**TraAM Former au *big data* et au *data mining* pour comprendre les dynamiques écosystémiques**

Equipe Collège : Camille PEYBERNES, Fabien PIKORKI

Les bases de données de biodiversité spécifique : un outil pour comprendre les dynamiques écosystémiques dans l’espace et le temps.

 Contexte : les données de biodiversité (nombre d’espèces d’un écosystème ou de la biosphère, évolution du nombre de groupes fossiles au cours du temps…) sont souvent présentées aux classes de collège sans explications de l’origine de ces données ni de la manière dont elles ont été collectées et agrégées. La base de données est donc vue comme une « boîte noire » d’où l’on tire des informations.

 Le but des activités présentées ici est de rendre plus explicites les pratiques scientifiques qui ont permis l’élaboration de ces bases de données et leur utilisation.

 La biodiversité des espèces est choisie comme sujet d’étude car c’est le type de diversité le plus facile à appréhender pour un élève de collège. Les deux ateliers proposés permettent de travailler sur l’évolution de la biodiversité du passé à l’échelle globale ainsi que sur la biodiversité actuelle à l’échelle de l’écosystème.

**Atelier 1 : *Escape game* les crises de la biodiversité (réalisé au collège de Pont-en-Royans en février 2021)**

 Les données de biodiversité fossiles sont très présentes dans les manuels de SVT mais correspondent le plus souvent à des graphiques qui n’explicitent pas l’origine des données brutes (exemple : courbe de Sepkoski montrant les 5 grandes crises biologiques).

 L’activité proposée se base sur une adaptation de la base de données de biodiversité fossile de Benton (*the Fossil Record*), développée par acces.inrp. Le format de l’activité (*escape game*) permet une approche ludique afin de faciliter la prise en main par les élèves des multiples fichiers Excel qui composent cette base de données.

Objectifs : Apprendre à utiliser une base de données pour répondre à un problème. Comprendre comment sont structurées les données brutes dans une base.

Niveau : 3ème

Matériel : Salle informatique, ensemble des fichiers de la base de données « *the fossil record 2* » (Lien de téléchargement : acces.ens-lyon.fr/acces/thematiques/limites/paleobiodiversite/developper/data/benton/Tous%20groupes.zip)

Déroulé :

*Le scénario proposé ici est donné à titre indicatif et correspond à des exemples vus par les élèves au cours de la séquence sur la biodiversité fossile et les crises biologiques. Chacun est donc libre de l’adapter en fonction de sa propre progression et de modifier les groupes étudiés ou l’intrigue. L’accompagnement du professeur est capital pour maintenir le rythme de l’escape game et ainsi faire durer l’engagement des élèves tout au long de l’activité.*

**1. Introduction :** « Le maléfique professeur Benton vous a enfermé dans la salle informatique du collège. Pour sortir vous devez trouver le mot de passe. Pour cela vous devez résoudre l’énigme en utilisant les informations cryptées contenues dans sa fameuse base de données de la biodiversité des temps géologiques : *the Fossil Record 2* » !

**2. Présentation des questions pour résoudre l’énigme :**

a) Quel groupe disparait à la fin du Trias ?

b)  Quand les trilobites étaient-ils le plus diversifiés ?

c) Quelle crise n’entraine pas d’extinction totale parmi les groupes étudiés ?

 d) Quel groupe a vécu le moins longtemps ?

 e) Quel groupe a survécu à toutes les crises ?

**3.** MAIS DE QUELS GROUPES S’AGIT-IL ? => Un mot est griffonné sur un papier au sol (préparé par le professeur et adapté aux groupes fossiles vus en classe) : ECHINOTRILOCONOPLACOAMMO

**4. Recherche collective des groupes concernés**, aide du professeur (prévoir des images ou des spécimens de chaque groupe biologique pour les rendre concrets pour les élèves).

**5.** Les élèves se rendent compte que l’énigme concerne cinq groupes biologiques : les échinodermes, les trilobites, les placodermes, les conodontes et les ammonites.

