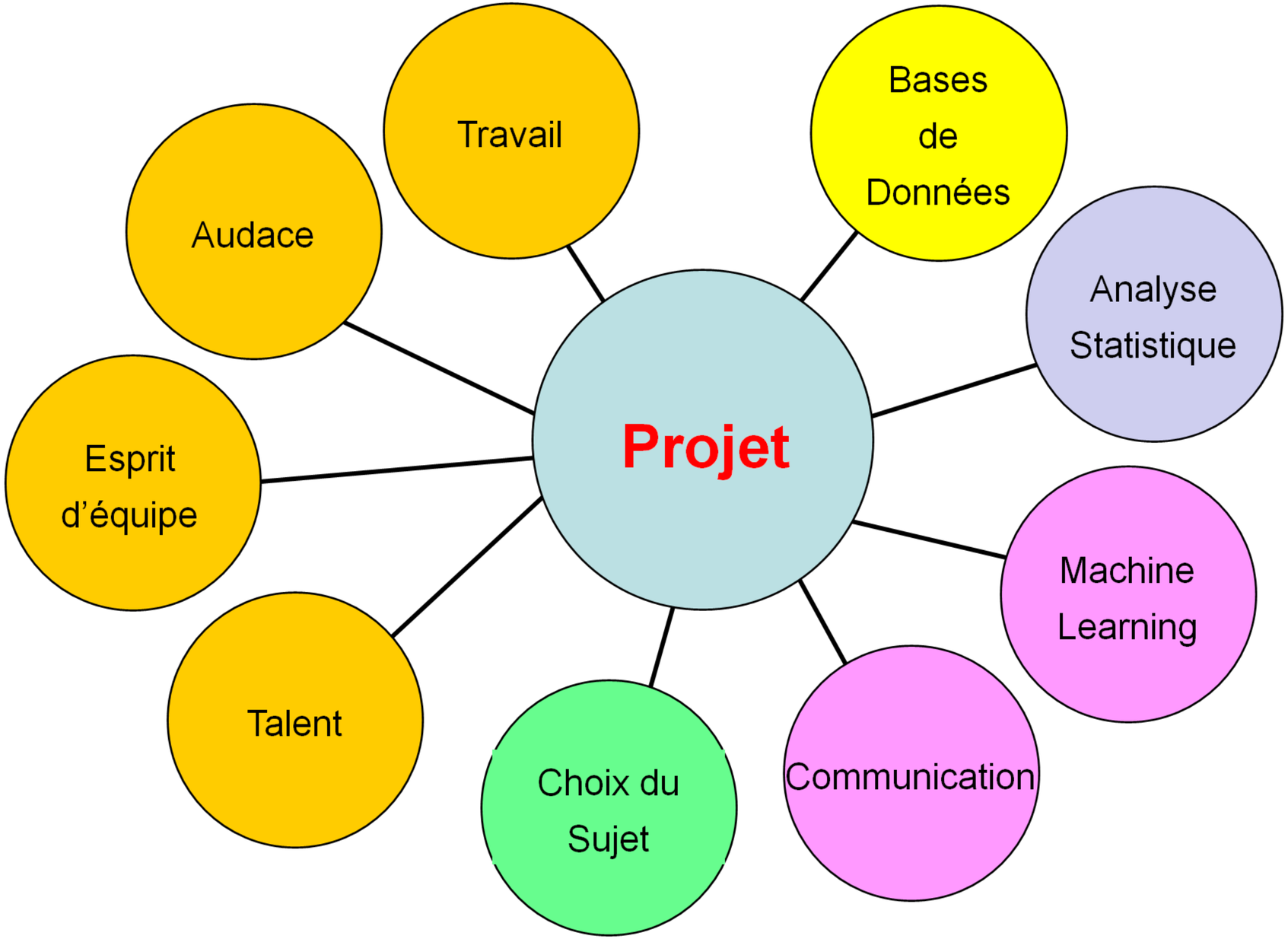
Projet Personnel d’Etudes et d’Insertion (PPEI)

⇨ une formation obligatoire à la conduite de projet

Les composantes de votre projet de Data Sciences



+ organisation

Les connaissances et les composantes qui vous permettront de concrétiser votre projet

Du jeu de données (data set) aux résultats obtenus

Le cheminement des données

Acquisition

(acquérir les données de façon intelligente, plan d'expérience, campagne de collecte, fiabilité des data)

Filtrer, organiser

Stockage

(organisation, formatage, labels, rôle central des **bases de données**

Nettoyer, décrire, regrouper

Traitement - Analyse

(codage, recodage, scoring, **analyse statistique**, modélisation, prédictions, **machine learning**, …)

Faire parler les données selon un objectif, un but précis

Communication

(conclusions, nouvelles connaissances, décisions, ordres à machines, prévention, ….)

