



Collecte des données et contrôle qualité avec Excel / Data Collection and Quality Control with Excel

Florèn Hugon, Xavier Navarro, Frank d'Amico, Charlotte Chazeau, Nicolas
Gasco

► To cite this version:

Florèn Hugon, Xavier Navarro, Frank d'Amico, Charlotte Chazeau, Nicolas Gasco. Collecte des données et contrôle qualité avec Excel / Data Collection and Quality Control with Excel. [Rapport Technique] Université de Pau et des Pays de l'Adour (UPPA), FRA.; Laboratoire de Mathématiques et de leurs Applications de Pau. 2022. hal-03695897

HAL Id: hal-03695897

<https://univ-pau.hal.science/hal-03695897>

Submitted on 15 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Webinaire Scientifique

"Collecte des données et Contrôle Qualité avec Excel"

Mardi 31 mai 2022, 9h-12h30

Organisé par
Florèn HUGON, Docteur Physiologie et Biologie des Organismes

En collaboration avec l'Université de Pau et des Pays de l'Adour (UPPA) et le Muséum National d'Histoire Naturelle (MNHN)

Avec les interventions de

Frank D'AMICO, Maître de Conférences
Laboratoire de Mathématiques et de leurs Applications de Pau (UPPA)

Xavier NAVARRO, Maître de Conférences,
Laboratoire Informatique de l'Université de Pau et des Pays de l'Adour (UPPA)

Nicolas GASCO, Chargé d'Études Scientifiques,
Laboratoire de biologie des organismes et des écosystèmes aquatiques (MNHN)

Charlotte CHAZEAU, Chargée d'Études Scientifiques
Laboratoire de biologie des organismes et des écosystèmes aquatiques (MNHN)

Hugon F., Navarro X., D'Amico F., Chazeau C. et Gasco N., Collecte des données et Contrôle Qualité avec Excel, Webinaire scientifique, UPPA – MNHN, à Pau, le 31/05/2022, <https://github.com/FloHugon/DataFormatting>.

Documents joints dans ce fichier pdf

- **Présentations dans l'ordre des interventions**

→ Introduction : quelques considérations sur la collecte des données, Frank D'AMICO
→ Introduction à la programmation sur Excel, Xavier NAVARRO
→ Excel: un outil adapté pour contrôler la qualité des données. Exemple sur des données de suivis d'espèces pour une modélisation de l'abondance, Florèn HUGON
→ Excel: un outil adapté à l'acquisition des données. Exemple sur des données de pêcheries pour implémentation dans une base de données, Nicolas GASCO et Charlotte CHAZEAU

- **Compte rendu des commentaires et questions** sur les présentations et lors de la section partage d'expérience (projets, difficultés rencontrées, ...)

- **Bilan des retours, proposition d'une seconde session et d'une formation**

Références bibliographiques

D'Amico F., Kermorvant C., Sanchez J.M. & Arizaga J. (2020). Optimal sampling design to survey riparian bird populations with low detection probability. *Bird Study*, 67:2, 160-167, DOI: <https://doi.org/10.1080/00063657.2020.1784090>

Hugon F., Navarro X., Rodriguez M., Das Neves G., Berroneau M. & D'Amico F. (2022). — Mise en forme et contrôle qualité de données, l'informatique au service de l'écologie. *Naturae* (2): 17-30. <https://doi.org/10.5852/naturae2>

Kermorvant C., Coube S., D'Amico F., Bru N. & Caill-Milly N. (2020). "Sequential process to choose efficient sampling design based on partial prior information data and simulations." *Spatial statistics* 38 : 100439. <https://doi.org/10.1016/j.spasta.2020.100439>

Kermorvant C., D'Amico F., Robertson B., Bru N. & Caill-Milly N. (2019). Spatially balanced sampling designs for environmental surveys. *Environmental Monitoring and Assessment* 191:524 / <https://doi.org/10.1007/s10661-019-7666-y>

Remerciements

Nous remercions Émilien CARUSO pour l'organisation logistique. Concernant les travaux présentés par Florèn HUGON, nous remercions Matthieu BERRONEAU pour l'acquisition des données, Matt RODRIGUEZ pour l'écriture du code informatique ainsi que le programme les Sentinelles du Climat et les financeurs, l'Union Européenne, la Région Nouvelle Aquitaine et les départements des Pyrénées Atlantiques et de la Gironde. Enfin, nous remercions l'ensemble des 45 participants pour leur intérêt et leurs remarques constructives.

Contact organisation

Florèn HUGON, floren.hugon@live.fr

Le document est également déposé sur ResearchGate (<https://www.researchgate.net/profile/Floren-Hugon>) et le sera prochainement dans HAL (lien indiqué dans la section Readme du dépôt GitHub).

Collecte des données et Contrôle Qualité avec Excel

Florèn HUGON, Docteur en Physiologie et Biologie des Organismes

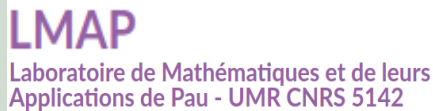
Nicolas GASCO, Chargé d'Études Scientifiques

Xavier NAVARRO, Maître de Conférences

Frank D'AMICO, Maître de Conférences

Charlotte CHAZEAU, Chargée d'Études Scientifiques

Mardi 31 mai 2022



Présentations des intervenants



LMAP
Laboratoire de Mathématiques et de leurs
Applications de Pau - UMR CNRS 5142

Florèn HUGON
Frank D'AMICO

LIUPPA
Laboratoire Informatique de l'Université de Pau
et des Pays de l'Adour - LIUPPA

Xavier NAVARRO



Nicolas GASCO
Charlotte CHAZEAU

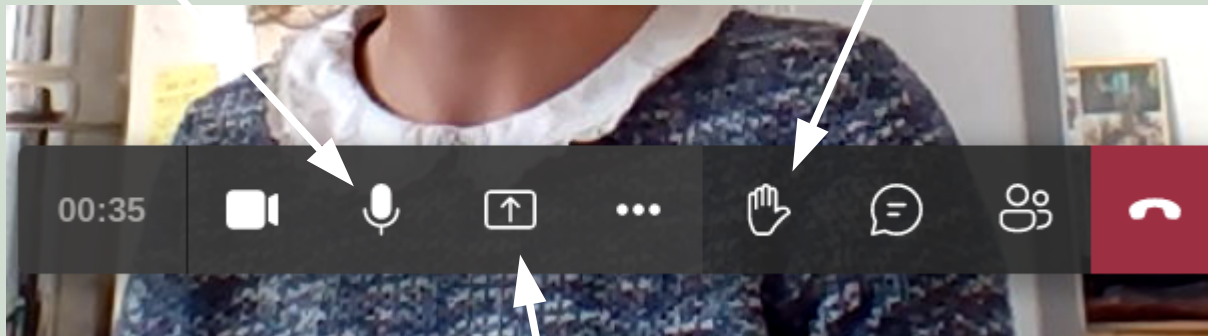
Programme détaillé

8h30 - 9h	Accueil des participants
9h - 9h15	Présentation des intervenants
9h15 - 9h45	Quelques considérations sur la collecte des données Frank D'AMICO (15min présentation + 15 min questions)
9h45 - 10h15	Introduction à la programmation sur Excel Xavier NAVARRO (15min P + 15 min Q)
10h15 - 10h50	Excel, un outil adapté pour la mise en forme et le contrôle qualité des données. Exemple sur des données collectées pour modéliser l'abondance Florèn HUGON (20min P + 15 min Q)
10h50 - 11h05	Pause
11h05 - 11h50	Excel, un outil adapté à l' acquisition qualitative des données . Exemple sur des données de pêcheries pour implémentation dans une base de données Nicolas GASCO et Charlotte CHAZEAU (30min P + 15min Q)
11h50 - 12h20	Interventions libres, discussion sur votre utilisation d'Excel et vos besoins
12h20 - 12h30	Conclusions et perspectives

Déroulement

Couper les micros

Lever la main pour prendre la parole au cours des temps d'échange



Partager son écran si besoin dans la session interventions libres

Envoi du pdf des présentations + Compte rendu des questions

Frank D'Amico

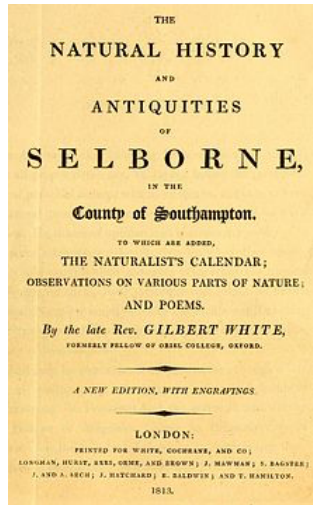
- Laboratoire de Mathématiques et de leurs Applications [UMR CNRS 5142] Equipe proba/stats
- Fédération de recherche 'Milieux & Ressources Aquatiques' [FED MIRA 4155]

Webinaire Collecte des données et Contrôle Qualité avec Excel

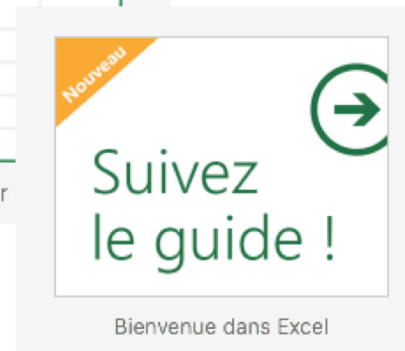
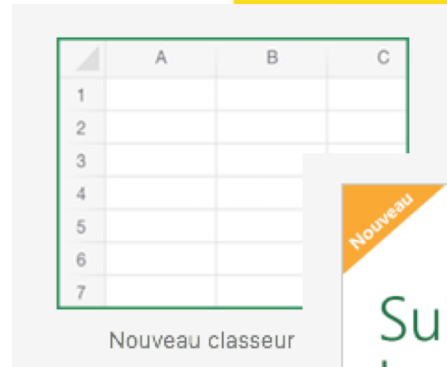
**9h15 – 9h45 : Quelques considérations sur la
collecte des données**



❑ Collecte des données: de la saisie au partage.