**6.** Les élèves sont ensuite répartis en 5 groupes, chacun étant responsable de l’étude d’un groupe d’animaux dans la base de données (un ordinateur par groupe). Les élèves doivent d’abord comprendre la structure de la base de données (4 dossiers, 45 fichiers) puis comprendre comment est structurée l’information dans chaque fichier (présence/absence des familles au cours des temps géologiques, étages et âges en Ma).

Exemple de fichiers à étudier (fragment) : occurrences des familles de conodontes au cours des temps géologiques



**7.** Pour répondre, les élèves doivent construire un graphique dans un tableur pour montrer l’abondance du groupe biologique (nombre de familles) au cours des temps géologiques. Prévoir une fiche méthode ou s’assurer au préalable que les élèves savent faire un graphique dans un tableur. *Astuce : pour chaque fichier, un onglet du tableur permet d’obtenir directement le graphique.*

Exemple de graphique à réaliser :



*Document de secours avec les 5 groupes sur le même graphique :*

 Source : fossilrecord.net

**8.** Une fois que les élèves ont obtenu leur graphique, il faut ajouter des informations pour mieux le comprendre en ajoutant les 5 grandes crises biologiques (extinctions de masse) au cours des temps géologiques.Les élèves doivent indiquerpar des flèches ces crises biologiques sur leur graphique.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Crise biologique | Fin de l’Ordovicien | Fin du Dévonien | Fin du Permien | Fin du Trias | Fin du Crétacé |
| Date (Ma : millions d’années) | -445 | -360 | -250 | -200 | -65 |

**9.** A l’aide du graphique obtenu, les élèves répondent collectivement aux questions. Les indices entre parenthèses fourniront les clés de l’énigme pour trouver le mot de passe:

a) Quel groupe disparait à la fin du Trias ? (première lettre)

 *Conodonte*: Première lettre *=> C*

b)  Quand est ce que les trilobites étaient le plus diversifié ? (cinquième lettre)

  *Cambrien*: Cinquième lettre *=> R*

c) Quelle crise n’entraine pas d’extinction totale parmi les groupes étudiés ? (sixième lettre)

  *Ordovicien* : sixième lettre *=>I*

d) Quel groupe a vécu le moins longtemps ? (dernière lettre)

 *Placodermes* : dernière lettre =>S

e) Quel groupe a survécu à toutes les crises ? (première lettre)

 *Echinodermes* : Première lettre *=> E*

 ***(Réponse : Crise)* Bravo ! Vous êtes sorti !**



Fin de séance : Une fois que les élèves se sont familiarisés avec les données brutes, on peut leur faire une démonstration de l’utilisation de *Paleobiology database* (<https://paleobiodb.org/#/>). Cette riche base de donnée en ligne (et en anglais) permet la visualisation et le téléchargement de très nombreuses données paléontologiques. Les élèves prennent alors conscience de l’immense quantité de données accumulées pour obtenir les connaissances actuelles sur l’évolution de la biodiversité au cours des temps géologiques.

**Atelier 2 : Contribuer à une base de données (réalisé au collège de Pont en Royans en Mai 2021)**

Objectif : Cet atelier vise à faire comprendre comment se construit une base de données et comment elle est enrichie en permanence par de nouvelles données. À la fin de l’atelier une utilisation de la base de données permet aux élèves d’appréhender l’intérêt des bases de données comme source d’information scientifiques ainsi que de comparer les données acquises sur le terrain avec les données agrégées de la base.

Niveau : 3eme

Matériel :

- Téléphones portables avec application PlantNet ;

- Piquets-ficelles afin de délimiter des quadrat d’échantillonnage (1 m2) ;

- Eventuellement une flore simplifiée.

Déroulé :

 L’activité a été proposée aux élèves dans le cadre d’une sortie dans l’Espace Naturel Sensible des Rimets (Rencurel, Isère).

 Elle fait suite à la visite du sentier géologique (http://www.rencurel-vercors.fr/183-vallee-fossile-des-rimets-espace-naturel-sensible-.htm) présent sur le site et qui permet de réinvestir les notions vues en classe sur la biodiversité du passé (voir atelier 1) et les paléoenvironnements.

 Cette activité peut être adaptée en fonction de l’ENS visité ou même réalisée dans un environnement proche du collège.

