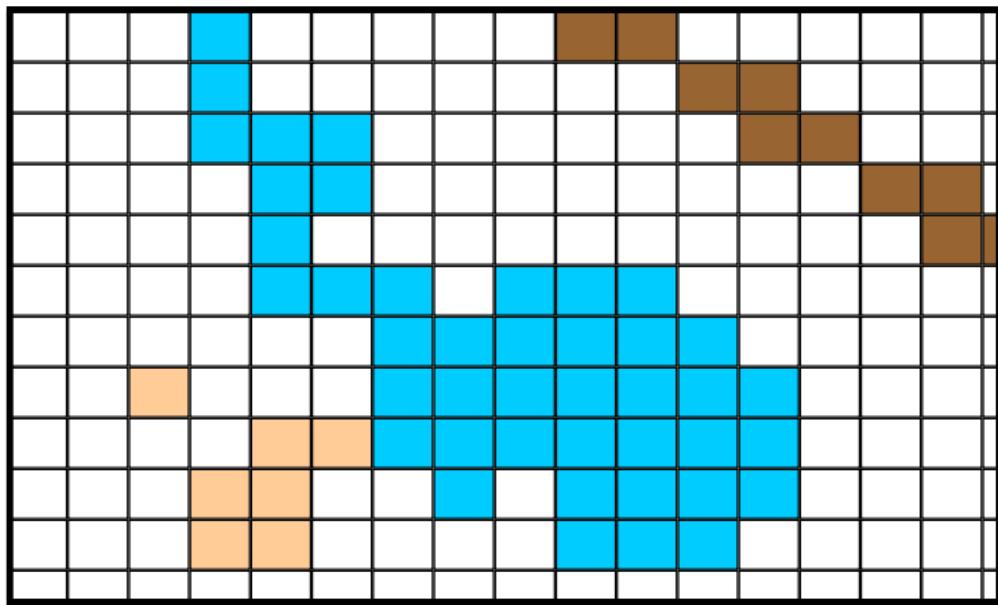
Conversion vecteur to raster



4. Les données spatiales

La représentation des phénomènes géographiques

Représentation raster



4. Les données spatiales

La représentation matricielle (**raster**, mais pas que...)

- Pavage régulier avec espace **discrétisé**

- ▶ Inclus le modèle **raster** (ou **image**) sans s'y limiter
- ▶ Unité d'observation : la **cellule** (maille ou pixel)
- ▶ Un objet est une cellule ou un agrégat de cellules partageant une même propriété
 - ★ l'objet est défini *a posteriori* (après pavage et interprétation)
- ▶ Les cellules sont porteuses d'une valeur exprimant une **propriété** thématique (ex. type sol, altitude...) ou un identifiant numérique.
 - ★ une seule valeur possible par cellule si mode raster
 - ★ toute cellule est porteuse d'une valeur
- ▶ La taille de la cellule (**Résolution spatiale**) définit le niveau de détail géométrique des objets représentés

4. Les données spatiales

La représentation matricielle (*ici, raster*)



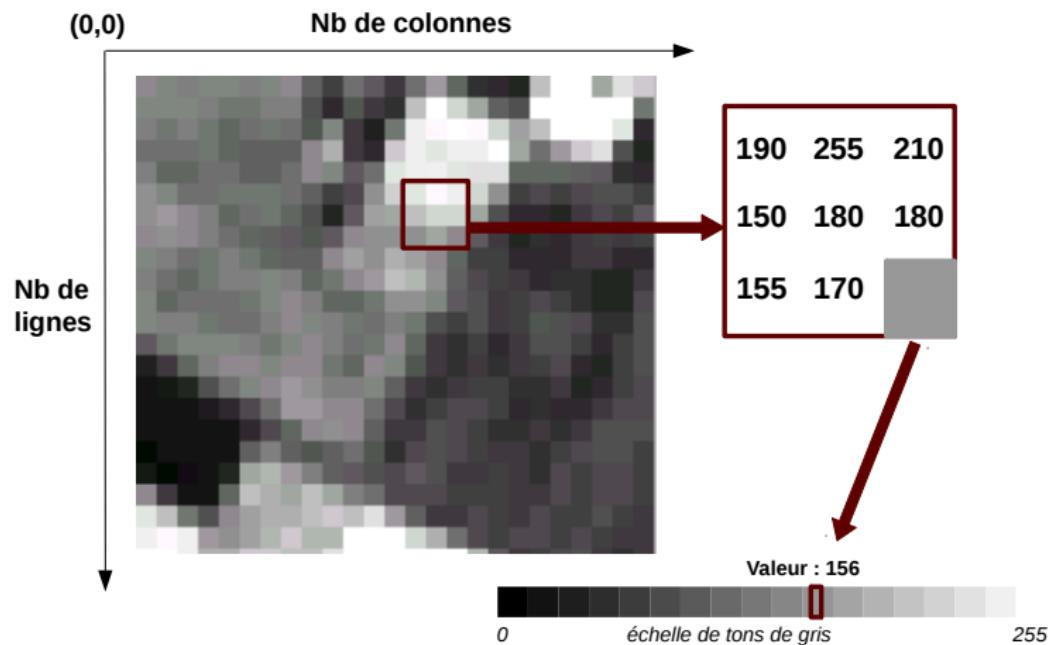
Raster non classé
(variation graduelle des valeurs RGB)