White, Gilbert [1789].
[The Natural History and Antiquities of Selborne](#).
London: Cassell & Company.



https://en.wikipedia.org/wiki/The_Natural_History_and_Antiquities_of_Selborne

<https://cybertracker.org/software-introduction/>

<https://play.google.com/store/apps/details?id=mnhn.CardObs.CarNat>

<https://cardobs.mnhn.fr/cardObs/>

❑ Quelques constats....

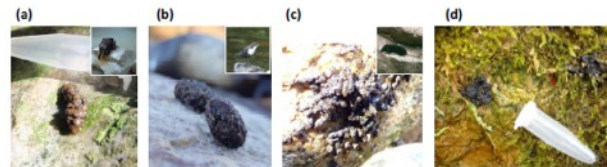
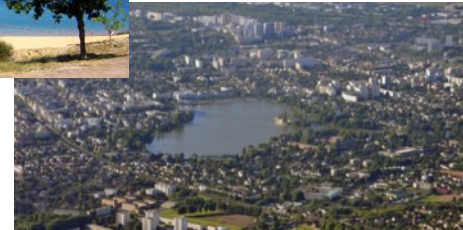
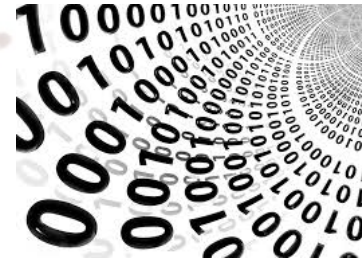
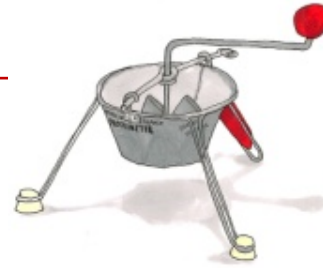
à nuancer en fonction des objectifs

- le passage des données « à la moulinette statistique »
- un certain tropisme naturaliste:
(des **1**... et des **0** !)
- l'environnement est hétérogène
- le processus d'observation* est imparfait (\Leftrightarrow erreurs!)

* et la saisie des données aussi...

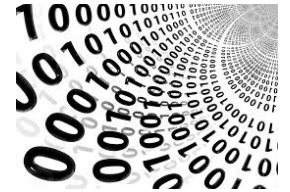
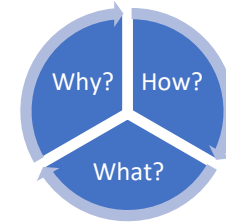


- un certain optimisme naturaliste



❑ Quelques préconisations....

à nuancer en fonction des objectifs



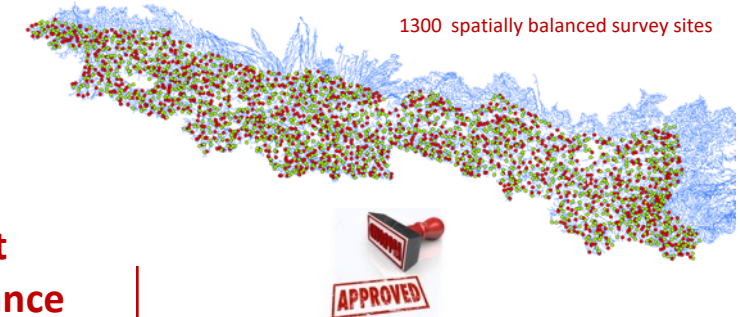
- le passage des données « à la moulinette statistique »
- gérer le tropisme naturaliste:
il faut des **1**... et des **0** !
- l'environnement est hétérogène
- le processus d'observation* est imparfait (\Leftrightarrow erreurs!)



* et la saisie des données aussi...

- gérer l'optimisme naturaliste

- **inverser la démarche habituelle et concevoir (travail d'équipe) l'analyse des résultats AVANT !**
- **recourir au maximum à la réplication spatiale et temporelle ; s'obliger à respecter un protocole probabiliste (BAS...) qui force à aller là où on n'irait pas (admettre l'ingratitude ou la frustration du travail de terrain)**
- **mesurer (modéliser) l'hétérogénéité**
- **recourir au maximum à la réplication spatiale et temporelle ; s'appuyer au maximum sur la puissance des modèles (approches *design-based* et *model-based*)**
- **modérer (modéliser) l'excès de confiance (s'appuyer sur les derniers développements statistiques)**





LANGAGE BASIC

SOUS EXCEL

XABI NAVARRO, MC INFORMATIQUE, COLLÈGE STEE, LIUPPA

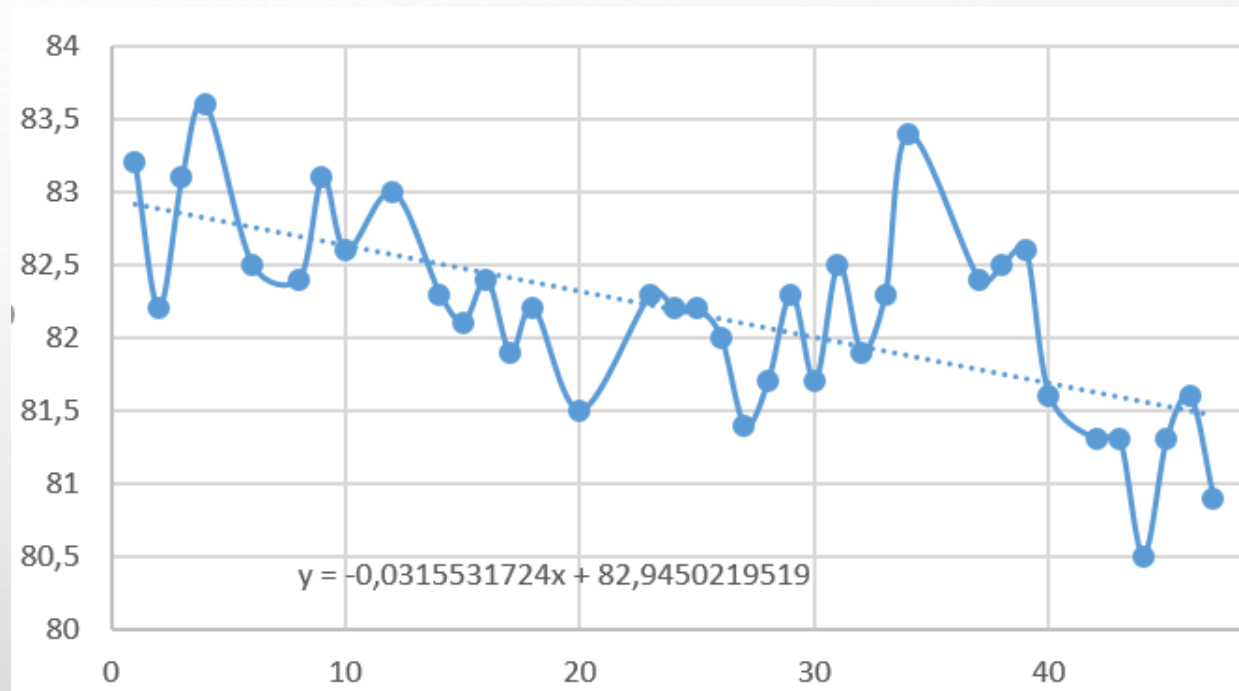


INFORMATIQUE

- TRAITEMENT AUTOMATIQUE DE L'INFORMATION
- RÉCOLTE DES DONNÉES, PRÉPARATION DES DONNÉES, TRAITEMENT DES DONNÉES
- RÉCOLTE AVEC UN OUTIL INFORMATIQUE, TRAITEMENT AVEC UN AUTRE
- PRÉPARATION DES DONNÉES AVEC LE TABLEUR EXCEL
- NÉCESSITÉ D'AVOIR QUELQUES RUDIMENTS DE PROGRAMMATION
- ILLUSTRATION AVEC UN EXEMPLE