**1. Séance de collecte de données sur le terrain**

 Mise en place et réalisation d’un protocole d’échantillonnage et de mesure : les élèves sont amenés à délimiter un carré de pelouse de 1 m par 1 m avec des piquets et une ficelle puis de compter et déterminer les espèces observées (insectes et plantes) notamment grâce à l’application PlantNet.

 Les élèves ont reconnu entre 7 et 14 espèces, principalement des plantes à fleurs. Cet exercice permet de mettre en lumière la richesse spécifique de cet écosystème. À la fin de la collecte, une réflexion est conduite avec les élèves pour essayer d’expliquer la disparité ces résultats (biais d’échantillonnage, variabilité des paramètres environnementaux …).

On peut aussi réfléchir avec les élèves sur la validité des déterminations proposées et sur l’esprit critique nécessaire pour utiliser des outils numériques comme PlantNet. Il est en effet important que les élèves comprennent que les déterminations proposées par PlantNet fonctionnent par reconnaissance d’image à partir d’une base de données en constante évolution et à laquelle ils contribuent.

**2. Intérêt et limites rencontrés sur le terrain.**

Cette séance est l’occasion de faire connaître des applications de sciences participatives comme PlantNet. Après discussions avec les élèves il s’avère que certains continuent d’utiliser l’application PlantNet installée dans le cadre de cette sortie. Cela permet donc de développer leur curiosité et de faire évoluer leur regard sur leur environnement proche.

Il parait indispensable de sensibiliser les élèves à l’environnement dans lequel ils font la recherche, ici un ENS. En effet, les élèves peuvent rapidement piétiner des espèces sans le vouloir ou être tentés d’en ajouter dans le carré d’échantillonnage.

 L’utilisation de l’application « PlantNet » nécessite d’avoir évidement une connexion internet et demande une gestion de l’outils numérique rigoureuse avec les élèves. L’idée de faire découvrir aux élèves l’utilisation d’une flore simplifiée peut être intéressante.

 Lorsque la cueillette est autorisée, on peut evisager faire réaliser un herbier « biodiversité ».

**3. Séance de tri, standardisation et utilisation des données recueillies.**

Une séance dédiée au tri des données recueillies, à leur standardisation et à la formalisation de la base de données a été envisagée mais n’a pas pu être réalisée. De même, il était prévu de faire remonter les informations collectées vers une base de données de référence ou vers un site de science participative puis d’utiliser cette base de données de référence (exemple : comparaison des données de biodiversité acquises localement et données de la base INPN de la commune : nouvelles espèces/bases ? espèces manquantes/bases ?). Cela n’a pas pu être réalisé cette année.

*N. B. : Dans le cadre du projet Regards Croisés sur l’environnement mené au collège de Pont en Royans, les données obtenues ont été partagées avec des élèves du collège sénégalais de Dindefelo. Ces derniers réaliseront des mesures similaires dans la réserve naturelle proche de leur collège. Ce partage permettra l’étude comparée de la biodiversité de chaque environnement, en France et au Sénégal. En parallèle des activités présentées, ici les élèves ont aussi échangé des images de mammifères capturées par des caméras pièges placées à proximité de leur collège, en France et au Sénégal.*

[*https://www.regardscroises.tetraktys-ong.org/college-dindefelo-college-raymond-guelen/*](https://www.regardscroises.tetraktys-ong.org/college-dindefelo-college-raymond-guelen/)

Lien avec des programmes de sciences participatives :

- <https://plantnet.org/#participer>



- INPN espèces : <https://inpn.mnhn.fr/docs/communication/flyer-INPN-Especes-2020.pdf>



- inventaire insecte pollinisateur : <https://www.spipoll.org/>

*Il se pose souvent le problème de l’accès aux données compilées dans la base des programmes de sciences participatives. Nous n’avons pas trouvé de moyen simple pour télécharger les données sur les sites des programmes de sciences participatives. Seule la base de l’INPN est facilement accessible.*

Partenaires :

 Département de l’Isère, Tetraktys (association mettant en relation des élèves de collèges isérois et sénégalais dans le cadre du projet Regards Croisés)