Raster classé
(variation abrupte des valeurs de classe)

4. Les données spatiales

La représentation matricielle (ici, *raster*)



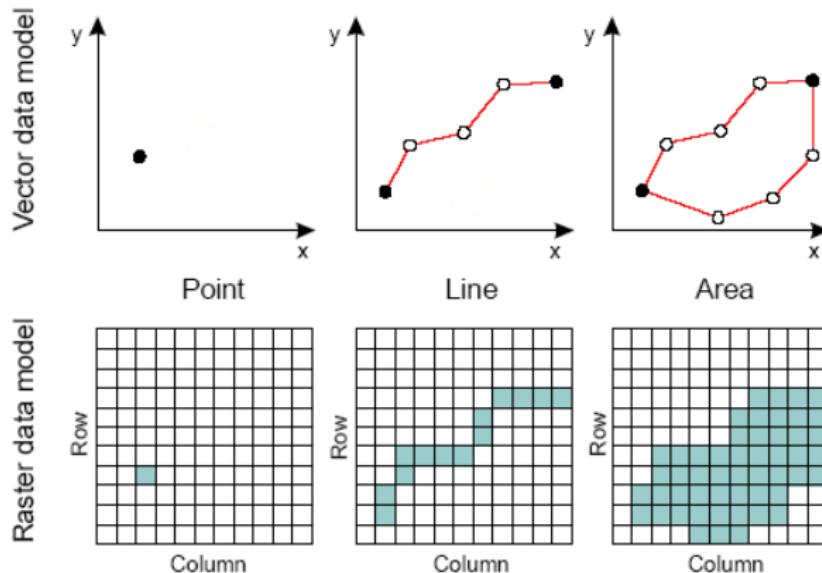
4. Les données spatiales

La représentation vectorielle ou *vecteur*

- Représentation **discrétisée** et **sélective** d'objets spatiaux
 - ▶ Unité d'observation : l'objet lui-même
 - ★ l'objet est défini *a priori* (au moment de sa sélection)
 - ▶ Un objet est représenté par une primitive géométrique :
 - ★ Point (0D) : couple de coordonnées (x, y)
 - ★ Ligne (1D) : ensemble de couples (x, y) avec noeuds et vertex
 - ★ Polygone (2D) : ensemble de couples (x, y) avec début = fin
 - ▶ Les objets sont porteurs d'un identifiant unique et d'attributs (thématiques, géométriques...)
 - ▶ Le niveau de détail des objets est défini par le degré de *généralisation cartographique* (niveau de simplification de la représentation)

4. Les données spatiales

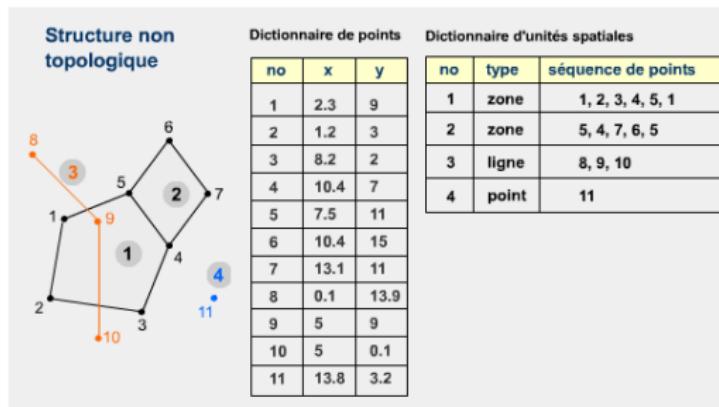
La représentation vectorielle ou *vecteur*



4. Les données spatiales

La représentation vectorielle ou *vecteur*

- Il existe différentes **structures de donnée** pour stocker les objets :
 - structure ***non topologique*** (spaghetti) : la plus simple
 - les relations spatiales entre les objets sont ignorées (implicites)
 - rapide à saisir mais redondance et artefacts (chevauchement, trous...)