Architecture des projets

Les Projets

Le but des **UEs «Projet Personnel d'Etudes et d'Insertion »** est de vous apprendre à construire et concrétiser un **projet en groupe** dans une thématique donnée (dans votre domaine scientifique de prédilection).

Dans cette **UE « Data Sciences »** vous réaliserez vos projets par groupes de 3 ou 4 (4 étant l'idéal). Attention, cela demande un réel travail partagé, une bonne **coordination**, un bon **esprit d 'équipe** et une bonne **organisation**!

Vous avez jusqu'à **mi mai** (date à définir en fonction de vos examens) pour **réaliser et présenter oralement votre projet.**

Cela va aller plus vite que vous ne l'imaginez.

Les Projets

Vous en sortirez **enrichis**, formés à la conduite de projet et pourrez valoriser cette expérience sur **votre CV** [YouTube ou un Podcast dédié (pour les **3 meilleurs** projets)]. Je réfléchis à un **prix pour doter les 3 meilleurs projets**.

Proposition de création de startup (en accompagnement total) pour les étudiants qui seront intéressés

Une courte **vidéo** explicative sera bientôt réalisée à propos de vos projets.

Les Projets, un travail d'équipe

Projets : les 4 compétences

> Base de données,

> Analyse statistique (avec R),

> Machine learning (en programmation python ou R)

> Communication

Et

> le/la chef(fe) de projet

(crédité de 1,5 points de bonus)

mais attention il s'agit d'une charge et d'une responsabilité importantes

Les rôles utiles des postit, du tableau blanc (paper board) et du story telling!

Les équipes-projets

Votre **équipe-projet** contiendra **4 membres**, chacun des membres se spécialisant plus sur une des 4 expertises/compétences (**Bases** de Données et SQL, **Analyse Statistique** et R, **Machine Learning** et Python, **Communication**, design & style écrit et oral).

L'un(e) de ces 4 membres assurera la fonction de **chef(fe) de projet** (créditée de points bonus, conditionnés au résultats obtenus).

Bref, vous réaliserez un **vrai travail d'équipe** comme il est de coutume de le faire en *data sciences* en réunissant des **compétences, des aspirations et des talents complémentaires qui servent un projet commun**, le votre.

Les équipes-projets

La thématique sera empruntée, au choix, **dans n'importe quel domaine de la biologie** qui intéresse les membres de l'équipe, projet biologique de "biodatasciences".

Votre **enseignant** sera à la fois **animateur**, supporter et **membre mobile de chaque équipe**. Il est là pour vous remonter le moral au cas où.

Vous aurez tous les mêmes enseignements mais vous pourrez décider de vous orienter vers un des aspects qui vous séduit le plus dans les data sciences. On choisira (peut-être) des jetons couleurs virtuels (je vous expliquerai) pour déterminer ce choix.

Synthèse

Projet

Base de données (ACCESS)

- ⊕ Analyse statistique (Rcmdr) - copies d'écran
- ⊕ Jeu de données (EXCEL) utilisé
- ⊕ Progamme python (Jupyter Notebook)
- ⊕ Rapport écrit (12 à 15 pages) - pdf
- ⊕ diaporama Powerpoint (sonorisé) et/ou Fichier vidéo mp4 maxi 10'