COLLECTE ET TRAITEMENT

	A	B	C
1	NOMBRE De Relevés	37	
2		Jours	Poids
3	jeudi 7 juin 2012	1	83,2
4	vendredi 8 juin 2012	2	82,2
5	samedi 9 juin 2012	3	83,1
6	dimanche 10 juin 2012	4	83,6
7	mardi 12 juin 2012	6	82,5
8	mercredi 13 juin 2012	8	82,4
9	jeudi 14 juin 2012	9	83,1
10	vendredi 15 juin 2012	10	82,6
11	samedi 16 juin 2012	12	83
12	dimanche 17 juin 2012	14	82,3



ALGORITHME DU TRAITEMENT

La méthode des moindres carrés indique³³ :

$$- \text{pente} = \frac{n * \sum_{i=1}^n (\text{jour}_i * \text{poids}_i) - \sum_{i=1}^n \text{jour}_i * \sum_{i=1}^n \text{poids}_i}{n * \sum_{i=1}^n \text{jour}_i^2 - (\sum_{i=1}^n \text{jour}_i)^2}$$

$$- \text{ordonnée} = \frac{1}{n} * \left(\sum_{i=1}^n \text{poids}_i - \text{pente} * \sum_{i=1}^n \text{jour}_i \right)$$

où n est le nombre de relevés.

TRAITEMENT AVEC FORMULES

- CONSTRUIRE COLONNE DES JOURS*POIDS
- CONSTRUIRE COLONNE DES JOURS*JOURS
- FAIRE LES SOMMES
- FORMULE DU NUMÉRATEUR
- FORMULE DU DÉNOMINATEUR

s	37		
	Jours	Poids	journspoids
	1	83,2	=B3*C3
2	2	82,2	164,4

Lundi 16 Juillet 2012	46
Mardi 17 Juillet 2012	47
Les sommes	=SOMME(B3:B39)

38	Lundi 16 Juillet 2012	46	81,6	3753,6	2116
39	Mardi 17 Juillet 2012	47	80,9	3802,3	2209
40					
41	les sommes	918	3040	75201,1	29868
42		842724			
43				numérateur	-8279,3
44				dénominateur	262392
45					
46				pente	-0,0315531724

$$y = -0,0315531724x + 82,9450$$

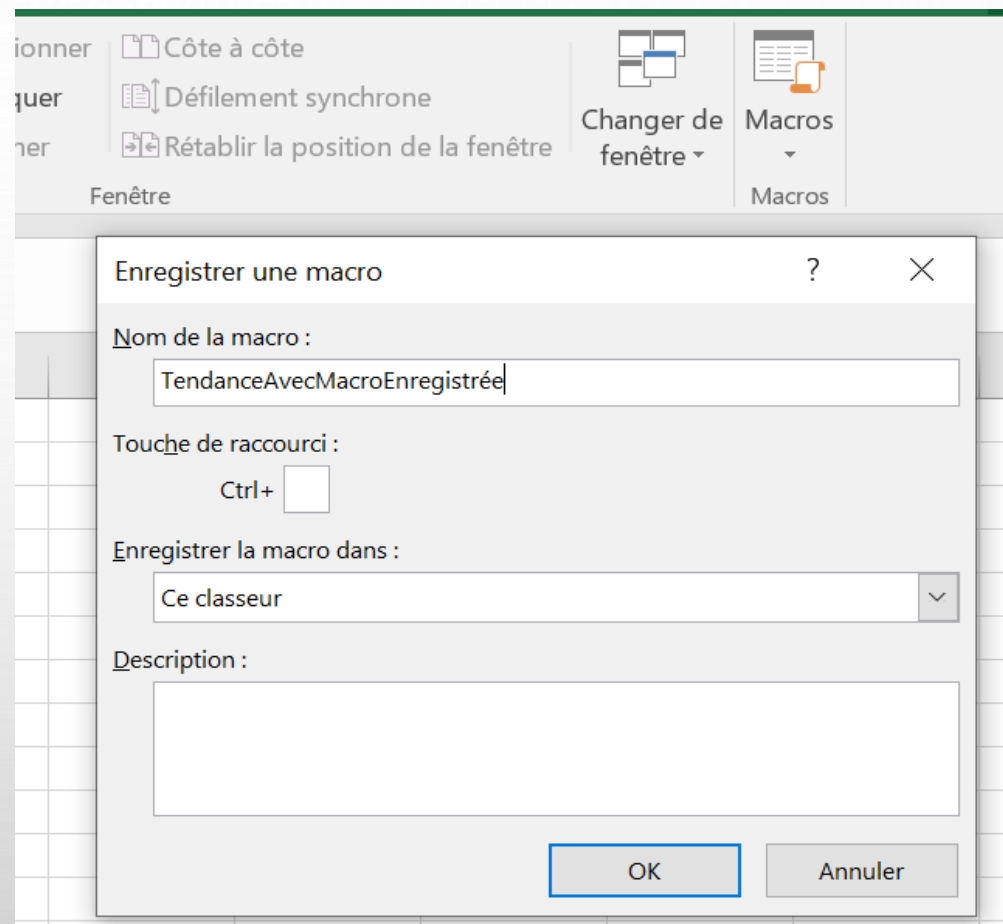
POSSIBILITÉ D'ENREGISTRER UNE MACRO

- AFFICHAGE, MACRO
 - DONNER UN NOM À LA MACRO
 - ENREGISTRER UNE MACRO
- ENREGISTREMENT DES « CLICS »
- ARRÊTER L'ENREGISTREMENT
- EXÉCUTER LA MACRO

LIMITE

**MACRO VALABLE UNIQUEMENT
POUR UN NOMBRE DE RELEVÉS
FIGÉ, 37**

Xabi Navarro, Collège STEE, LIUPPA



ADAPTATION DE LA MACRO

QUELQUE SOIT LE NB DE RELEVÉS

- COMPRÉHENSION DU CODE GÉNÉRÉ
 - SÉLECTION CELLULE
 - METTRE DANS LA CELLULE ACTIVE
 - RC[-2]*RC[1]
 - R : MÊME LIGNE
 - C[-2] : 2 COLONNES AVANT
 - ...
- MODIFICATION *TRÈS ARDUE*
 - RANGE("D4:D39").SELECT POUR DIRE QUE 39 EST LA CELLULE B1+2

```
Sub TendanceAvecMacroEnregistrée()  
'  
' TendanceAvecMacroEnregistrée Macro  
'  
  
Range("D2").Select  
ActiveCell.FormulaR1C1 = "Jour*Poids"  
Range("D3").Select  
ActiveCell.FormulaR1C1 = "=RC[-2]*RC[-1]"  
Range("D3").Select  
Selection.Copy  
Range("D4:D39").Select  
ActiveSheet.Paste  
Range("E2").Select  
Application.CutCopyMode = False  
ActiveCell.FormulaR1C1 = "jour*jour"  
Range("E3").Select  
ActiveCell.FormulaR1C1 = "=RC[-3]*RC[-3]"  
Range("E3").Select
```


LA PROGRAMMATION AVEC UN LANGAGE IMPÉRATIF : BASIC

- BASÉ SUR LA NOTION DE VARIABLE
- UNE VARIABLE
 - A UN NOM : POIDS
 - A UNE ET UNE SEULE VALEUR À UN INSTANT DONNÉ DE L'EXÉCUTION DU PROGRAMME
 - A UN ENSEMBLE DE VALEURS POSSIBLES ; ENSEMBLE DÉFINI PAR SON TYPE
- UNE VARIABLE EST AFFECTÉE PAR UNE EXPRESSION
 - $\text{JOURSPIDS} = \text{JOURS} * \text{POIDS}$

LES VARIABLES

- TYPES ENTIERS
 - **INTEGER, LONG** SUIVANT LE NB D'OCTETS
 - INTEGER SUR 2 OCTETS : -32768 À 32767 , -2^{15} À $2^{15}-1$
 - *ON NE PEUT PAS AVOIR DES NOMBRES JUSQU'À L'INFINI*
- TYPES DÉCIMAUX
 - **SINGLE, DOUBLE** PRÉCISION
 - *ON NE PEUT PAS AVOIR TOUS LES DÉCIMAUX QU'IL EXISTE ENTRE 2 DÉCIMAUX*
- TYPES CHAÎNE DE CARACTÈRES
 - **STRING**
- ...
- AFFECTATION : $\text{JOURSPIDS} = \text{JOURS} * \text{POIDS}$
 - ÉVALUATION DE L'EXPRESSION, PARTIE DROITE
 - AFFECTATION PROPREMENT DITE À LA VARIABLE PRÉSENTE DANS LA PARTIE GAUCHE

