

Science des données I



First class meeting

Philippe Grosjean & Guyliann Engels

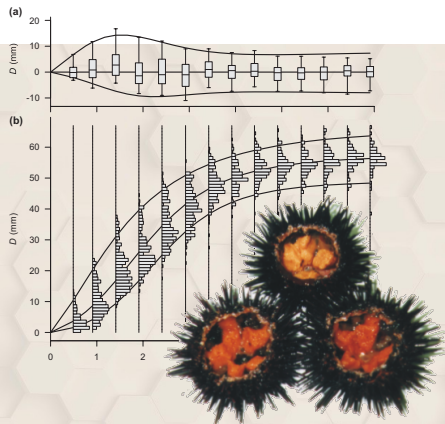
Université de Mons, Belgique
Service d'Écologie numérique



<https://wp.sciviews.org>
sdd@sciviews.org

Qui sommes nous ?

Prof. Philippe Grosjean



- Bioingénieur + thèse de doctorat en biologie marine (croissance d'oursins)
- Capacités supplémentaires développées en science des données durant des post-docs et via de la consultance pendant quatre ans partout en Europe
- Service EcoNum créé en 2004 à l'Université de Mons
- Intéressé par des travaux interdisciplinaires : biologie, chimie, modélisation, statistiques, informatique

Guyliann Engels

- Master en Biologie des Organismes et Écologie à l'UMONS.
- Mémoire effectué dans le service EcoNum sur l'écophysiologie et l'écotoxicologie de la posidonie (*Posidonia oceanica*, une plante marine) en Méditerranée.
- Assistant dans le service EcoNum de 2017-2023.
- Adjoint pédagogique en biologie à l'UMONS depuis septembre 2023.



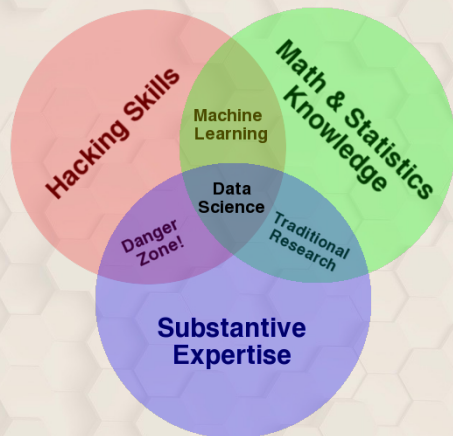
Science des données

Science des données : une approche pragmatique

A data scientist is a
statistician who is useful.
— *Hadley Wickham*

Science des données : à l'interface entre plusieurs disciplines

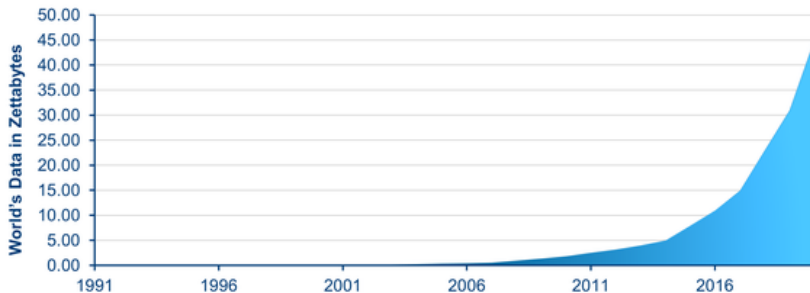
- La Science des Données, c'est la discipline qui s'intéresse à l'analyse de données sous toutes ses formes
- Très large et interdisciplinaire :
 - (Bio)statistiques et visualisation
 - Utilisation d'outils informatiques
 - Expertise dans le domaine (biologie)
- Il faut maîtriser simultanément les trois domaines pour être un scientifique des données.



Pourquoi la science des données ?

- Discipline à la fois ancienne et récente
 - Évolution des statistiques, avec ses prémices dans les années 1960 (John Tukey).
 - Émerge comme science à part : 2001 William S. Cleveland, “Data Science : An Action Plan for Expanding the Technical Area of Field of Statistics”.
 - Le terme Data Scientist n’est d’usage courant que depuis 2008.
- Besoin issu de la quantité de données disponibles (1 zettabyte = 1 milliard de terabytes = 1 000 000 000 000 000 000 octets).