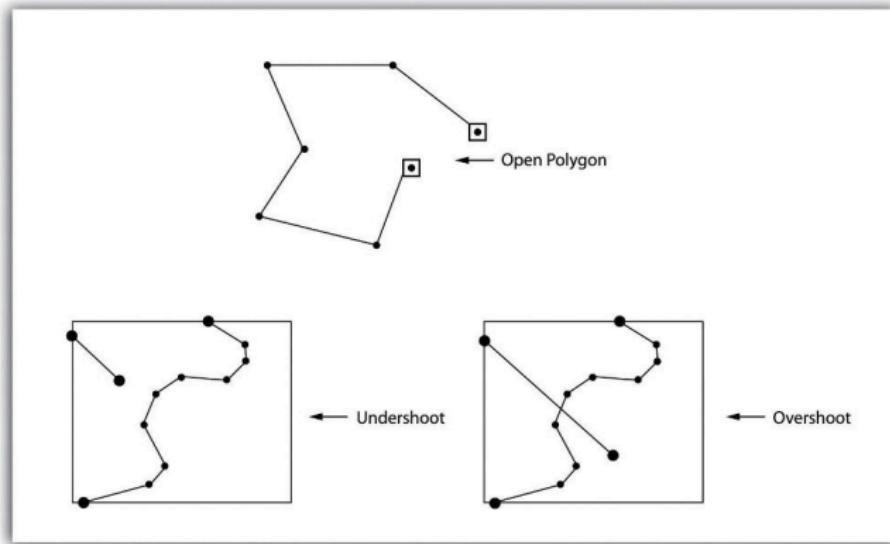


(Source : Gitta)

4. Les données spatiales

La représentation vectorielle ou *vecteur*

Erreurs de saisie fréquentes avec structure non topologique :



(Source : Saylor academy)

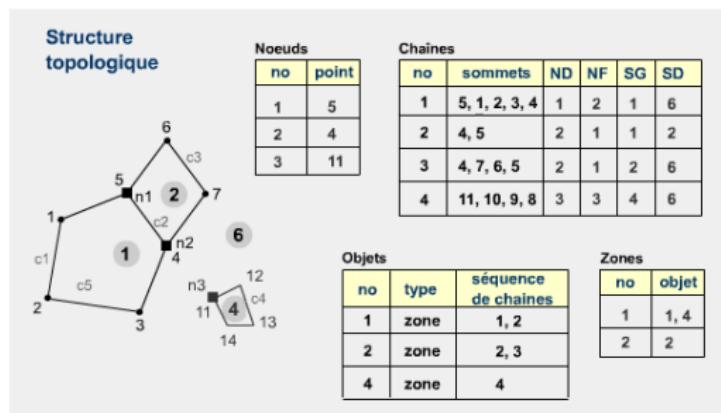
4. Les données spatiales

La représentation vectorielle ou *vecteur*

- Il existe différentes **structures de donnée** pour stocker les objets :

- structure ***topologique*** :

- ★ relations de **connectivité** (arc-noeud)
- ★ relations de **voisinage** (polygones-polygones)
- ★ relations de **coincidence** (partage de géométrie)



(Source : Gitta)

4. Les données spatiales

Comparatif des représentations *vecteur* et *raster*

Vecteur

Localisation et géométrie liées aux (x, y)

Volume réduit des données (selon nombre d'objets)

Représente une partie de l'espace (objets saisis)

Enregistrement de nombreux attributs

Structure de données plus ou moins complexe

Description topologique implicite ou explicite

Gère les réseaux

Analyse spatiale via calculs complexes (algorithmie géométrique)

Qualité graphique dépendante du degré de généralisation

Mise à jour facile (édition)

Raster

Localisation et géométrie liées aux pixels

Volume important (selon nombre de pixels)