VERSION 000

```
Sub analyseV000()  
'Titre : Calcul de la tendance à partir de relevés de poids  
'Auteur: Xabi Navarro  
'Résultat pente attendu= -0,0315531724  
'Avec la méthode des moindres carrés  
  
'Variables en entrée : les jours et les poids  
'Variables de traitement : joursjours et jourspoids  
'Variable en sortie : la pente  
  
'Construction de la colonne jourspoids  
'Construction de la colonne joursjours  
'Les Sommes  
'  la somme des jours  
'  la somme des poids  
'  la somme des jourspoids  
'  la somme des jours * jours  
' le numérateur  
' le dénominateur  
' la pente  
End Sub
```

VERSION 001

METTRE DU TEXTE DANS UNE CELLULE

- `CELLS(NUMÉRO DE LIGNE, NUMÉRO DE COLONNE) = " ... "`

```
Sub analyseV001()  
'Titre : Calcul de la tendance à partir de relevés de poids  
'Auteur: Xabi Navarro  
'Résultat pente attendu= -0,0315531724  
'Avec la méthode des moindres carrés  
  
'Variables en entrée : les jours et les poids  
'Variables de traitement : joursjours et jourspoids  
'Variable en sortie : la pente  
  
'Construction de la colonne jourspoids  
Cells(2, 4) = "jourspoids"
```

	jourspoids
83,2	83,2
82,2	164,4

VERSION 002 (1)

LE PREMIER JOURS*POIDS

- QUEL EST LE TYPE DE JOURS ?
- QUEL EST LE TYPE DE POIDS ?
- QUEL EST LE TYPE DE JOURSPIDS ?
- ... EFFECTUER LES DÉCLARATIONS DES VARIABLES EN CONSÉQUENCE

```
'Variables en entrée : les jours et les poids  
Dim jours As Integer  
Dim poids As Single  
'Variables de traitement : |jourspoids  
Dim jourspoids As Single
```

VERSION 002 (2)

LE PREMIER JOURS*POIDS

- RÉCUPÉRER LA VALEUR DU 1^{ER} JOUR
- RÉCUPÉRER LA VALEUR DU 1^{ER} POIDS
- FAIRE LE CALCUL
- METTRE LE RÉSULTAT DANS LA CELLULE ADÉQUATE

```
'Cells(3, 4) = Cells(3, 2) * Cells(3, 3)
jours = Cells(3, 2)
poids = Cells(3, 3)
jourspoids = jours * poids
Cells(3, 4) = jourspoids
```

B	C	D
37		
Jours	Poids	jourspoids
1	83,2	83,2

VERSION 002 (3)

LE PREMIER JOURS*POIDS

- TESTER POUR VÉRIFIER

37		
	Poids	joursoids
1	83,2	83,2
2	82.2	

- TESTER LA PRÉCISION VOULUE POUR VÉRIFIER : 10 DÉCIMALES

	Poids	joursoids
1	83,2	83,1999969482
2	82.2	

- LA VARIABLE POIDS AVEC 10 DÉCIMALES NÉCESSITE UNE DOUBLE PRÉCISION
EN CONSÉQUENCE, JOURSPIDS AUSSI

VERSION 003

TOUS LES JOURS*POIDS

- EFFECTUER LE CALCUL DE LA LIGNE 3 POUR LES LIGNES DE 3 À 39
 - METTRE UNE BOUCLE FOR AVEC UN INDICE DE BOUCLE
 - REMPLACER TOUS LES 3 CORRESPONDANTS AU NUMÉRO DE LIGNE PAR L'INDICE DE BOUCLE

```
jours = Cells(3, 2)  
poids = Cells(3, 3)  
jourspoids = jours * poids  
Cells(3, 4) = jourspoids
```

```
Dim jourspoids As Double  
Dim i As Integer  
'Variable en sortie : la pente  
  
'Construction de la colonne jourspoids  
Cells(2, 4) = "jourspoids"  
'Cells(3, 4) = Cells(3, 2) * Cells(3, 3)  
For i = 3 To 39  
    jours = Cells(i, 2)  
    poids = Cells(i, 3)  
    jourspoids = jours * poids  
    Cells(i, 4) = jourspoids  
Next
```


VERSION 004

UNE SOMME

- EFFECTUER LA SOMME DES JOURS
 - INITIALISER LA SOMME À 0 : SOMJOURS = 0
 - AJOUTER À CHAQUE FOIS À LA SOMME, LA VALEUR DU JOUR : SOMJOURS = SOMJOURS+JOURS
 - NE PAS OUBLIER DE DÉCLARER SOMJOURS : DIM SOMJOURS AS INTEGER

```
'Les Sommes
Cells(41, 1) = "Les sommes"
'   la somme des jours
somsjours = 0
For i = 3 To 39
    jours = Cells(i, 2)
    somsjours = somsjours + jours
Next
Cells(41, 2) = somsjours|
```

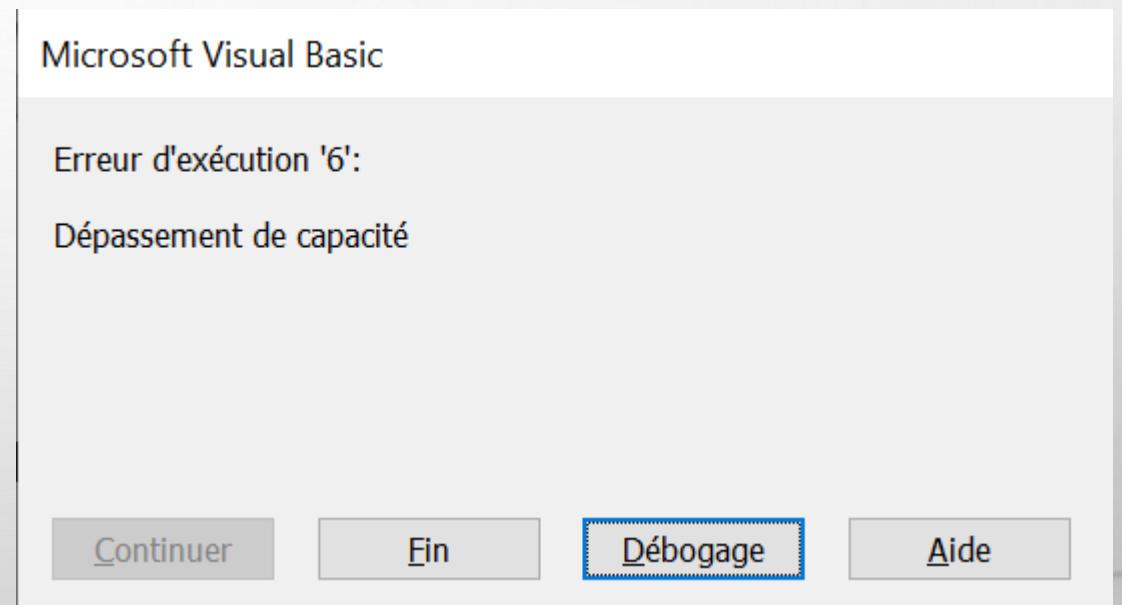
PROBLÈMES AVEC LES ENTIERS (1)

- ATTENTION SUR LES SOMMES D'ENTRIERS

LA SOMME DES JOURSJOURS APPROCHE DE LA LIMITE 32767 DES "INTEGER"

2,3	2209	
1,1	29868	
	-82793	

- ATTENTION AUX DÉPASSEMENTS DE CAPACITÉ



PROBLÈMES AVEC LES ENTIERS (2)

- ATTENTION SUR LES CALCULS D'ENTRIERS

LES CALCULS S'EFFECTUENT AVEC LES RÈGLES DE PRIORITÉS CLASSIQUES **DANS**
L'ENSEMBLE VISIBLE : SOMJOURS*SOMJOURS S'EFFECTUE DANS LA LIMITE DES "INTEGER"
UN DÉPASSEMENT DE CAPACITÉ EST SIGNALÉ ET LA LIGNE INCRIMINÉ EST POINTÉ

```
nb = Cells(1, 2)
⇒ dénominateur = nb * somjoursjours - (somjours * somjours)
Cells(43, 2) = dénominateur
! la porte
```

VERSION 005

QUELQUE SOIT LE NOMBRE DE RELEVÉS

- UTILISER UNE VARIABLE NBRELEVES
- POSITIONNER CETTE VARIABLE AU DÉBUT DE LA MACRO AVEC LA CELLULE B1
- REMPLACER TOUTES LES VALEURS QUI DÉPENDENT DU NB DE RELEVÉS PAR UNE EXPRESSION CONTENANT LA VARIABLE NBRELEVES : 39 DEVIENT NBRELEVÉS + 2

```
'Les Sommes
Cells(41, 1) = "Les sommes"
'    la somme des jours
somjours = 0
For i = 3 To 39
    jours = Cells(i, 2)
    somjours = somjours + jours
Next
Cells(41, 2) = somjours
```

```
nbreleves = Cells(1, 2)
```

```
'Les Sommes
Cells(nbreleves + 4, 1) = "Les sommes"
'    la somme des jours
somjours = 0
For i = 3 To nbreleves + 4
    jours = Cells(i, 2)
    somjours = somjours + jours
Next
```