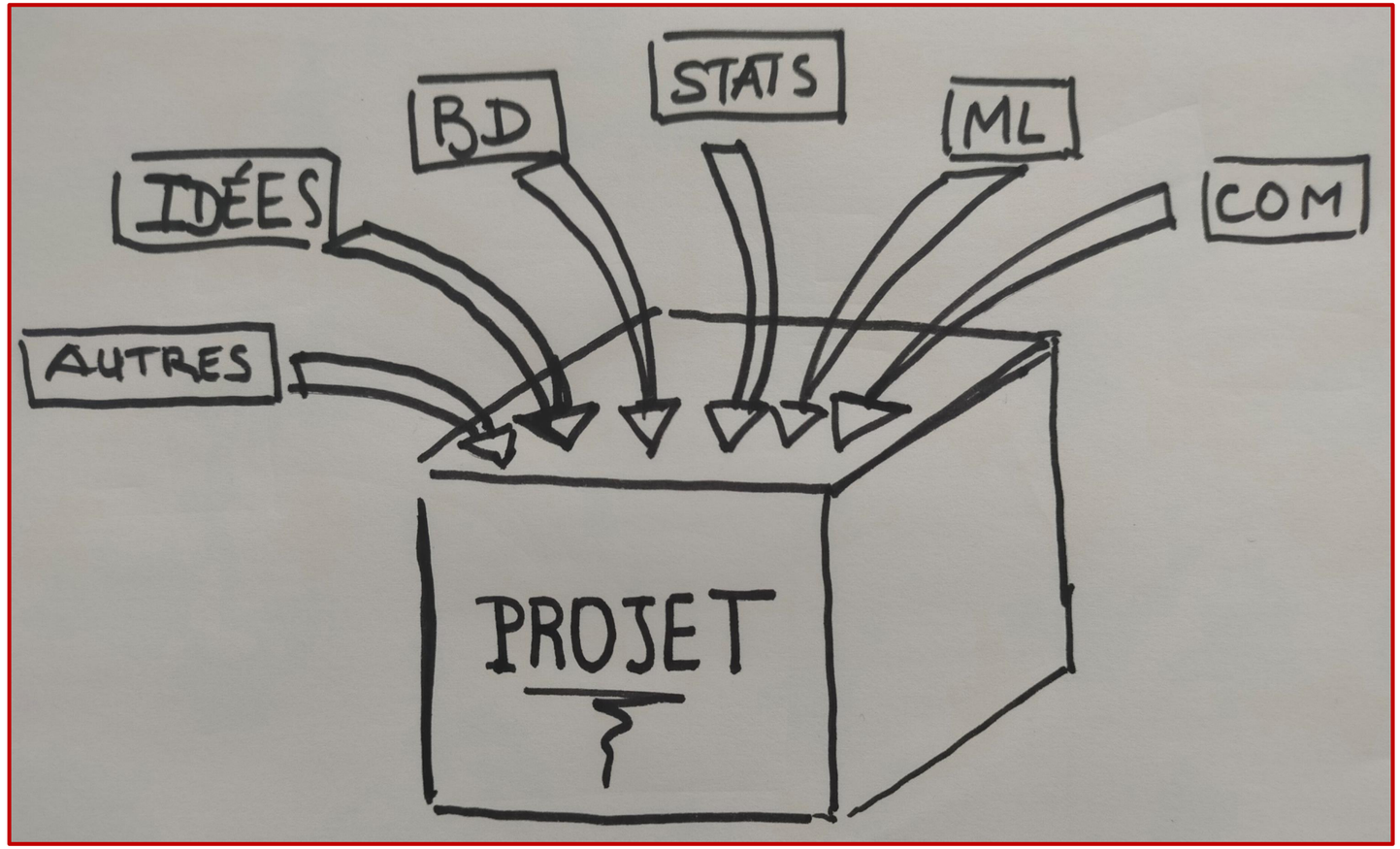
documents rangés dans des répertoires dédiés
et l'ensemble mis dans un répertoire racine portant le nom :
" noms des coequipiers - projet2022 "

↓

Fichier compressé Zip " noms coequipiers - projet2022.zip "