Data growth



La science des données biologiques

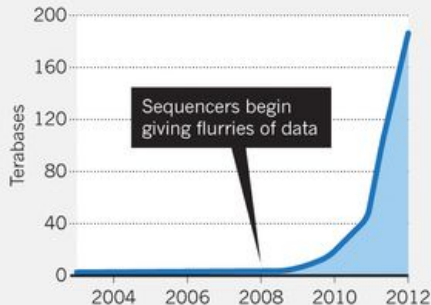
La biologie n'échappe pas au besoin d'analyser des (gros) jeux de données :

- Génétique, bases immenses
- Biodiversité animale et végétale
- Études écologiques avec images satellites, capteurs haute vitesse
- Littérature scientifique
- etc.

Un biologiste analyse des données pratiquement quotidiennement sous une forme ou l'autre !

DATA EXPLOSION

The amount of genetic sequencing data stored at the European Bioinformatics Institute takes less than a year to double in size.



Objectifs

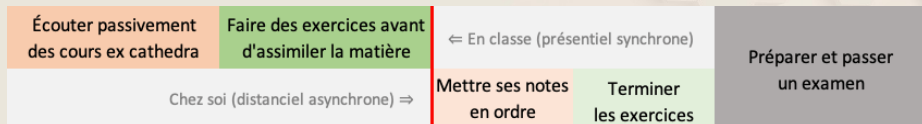
Nos objectifs principaux durant votre formation en **science des données biologiques** est de vous rendre capable de :

- réaliser des analyses biologiques usuelles,
- présenter clairement vos résultats de manière reproductible avec des outils informatiques et statistiques professionnels
- développer votre esprit critique

Vous pouvez retrouver via la fiche du cours le détails de tous les objectifs.

Comment se déroule un cours ?

Cours ex cathedra + Travaux pratiques



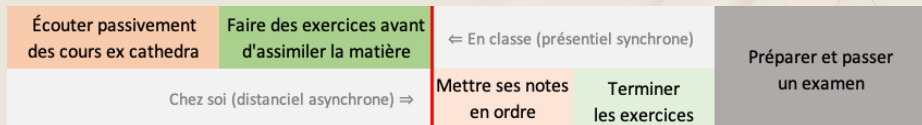
- Le réel apprentissage se déroule après les séances de cours et d'exercices
- Un examen est nécessaire pour vérifier vos acquis

D'après la science lors d'un cours ex cathedra, vous n'apprenez rien

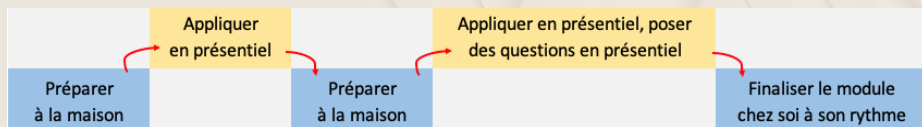
-> Quelle perte de temps :(

Pédagogie active et classe inversée

Cours classique ex cathedra + séances d'exercices



Approche en classe inversée



- Aucune séance en présentiel sans préparation
- Chaque heure de travail pleinement consacrée à l'apprentissage
- Vous êtes actifs tout le temps et vous gérez à votre rythme
- Pas besoin d'un examen à la fin : travail évalué durant l'année

C'est quoi la classe inversée ?

Vous avez une minute ?

Pour comprendre
La classe inversée

0:03 / 1:18



(lien vers la vidéo)

UMONS

Classe inversée et pédagogie active

Notre approche : pédagogie active en classe inversée (vous apprenez d'abord à la maison, nous appliquons ensuite en présentiel).

I hear and I forget.

I see and I remember.

I do and I understand.

— Confucius

C'est quoi la pédagogie active ?

Les
pédagogies
actives
pourquoi ne
pas essayer?



0:00 / 3:05



(lien vers la vidéo)

UMONS

Et moi, je fais quoi dans tout cela ?



- Vous êtes acteur de votre apprentissage, les enseignants sont des facilitateurs (plus en retrait par rapport à l'approche classique).
- Plus de séparation entre cours théorique et exercices ; vos échanges avec le professeur et l'assistant sont similaires.
- Vous posez les questions, et vos enseignants vous répondent individuellement.