TEST DE LA VERSION 005

- TESTER UNIQUEMENT AVEC 37 RELEVÉS NE DÉTECTE PAR L'OUBLI DU REMPLACEMENT D'UNE VALEUR
- LE TEST DE LA VERSION « QUELQUE SOIT LE NOMBRE DE RELEVÉS » DOIT ÊTRE EFFECTUÉ
- AVEC
 - 37 RELEVÉS
 - MOINS QUE 37 RELEVÉS
 - PLUS QUE 37 RELEVÉS

VERSION 006

QUELQUE SOIT LE NOMBRE DE RELEVÉS

CALCUL DU NB DE RELEVÉS

- LE CALCUL DU NB DE RELEVÉS EFFECTUÉ PAR PROGRAMME
 - AVANTAGE : LE « BON » PROGRAMME NE SE TROMPE PAS
 - PRENDRE LA DATE DE LA COLONNE 1 ET TANT QUE LA CELLULE N'EST PAS VIDE DANS CETTE COLONNE, PASSER À LA COLONNE DU DESSOUS, JUSQU'À CE QU'ELLE SOIT VIDE
 - ATTENTION, LA DATE EST DE TYPE STRING ET LA VARIABLE NE PEUT S'APPELER DATE CAR DATE EST UN MOT RÉSERVÉ

7	Dimanche 15 Juillet 2012	
8	Lundi 16 Juillet 2012	
9	Mardi 17 Juillet 2012	
0		
1		

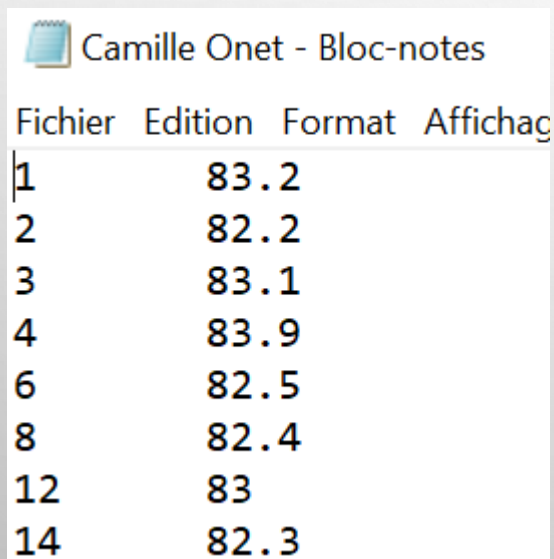
```
Dim datte As String
```

```
'Calcul du Nombre de relevés  
nbreleves = 0  
datte = Cells(nbreleves + 3, 1)  
Do While (datte <> "")  
    nbreleves = nbreleves + 1  
    datte = Cells(nbreleves + 3, 1)  
Loop
```

VERSION 007

RÉCUPÉRATION DE DONNÉES À PARTIR D'UN FICHIER

- LORSQUE L'OPÉRATEUR, LE PATIENT, SAISIT LES DONNÉES,
IL EST IMPORTANT DE PRÉCISER UN FORMAT DE SAISIE



Fichier	Edition	Format	Affichage
1	83.2		
2	82.2		
3	83.1		
4	83.9		
6	82.5		
8	82.4		
12	83		
14	82.3		

```
Dim filenum As Integer
filenum = FreeFile()
i = 3
Open "S:\internet\intranet\Base de Progr
Do While Not EOF(filenum)
    Input #filenum, jours
    Cells(i, 2) = jours
    Input #filenum, poids
    Cells(i, 3) = poids
    i = i + 1
Loop

'Affichage du nombre de relevés
nbreleves = i - 3
Cells(1, 1) = "NOMBRE D'EXPERIENCE"
Cells(1, 2) = nbreleves
```

CONCLUSION

- APRÈS CES QUELQUES RUDIMENTS : LA NOTION DE VARIABLE
- MÉTHODOLOGIE
 - PRATIQUER, ESSAYER ET SURTOUT TESTER ...
 - NE PAS HÉSITER À ÉCRIRE DES PETITES MACROS À COTÉ DE L'APPLICATION DÉVELOPPÉE POUR SE RASSURER ET TESTER DES CONCEPTS, NOTIONS, FONCTIONNALITÉS, ... MODÈLE EN Y
 - AVEC UN TABLEUR, NE PAS HÉSITER À ENREGISTRER DES MACROS POUR RÉCUPÉRER DES CODES GÉNÉRÉS ... POUR UN ASPECT VISUAL

Excel, un outil adapté pour la mise en forme et le contrôle qualité des données

Cas d'étude : Données collectées pour modéliser l'abondance

Florèn HUGON, Docteur en Physiologie et Biologie des Organismes

Laboratoire de Mathématiques et de leurs Applications de Pau, CNRS, UMR 5142

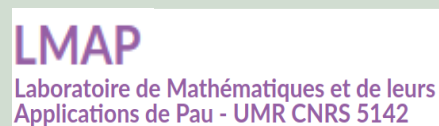
Xavier NAVARRO, Maître de Conférence, Laboratoire d'Informatique de l'Université de Pau et Pays Adour, CNRS

Matt RODRIGUEZ & Gabriel DAS NEVES, Étudiant Licence Informatique pour le Vivant

Matthieu BERRONEAU, Chargé d'Études Cistude Nature

Frank D'AMICO, Maître de Conférence, Laboratoire de Mathématiques et de leurs Applications de Pau, CNRS, UMR 5142

Mardi 31 mai 2022



Contexte – État des lieux des données à analyser

Objectif : Évaluer l'abondance des têtards de Grenouille



7 torrents découpés en 4 à 28 sites → 97 sites

Suivis annuels lancés en 2017

Problématiques

Besoin d'un tableau ligne = site et colonne = variable MAIS ...

Une feuille comptage et covariables d'échantillonnage + Une covariables de site

AspeEstl_Aval_2019.xlsx - LibreOffice Calc							
A1 f_x Σ = Tronçon n° :							
A	B	C	D	E	F	G	H
1	Tronçon n° :	1					
2							
3	Passage	1	2	3			
4	Observateur	Valentin Lasserre	Valentin Lasserre	Valentin Lasserre			
5	Date	29/04/2019	04/06/2019	05/07/2019			
6	Heure	Début	16h03	10h32	9h43		
7		Fin	16h12	10h37	9h51		
8	Météo et état du cours d'eau	T°	15,7	17,1	18,3		
9		T° 7 jours	17,57	25,18	27,71		
10		T° eau	13,3	13,5	16,1		
11		Nébulosité	2	2	1		
12		Pluvio.	non	non	non		
13		Pluvio. 7 jours	0mm	4mm	1,2mm		
14		Profondeur max.	18	21	17		
15		Turbidité max.					
16		Débit	faible	faible	faible		
17	Contact G. des Pyrénées	Nbre têtards	236	120	32		
18		Nbre groupes	2	1	3		
19		Nbre adultes	0	0	0		
20		Nbre pontes	0	0	0		

AspeEstl_Aval_2019.xlsx - LibreOffice Calc							
A1 f_x Σ = Tronçon n°							
A	B	C	D	E	F	G	H
1	Tronçon n°	1	Observateur :	Valentin Lasserre	Année :	2019	
2		N-O			N-E		
3		S-O			S-E		
4							
5	Aide localisation :						
6							
7							
8							
9							
10							
11	Altitude (m)			539			
12	Largeur max. (m)			4			
13	Largeur min. (m)			1,5			
14	Dénivelé (m)			0,4			
15	Exposition						
16	Nombre de vasques			2			
17	Diamètre de la plus grande vasque			400 cm			
18	Substrat / Granulométrie			roche nue, gravier, sable			
19	Distance source						
20	Présence connue de <i>P. leniusculus</i> dans le ruisseau			non			
21	Hiver précédent		Température	moyenne : 8,17 / min : -5,2 / max : 25,8			
22			Pluviométrie	348,5 mm			
23	Boisement feuillu			x			
24	Boisement résineux						
25	Boisement mixte						

Problématiques

Besoin d'un tableau ligne = site et colonne = variable MAIS ...