Représente la totalité de l'espace

Enregistrement d'une valeur par pixel

Structure de données simple

Aucune description topologique explicite

Inadapté aux réseaux

Analyse spatiale via traitements simples (algèbre de carte)

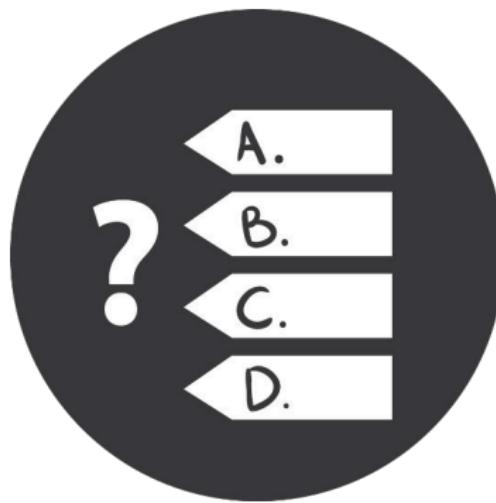
Qualité graphique dépendante de la résolution spatiale

Mise à jour plus complexe (changement valeur pixel)

Conversion possible : raster \longleftrightarrow vecteur

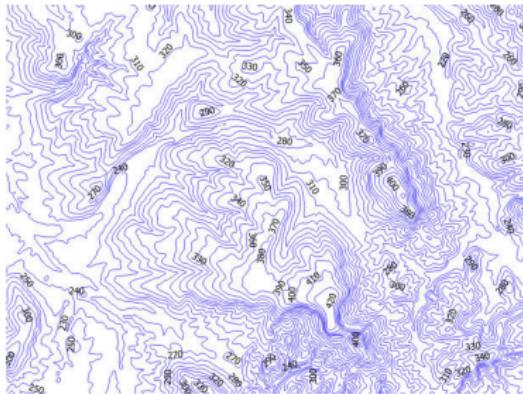
4. Les données spatiales

Quizz !



4. Les données spatiales

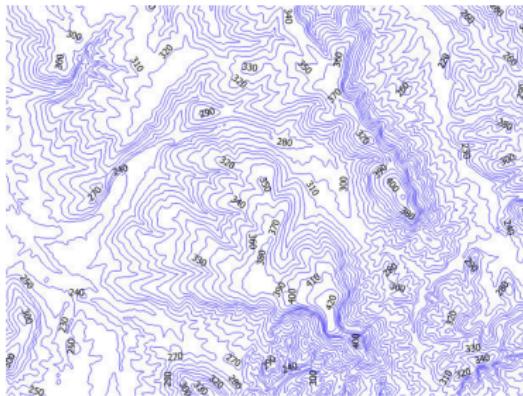
Q1. De quelle nature est le phénomène géographique ? Quel est son mode de représentation ?



- ① phénomène discret et représentation vecteur
- ② phénomène discret et représentation raster
- ③ phénomène continu et représentation vecteur

4. Les données spatiales

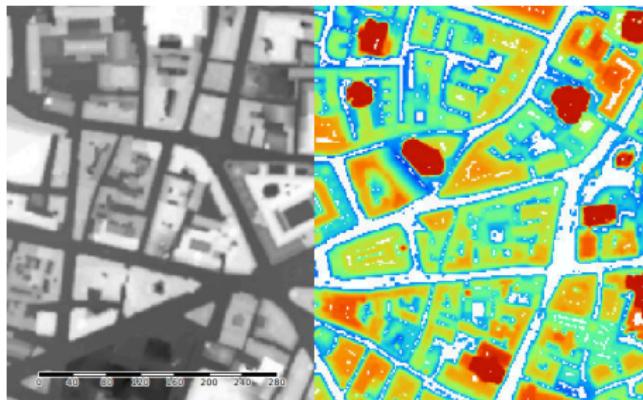
Q1. De quelle nature est le phénomène géographique ? Quel est son mode de représentation ?