Organisation du cours

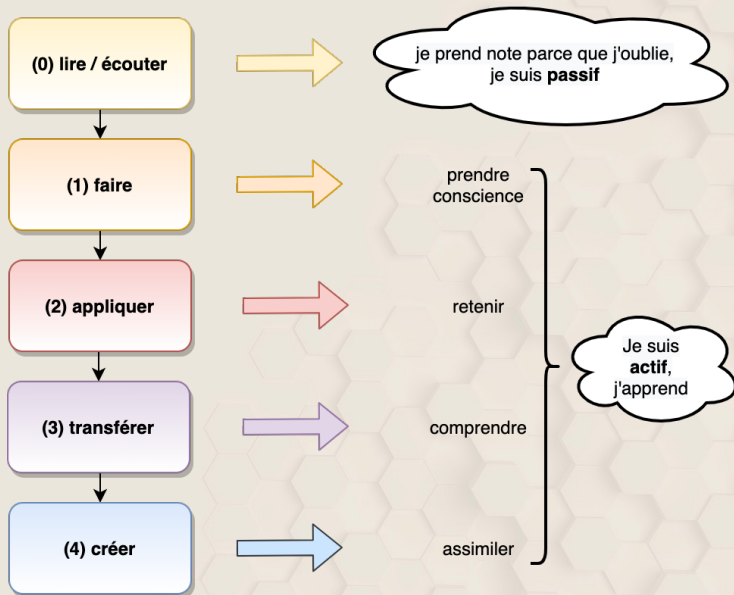
ECTS

European Credits Transfer System, créé en 1988 de manière standardisée par la Commission Européenne comme correspondant à une charge de travail totale pour l'étudiant. Ce cours comprend 12 modules pour 6 ECTS (0.5 ECTS/module)

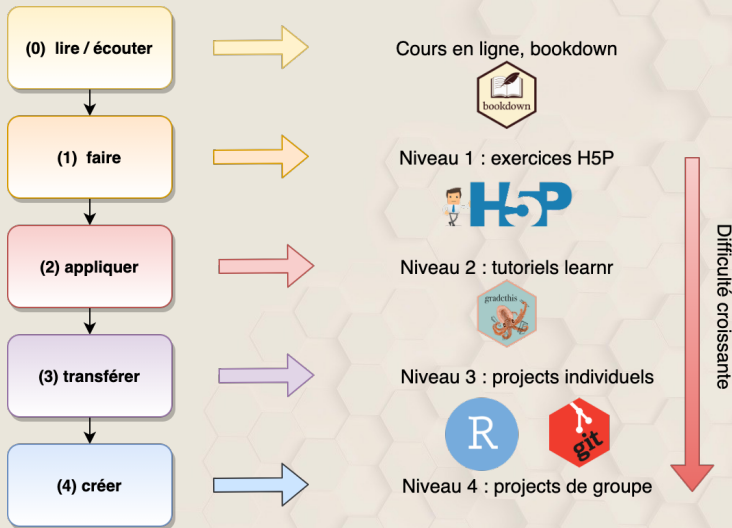


Apprentissage en quatre niveaux

Niveaux d'exercice = Apprendre, niveaux 1 à 4



Quatre niveaux d'exercices



Évaluation continue & planning

- Le planning des séances est synthétisé dans le cours en ligne
- Vous pouvez suivre votre progression sur Moodle ou à la fin de chaque module du cours en ligne
- Vous aurez des projets (analyse de données biologiques) en séance
- Vous aurez deux interrogations écrites par AA et un challenge
- Note de l'AA (voir plan de cours) :
 - 25% des points pour les exercices (5% H5P-Shiny-Learnr, 20% challenge)
 - 25% des points pour les projets (5% individuels, 20% de groupe)
 - 25% première interrogation
 - 25% seconde interrogation

Avez-vous des questions ?



Ressources utiles :

- Site web du cours : <http://bds.sciviews.org/>
- Cette présentation : https://github.com/BioDataScience-Course/sdd_lessons/tree/2023-2024/A00/presentations