Une feuille comptage et covariables d'échantillonnage + Une covariables de site

AspeEstl_Aval_2019.xlsx - LibreOffice Calc							
A1 f_x Σ = Tronçon n° :							
A	B	C	D	E	F	G	H
1	Tronçon n° :	1					
2							
3	Passage	1	2	3			
4	Observateur	Valentin Lasserre	Valentin Lasserre	Valentin Lasserre			
5	Date	29/04/2019	04/06/2019	05/07/2019			
6	Heure	Début	16h03	10h32	9h43		
7		Fin	16h12	10h37	9h51		
8	Météo et état du cours d'eau	T°	15,7	17,1	18,3		
9		T° 7 jours	17,57	25,18	27,71		
10		T° eau	13,3	13,5	16,1		
11		Nébulosité	2	2	1		
12		Pluvio.	non	non	non		
13		Pluvio. 7 jours	0mm				
14		Profondeur max.	18				
15		Turbidité max.					
16		Débit	faible				
17		Nbre têtards	236				
18	Contact G. des Pyrénées	Nbre groupes	2	1	3		
19		Nbre adultes	0	0	0		
20		Nbre pontes	0	0	0		
21							

AspeEstl_Aval_2019.xlsx - LibreOffice Calc

Automatiser la mise en forme avec un programme informatique

Problématiques

Besoin de données qualitatives

Est ce que toutes les informations ont été collectées ?

Est ce qu'elles sont collectées correctement ?

Est ce qu'elles sont cohérentes ?

	A	B	C	D	E	F	G	H
1	Tronçon n°		1	Observateur :	Valentin Lasserre	Année :	2019	
2			N-O			N-E		
3			S-O			S-E		
4								
5				Aide localisation :				
6								
7								
8								
9								
10								
11				Altitude (m)		539		
12				Largeur max. (m)		4		
13				Largeur min. (m)		1,5		
14				Dénivelé (m)		0,4		
15				Exposition				
16				Nombre de vasques		2		
17				Diamètre de la plus grande vasque		400 cm		
18				Substrat / Granulométrie		roche nue, gravier, sable		
19				Distance source				
20				Présence connue de <i>P. leniusculus</i> dans le ruisseau		non		
21				Hiver précédent	Température	moyenne : 8,17 / min : -5,2 / max : 25,8		
22					Pluviométrie	348,5 mm		
23				Boisement feuillu		x		
24				Boisement résineux				
25				Boisement mixte				

Problématiques

Besoin de données qualitatives

Est ce que toutes les informations ont été collectées ?

Est ce qu'elles sont collectées correctement ?

Est ce qu'elles sont cohérentes ?

AspeEsti_Aval_2019.xlsx - LibreOffice Calc

A1 f_x Σ = Tronçon n°

	A	B	C	D	E	F	G	H
1	Tronçon n°		1	Observateur :	Valentin Lasserre		Année :	2019
2			N-O			N-E		
3			S-O			S-E		
4								
5				Aide localisation :				
6								
7								
8								
9								
10								
11			Altitude (m)			539		
12			Largeur max. (m)			4		
13			Largeur min. (m)			1,5		
14			Dénivelé (m)			0,4		
15			Exposition					
16			Nombre de vasques			2		
17			Diamètre de la plus grande vasque			400 cm		
18			Substrat / Granulométrie			roche nue, gravier, sable		
19			Distance					
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								
30								
31								
32								
33								
34								
35								
36								
37								
38								
39								
40								
41								
42								
43								
44								
45								
46								
47								
48								
49								
50								
51								
52								
53								
54								
55								
56								
57								
58								
59								
60								
61								
62								
63								
64								
65								
66								
67								
68								
69								
70								
71								
72								
73								
74								
75								
76								
77								
78								
79								
80								
81								
82								
83								
84								
85								
86								
87								
88								
89								
90								
91								
92								
93								
94								
95								
96								
97								
98								
99								
100								

Effectuer un contrôle qualité avec un programme informatique

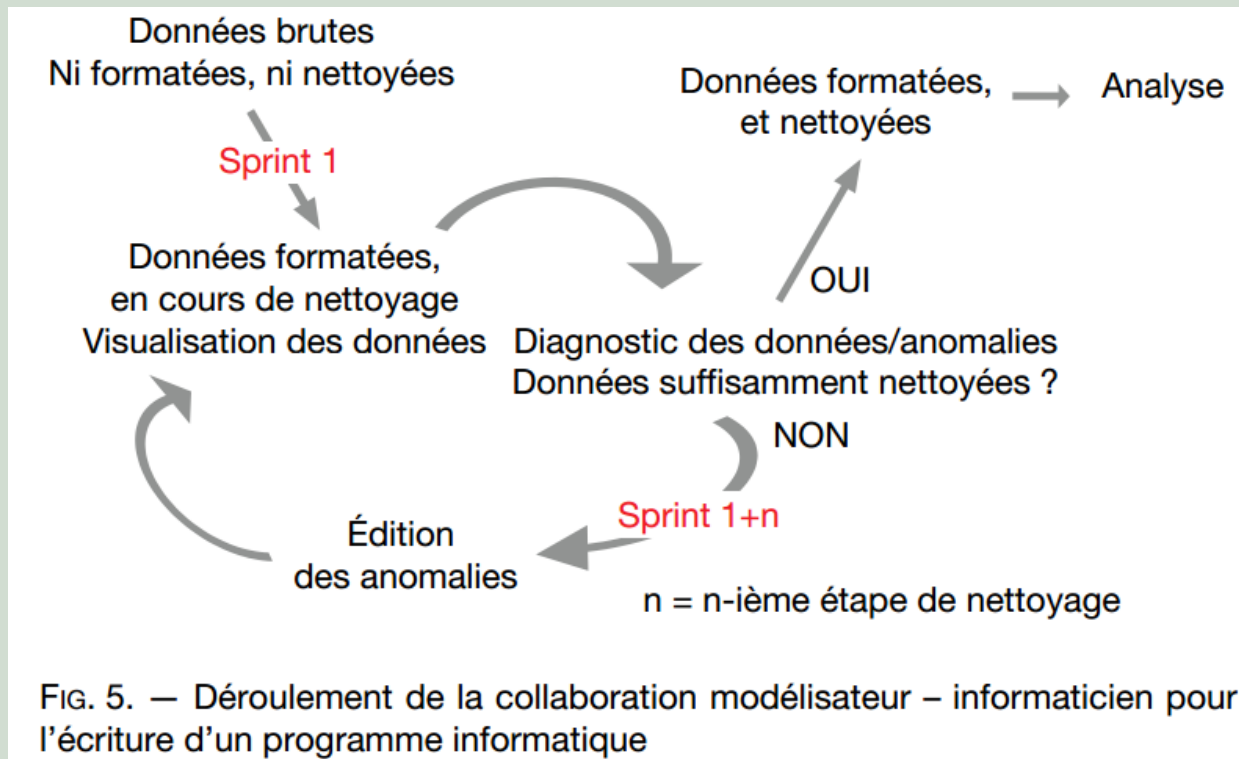
Initialisation du projet informatique

1. Rédaction du projet

2. Échanges entre informaticien et écologue

Besoins de l'informaticien pour réaliser le projet ? Livrables attendus par l'écologue ?

Méthode Agile



Facilité de mise en œuvre

Adaptabilité

Compréhension par étape

Déroulement par sprints

Tableau de critères, rédigé par l'écologue pour l'informaticien

Sprint	Temps dédié	Description
1	2 jours	Copie des données du fichier brut dans le fichier formaté, permet de détecter et lister des anomalies par le test du programme
2	1/2 jour	Vérification de la présence de la donnée
3	1/2 jour	Vérification des ordres de grandeurs des variables quantitatives
4	1/2 jour	Traduction des variables qualitatives en variable quantitative
5	1/2 jour	Vérification de la qualité de certaines variables par rapport à d'autres, largeur maximale > largeur minimale, somme des pourcentages égale à 100, etc.
6	1/2 jour	Traitement spécifique à certaines variables, agrégation de données brutes en une unique variable pour les variables relatives aux présences d'espèces prédatrices, gestion des unités, etc.

Variables	Critère
ObsV1-V2-V3	Nbr positif, sinon écrire en rouge
TairV1-V2-V3	De 0 à 40, sinon écrire en rouge
TeauV1-V2-V3	De 0 à 30, sinon écrire en rouge
CloudV1-V2-V3	Valeur = 0, 1, 2, 3, 4 ou 5, sinon écrire en rouge
DepthV1-V2-V3	Nbr positif, sinon écrire en rouge
FlowV1-V2-V3	Expliquer « faible » = 1, « moyen » = 2, « fort » = 3. Écrire « Faible », « Moyen », « Fort » est aussi accepté. Si autre mot, écrire en rouge
Alti	Nbr positif, sinon écrire en rouge
LargMax, LargMin	Nbr positif pour les deux, ET Vérifier que LargMin <= LargMax, sinon écrire les 2 en rouge
Drop	Nbr positif, sinon écrire en rouge
Expo	Pas de critère de cohérence pour cette variable
nBasins	Nbr positif, écrire en rouge
MaxSize	Nbr positif, écrire en rouge
SourceDist	Nbr positif, sinon écrire en rouge
<i>P.leniusculus</i>	Prendre l'information sur la feuille A, ligne 27 (pas sur la B) Si trois « 0 », indiquer 0. Si au moins une observation de l'espèce sur les trois visites (même si manque l'info sur une ou deux des cases), indiquer 1 (i.e. si somme des 3 cases >= 1). Si autre cas, écrire « error » en rouge
<i>S. trutta</i>	Si trois « A », indiquer 0. Si au moins un « P » sur les trois visites (même si manque l'info sur une ou deux des cases), indiquer 1. Si autre cas, écrire « error » en rouge
<i>N.maura</i>	Si trois « A », indiquer 0. Si au moins un « P » sur les trois visites (même si manque l'info sur une ou deux des cases), indiquer 1. Si autre cas, écrire « error » en rouge
Farming	Expliquer « non » = 0, « oui » = 1. Écrire « Non », « Oui » est aussi accepté. Si autre mot, écrire en rouge
Moss, Herb, Shrub, Trees	Chacun de 0 à 100, sinon écrire en rouge ET Vérifier que somme des 4 = 100, sinon écrire les 4 en rouge (si somme des 4 = 100 même avec une donnée manquante alors c'est ok quand même)

Sur toutes les variables sauf espèces. Si absence de données, écrire NA.