- ① phénomène discret et représentation vecteur
- ② phénomène discret et représentation raster
- ③ [X] phénomène continu et représentation vecteur =*courbes de niveau*

4. Les données spatiales

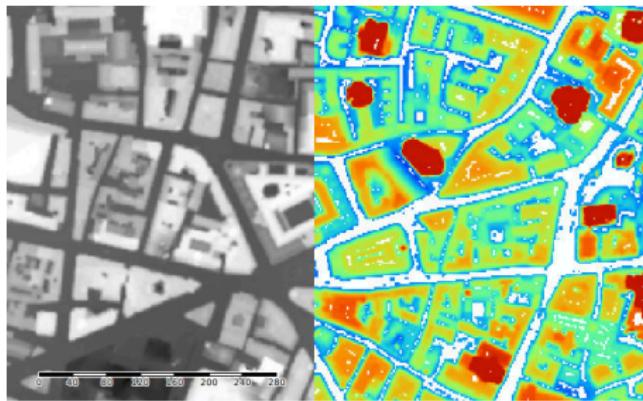
Q2. De quelle nature est le phénomène géographique ? Quel est son mode de représentation ?



- ① phénomène continu et représentation raster
- ② phénomène discret et représentation raster
- ③ phénomène continu et représentation vecteur

4. Les données spatiales

Q2. De quelle nature est le phénomène géographique ? Quel est son mode de représentation ?



- ① [X] phénomène continu et représentation raster =MNS
- ② phénomène discret et représentation raster
- ③ phénomène continu et représentation vecteur

4. Les données spatiales

Q3. Si j'enregistre des points avec un GPS, quel sera le mode de représentation des données une fois importées sous SIG ?

- ① vecteur
- ② raster

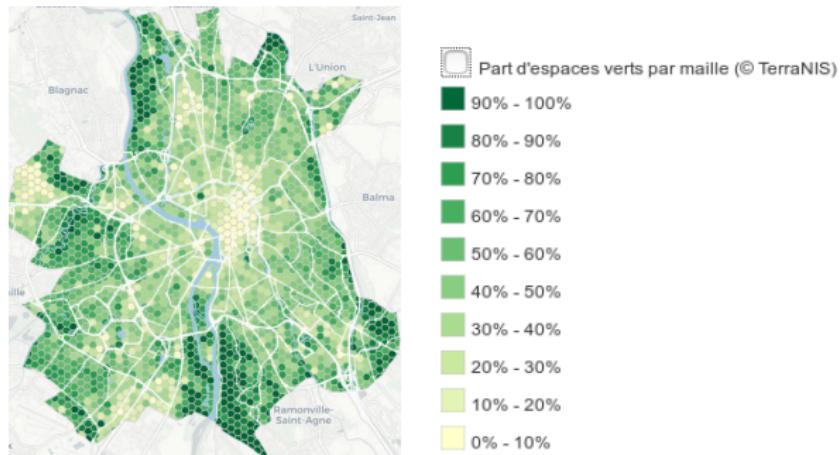
4. Les données spatiales

Q3. Si j'enregistre des points avec un GPS, quel sera le mode de représentation des données une fois importées sous SIG ?

- ① [X] vecteur
- ② raster

4. Les données spatiales

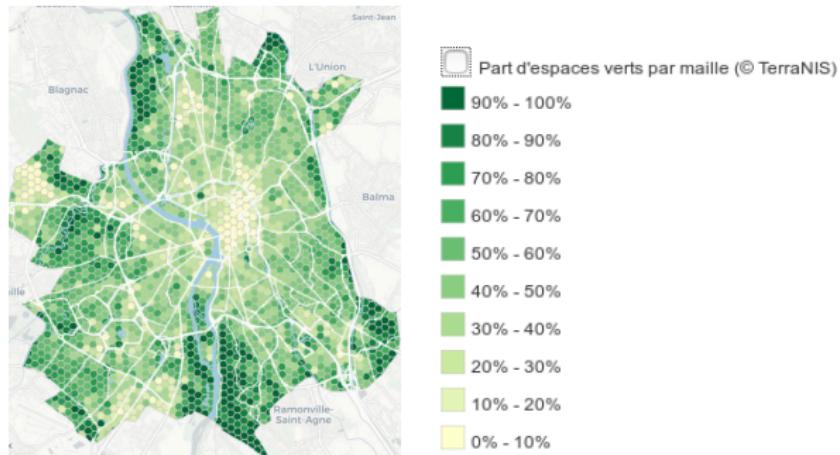
Q4. De quelle nature est le phénomène géographique ? Quel est son mode de représentation ?