↓

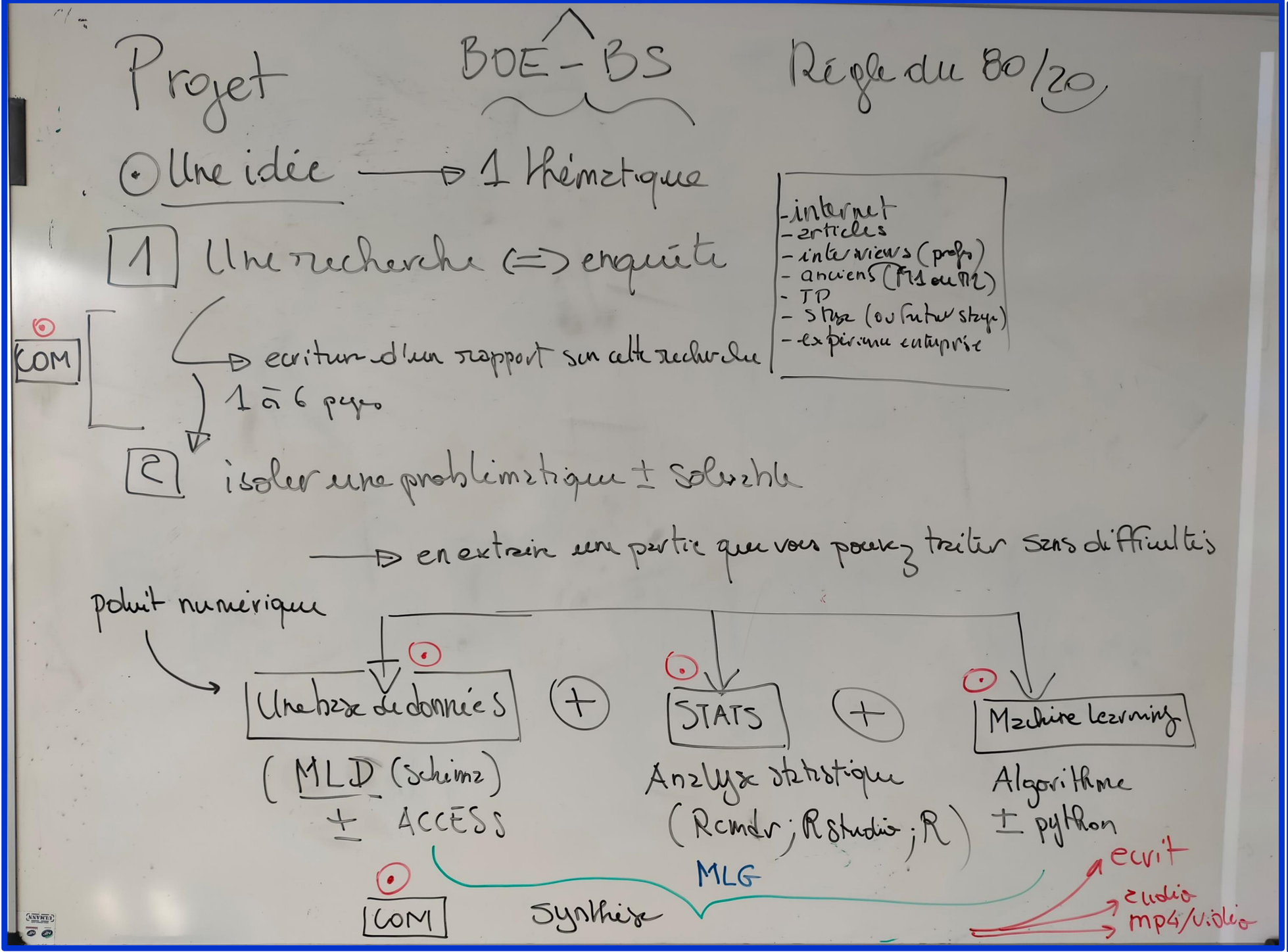
Envoi du lien de ce fichier (transmission fichiers volumineux)
par email à : pascal.rigolet@u-psud.fr



Ne pas élider vos idées
dans votre projet

Les différentes branches de votre projet peuvent être totalement déconnectées les unes des autres ; c'est le rapport écrit qui les met en scène et fait une synthèse du travail réalisé, mettant également en valeur toutes les idées que vous avez eues mais que vous n'avez pas concrétisées (ou pas pu concrétiser) en DB, Stats ou ML. Comme pour les défis : no limit (vous êtes des scientifiques, on ne doit pas entraver vos idées, votre élan ou votre enthousiasme).