Premier test, ai je une donnée ? Non → « NA » / Oui → deuxième test pour vérifier sa cohérence selon le tableau de critères. Si incohérence détectée à l'aide des critères, écrire la donnée en rouge

Sur les variables « espèces » (*P. leniusculus*, *S. trutta* et *N. maura*), suivre scrupuleusement les indications du tableau.

Écriture du code informatique au fil des sprints

Exemple : code pour la variable Nombre de têtards observés à la visite 1, ObsV1

Sprint	Temps dédié	Description
1	2 jours	Copie des données du fichier brut dans le fichier formaté, permet de détecter et lister des anomalies par le test du programme
2	1/2 jour	Vérification de la présence de la donnée
3	1/2 jour	Vérification des ordres de grandeurs des variables quantitatives
4	1/2 jour	Traduction des variables qualitatives en variable quantitative
5	1/2 jour	Vérification de la qualité de certaines variables par rapport à d'autres, largeur maximale > largeur minimale, somme des pourcentages égale à 100, etc.
6	1/2 jour	Traitement spécifique à certaines variables, agrégation de données brutes en une unique variable pour les variables relatives aux présences d'espèces prédatrices, gestion des unités, etc.

Erreurs_Anomalie et Action_Anomalie

Fonctions qui permettent l'édition d'un fichier qui rend compte du traitement informatique.

Box : Traitement d'une variable

```
' Traitement de la variable métier, ObsV1 = Nombre de têtards observés à la visite 1
' Détection cellule vide équivalent à l'absence de la donnée
If Workbooks(fichierOrigine).Worksheets(nomOngletA).Range("C10").Value = "" Then
    ' Traitement de l'absence de données
    Workbooks(fichierDest).Worksheets("Feuil1").Cells(i + 1, 2) = NA
    'Compte-rendu de l'absence de données
    Erreurs_Anomalie Tborder, "ObsV1", "Integer", "C10", "B" + CStr(i + 1), CStr(i),
    nomOngletA, "vide"
    Action_Anomalie Tborder, "la valeur est un entier positif", "donnee absente donc NA"
Else
    ' Récupération de la donnée
    ObsV1 = Workbooks(fichierOrigine).Worksheets(nomOngletA).Range("C10").Value
    ' Détection anomalie
    If ObsV1 < 0 Then
        ' Traitement anomalie : mise en rouge de la cellule
        Workbooks(fichierDest).Worksheets("Feuil1").Range("B" + CStr(i + 1)).Font.Color =
        RGB(255, 0, 0)
        ' Compte-rendu de l'anomalie
        Erreurs_Anomalie Tborder, "ObsV1", "Integer", "C10", "B" + CStr(i + 1), CStr(i),
        nomOngletA, CStr(ObsV1)
        Action_Anomalie Tborder, "la valeur est un entier positif", CStr(ObsV1)
    End If
    Workbooks(fichierDest).Worksheets("Feuil1").Cells(i + 1, 2) = ObsV1
End If
```


Exécution du programme

	A	B	C	D	E		A	B	C	D	E
1						1					
2		Visite	1	2	3	2		Visite	1	2	3
3		Observateur	Prénom Nom	Prénom Nom	Prénom Nom	3		Observateur	Prénom Nom	Prénom Nom	Prénom Nom
4		Date	JJ/MM/AAAA	JJ/MM/AAAA	JJ/MM/AAAA	4		Date	JJ/MM/AAAA	JJ/MM/AAAA	JJ/MM/AAAA
5		Tair	15,7	17,1	18,3	5		Tair	17,2	17	18,5
6		Teau	13,3	13,5	16,1	6		Teau	13,2	13,5	16
7		Nébulosité	2	2	1	7		Nébulosité	3	4	3
8		Pluviométrie	non	non	non	8		Pluviométrie	non	non	non
9		Débit	faible	faible	faible	9		Débit	faible	moyen	faible
10		Nombre têtards	236	-120	32	10		Nombre têtards	1	0	
11		Nombre adultes	0	0	0	11		Nombre adultes	0	0	0
12		Nombre pontes	0	0	0	12		Nombre pontes	0	0	0
13		Nombre P.leniuculus	1	0	0	13		Nombre P.leniuculus	0	0	0
14		Salmo trutta, P/A	A	A	A	14		Salmo trutta, P/A	A	P	A
15		Natrix maura, P/A	A	A	A	15		Natrix maura, P/A	A	A	A
16		Commentaires				16		Commentaires			

Input : Fichier entrée

Programme
VBA Excel

Output : Fichiers sortie

FIG. 1. — Extrait du fichier de données brutes stockées dans plusieurs feuilles d'un même fichier tableur pour un torrent découpé en quatre sites d'étude. Mise en évidence de la variable « nombre de têtards observés », comptabilisés à chaque visite sur les sites 1 (onglet A1) et 2 (onglet A2). Les onglets A stockent le nombre d'observation et les covariables d'échantillonnage ; les onglets B, les covariables de site.

	A	B	C	D	E	F	G
1		ObsV1	ObsV2	ObsV3	TairV1	TairV2	TairV3
2	Site1	236	-120	32	15,7	17,1	18,3
3	Site2	1	0	NA	17,2	17	18,5
4	Site3	5	10	4	16,8	15,8	18,5
5	Site4	18	19	5	17,3	17,2	18,2

FIG. 2. — Extrait du fichier formaté, un tableur avec en ligne les sites d'étude et en colonne les variables étudiées. Les suffixes V1, V2, V3 indiquent la visite relative à chaque donnée. La valeur -120 ne respecte pas le critère de qualité de la variable « nombre de têtards observés », elle est donc indiquée en rouge. Il n'y avait pas de données de comptage pour le site 2 à la visite 3, la valeur indiquée est donc NA.

```
fichier origine : Nom_du_fichier_input_brut.xlsx
fichier de destination : Nom_du_fichier_output_formaté.xlsx

----- Site 1-----

Erreur : Variable ObsV2 (type Integer) (Position d'origine : Cellule D10
-> Position de destination : Cellule C2) du site 1 (Feuille A1) : -120
-> Conditions de cohérence : la valeur est un entier positif
-> Remplissage effectué : incohérence donc -120 en rouge gras

----- Site 2-----

Erreur : Variable ObsV3 (type Integer) (Position d'origine : Cellule E10
-> Position de destination : Cellule D3) du site 2 (Feuille A2) : vide
-> Conditions de cohérence : la valeur est un entier positif
-> Remplissage effectué : donnée absente donc NA
```

FIG. 3. — Extrait du fichier texte rendant compte des anomalies rencontrées.