- ① phénomène continu et représentation raster
- ② phénomène discret et représentation raster
- ③ phénomène discret et représentation vecteur

4. Les données spatiales

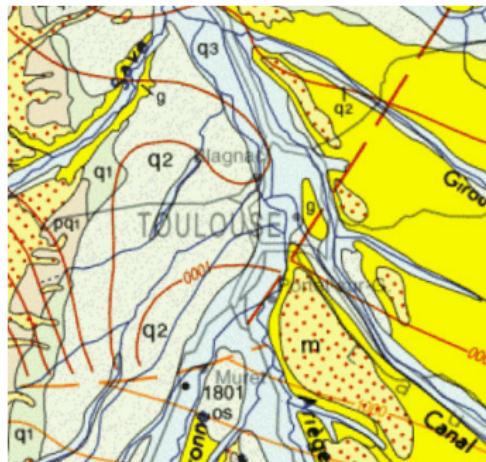
Q4. De quelle nature est le phénomène géographique ? Quel est son mode de représentation ?



- ① phénomène continu et représentation raster
- ② phénomène discret et représentation raster
- ③ [X] phénomène discret et représentation vecteur **par pavage régulier !**

4. Les données spatiales

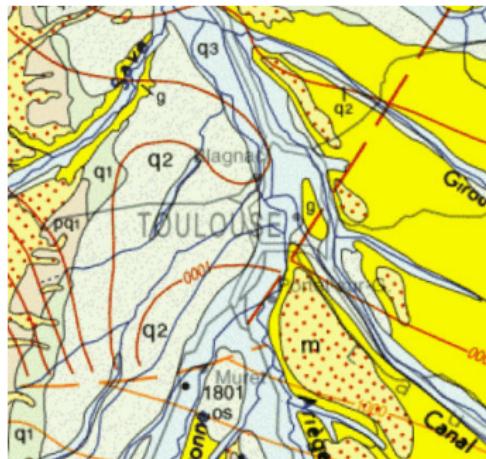
Q5. De quelle nature est le phénomène géographique ? Quel est son mode de représentation ?



- ① phénomène discret et représentation raster
- ② phénomène continu et représentation vecteur
- ③ phénomène continu et représentation raster

4. Les données spatiales

Q5. De quelle nature est le phénomène géographique ? Quel est son mode de représentation ?



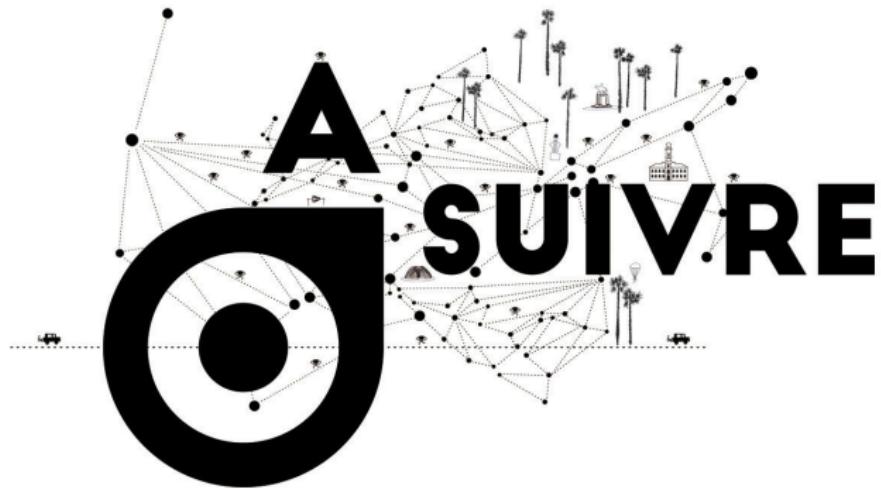
- ① phénomène discret et représentation raster
- ② [X] phénomène continu et représentation vecteur **par pavage irrégulier !**
- ③ phénomène continu et représentation raster

4. Les données spatiales

Attention

- La **représentation matricielle** n'est **pas réservée au mode raster**. Un pavage régulier (et irrégulier) de l'espace peut être réalisé en mode vecteur.
- Le **mode raster** n'est **pas uniquement réservé à la représentation de phénomènes continus**. Il peut aussi être utilisé pour représenter des objets (avec une partition complète de l'espace)
- Le **mode vecteur** n'est **pas strictement réservé à la représentation de phénomènes discrets** (à objets). Il peut aussi être utilisé pour représenter des phénomènes continus.

Donc : bien distinguer la nature du phénomène et sa représentation...



(Source : Asso Omnibus)

Ce document est mis à disposition selon les termes de la licence Creative Commons "CC-BY-SA"