Et si vous avez peur d'être ridicules dans vos idées, n'oubliez pas que le ridicule ne tue pas…



- ⇨ Base de données
- ⇨ Traitement statistique (copies d'écran + commentaires)
- ⇨ Algorithme (extrait)
- ⇨ Programme python
- ⇨ Rapport écrit (~12 pages)
- ⇨ Diaporama sonorisé/ vidéo mp4 (5 à 10 dias)

Organisation soignée

(répertoires dédiés)

Envoi du **lien** (par mail)

du fichier compressé zip

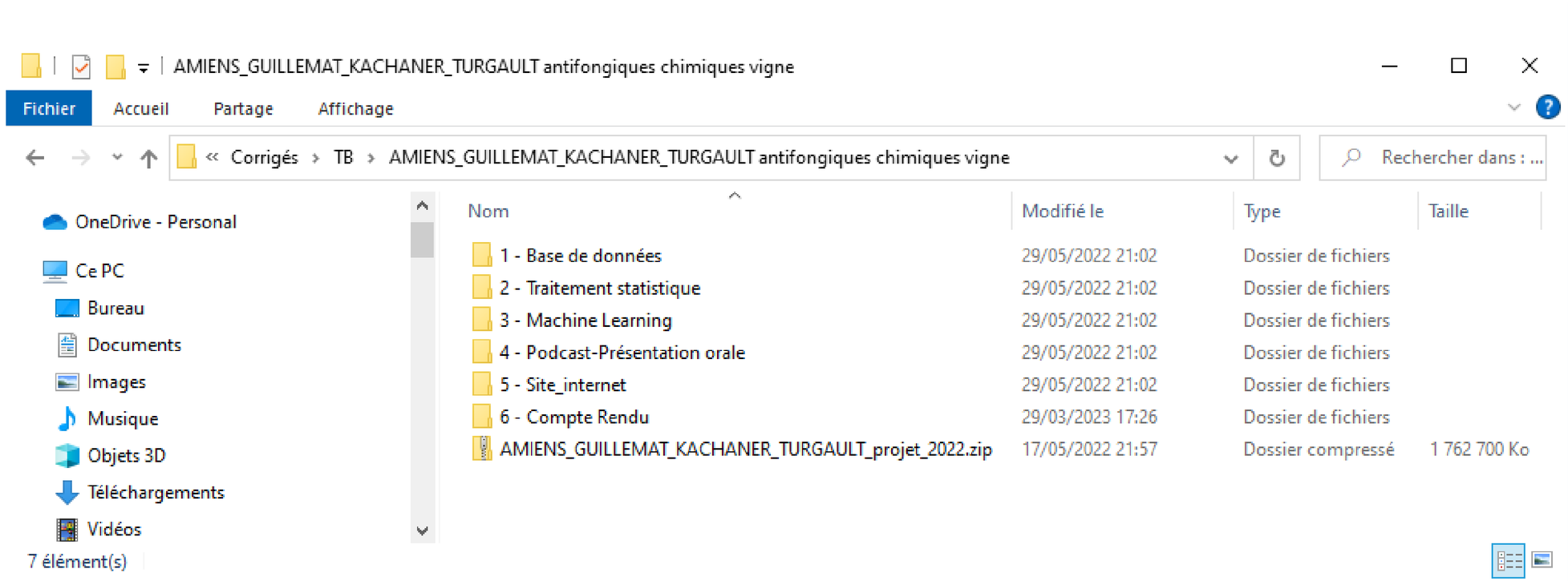
contenant tous les fichiers

- ⇨ Deadline :

mercredi 15 mai – 21h

envoi notes administration :

vendredi 24 mai



- ⇨ Base de données
- ⇨ Traitement statistique (copies d'écran + commentaires)
- ⇨ Algorithme (extrait)
- ⇨ Programme python
- ⇨ Rapport écrit (~12 pages)
- ⇨ Diaporama sonorisé/ vidéo mp4 (5 à 10 dias)

Organisation soignée

(répertoires dédiés)

Envoi du **lien** (par mail)

du fichier compressé zip

contenant tous les fichiers

- ⇨ Deadline :

mercredi 15 mai à 21h

envoi notes administration :

vendredi 24 mai