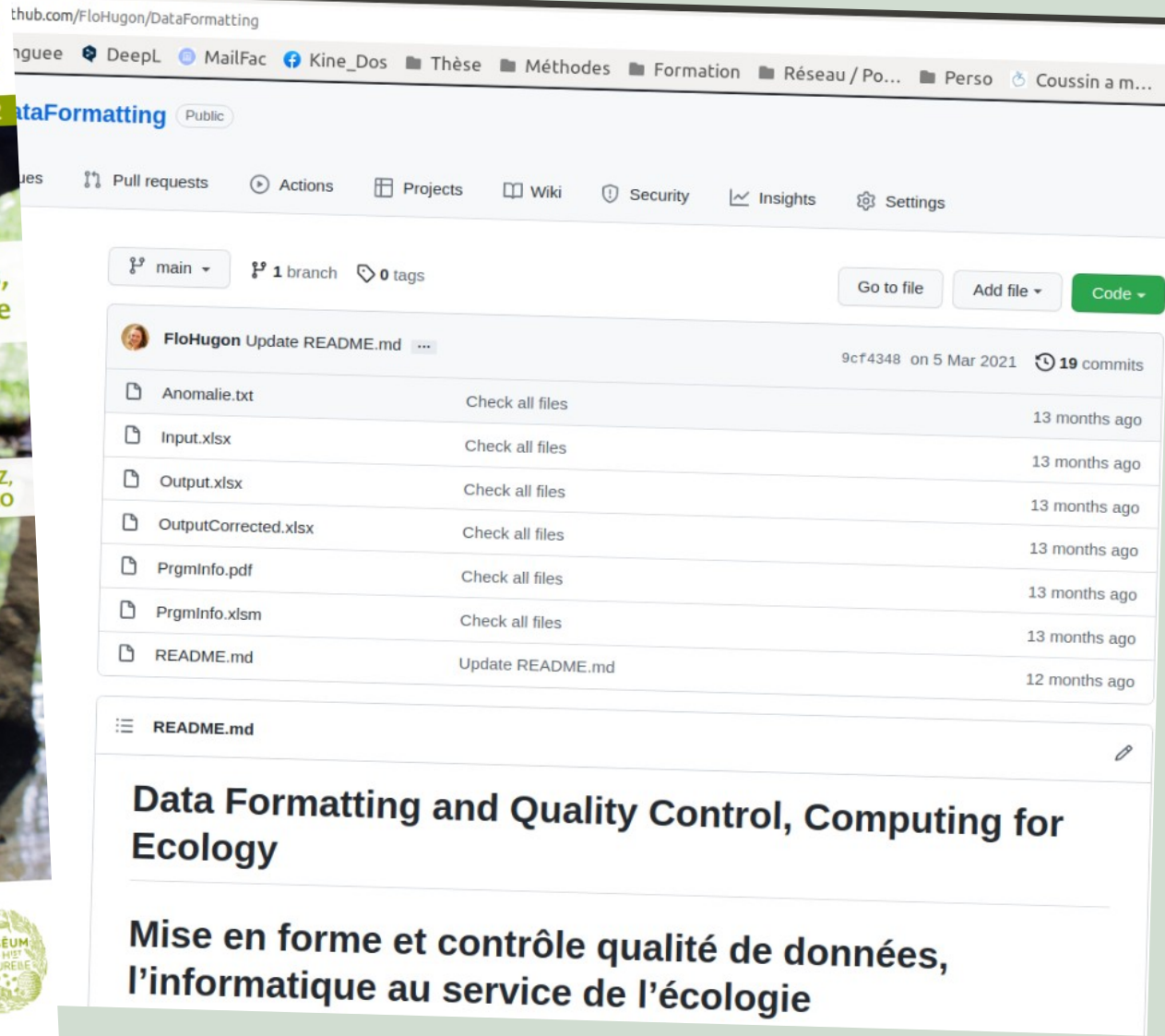
Démonstration : Exécution du programme informatique



https://youtu.be/RA2VQUC_A2s

CCL : Peu d'investissement pour beaucoup de résultat qualitatif

Valorisation des travaux



Perspectives

Ré-utilisation du programme informatique

Ajout de fonctionnalités

Adaptation des codes initiaux à d'autres besoins

Programmation pour des problématiques similaires sur demande

A l'écoute de vos questions



Florèn HUGON

floren.hugon@live.fr

En Recherche d'Opportunités, Salariat et/ou Prestations



<https://www.researchgate.net/profile/Floren-Hugon>



<https://www.linkedin.com/in/floren-hugon/>



<https://github.com/FloHugon>

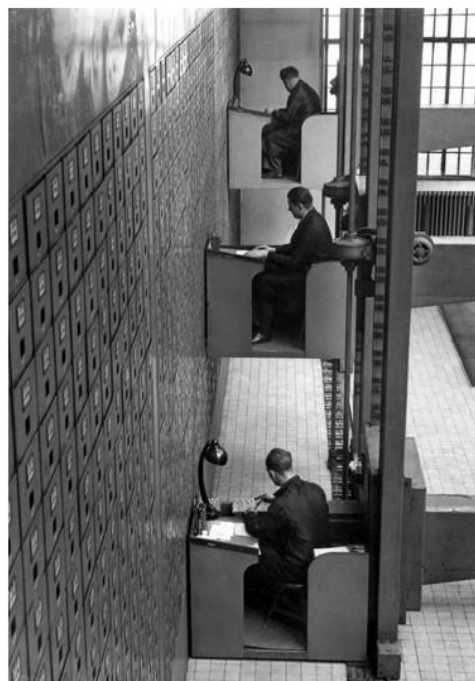
Un outil adapté à l'acquisition des données. Exemple sur des données de pêcheries pour implémentation dans une base de données

Webinaire du 31/5/2022

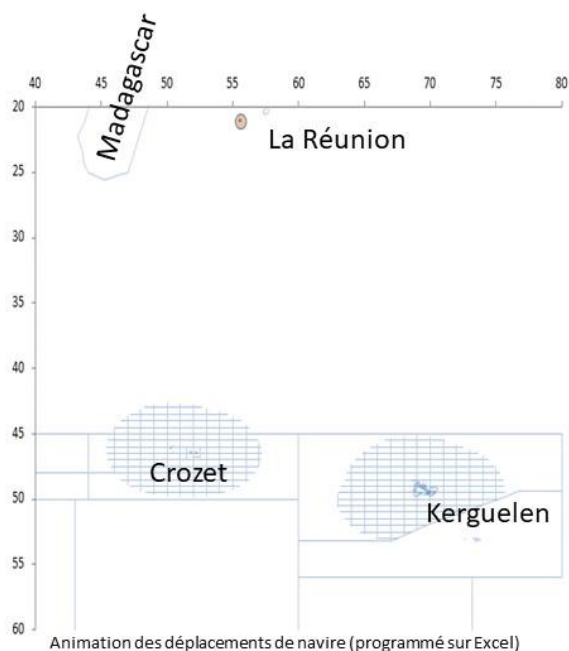
N. Gasco & C. Chazeau
MNHN



« vous n'imaginez pas tout ce qu'Excel peut faire pour vous! »



LE CONTEXTE



navires de pêche

25

campagnes en mer
(de trois mois) par an

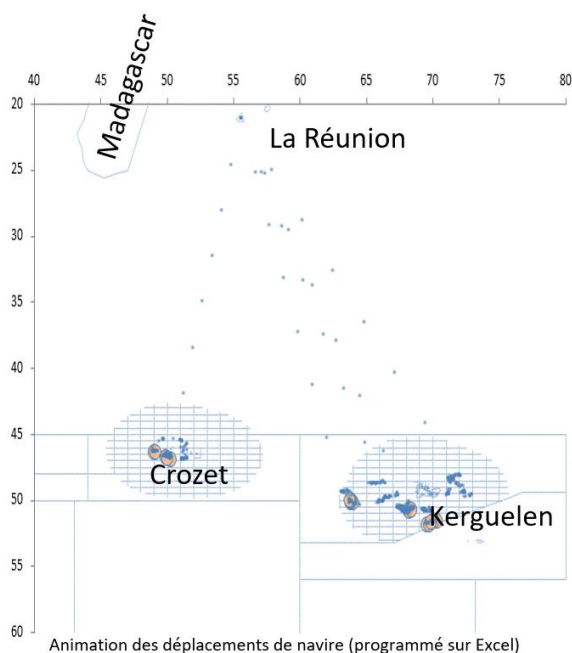


Une pêche à l'hameçon



« légine australe »

➔ LE CONTEXTE



100% de couverture par un observateur scientifique

Par an:



250 000 biométries



6 000 marquages



3 000 échantillons

4000

Stations
(30 000 lat long
prof)



PROGRAMME D'OBSERVATION ECOSYSTEMIQUE DES PECHERIES AUSTRALES³

➔ LE CONTEXTE



2,5 millions de cellules Excel / an



Base de donnée « PECHEKER »



Analyses et modélisations



Avis scientifique destiné au gestionnaire
(TAAF)



PROGRAMME D'OBSERVATION ECOSYSTEMIQUE DES PECHERIES AUSTRALES⁴

La collecte des données sur le terrain....



- Faciliter au maximum la vie de l'utilisateur sur le terrain et lui faire gagner du temps pour qu'il puisse se concentrer sur la collecte des données.
- Lui fournir tous les outils nécessaires pour détecter les erreurs et les corriger en temps réel (après le retour du bateau c'est souvent trop tard) de façon à obtenir un jeu de données fiable et de « haute qualité » prêts à être chargés en base pour les analyses.
- Une récolte de données compatible avec les standards internationaux



- Aujourd'hui une présentation de quelques possibilités d'excel à travers un exemple de fichier de terrain
- Un développement au fil de l'eau depuis une vingtaine d'années sans formation en informatique



➤ LE FICHIER DE SAISIE: « Carnet de Pêche »

➤ CONTRAINDRE LA SAISIE

Exemple sur une variable numérique décimale:

longueur	
15,2	Ok
15,,2	Erreur de virgule
assez grand mais j'ai perdu ma règle	inutilisable

➤ LE FICHIER DE SAISIE: « Carnet de Pêche »

➤ CONTRAINDRE LA SAISIE

Données Révision Affichage Aide Dites-nous ce que vous voulez faire

Tri et filtrer: Trier, Filtrer, Effacer, Réappliquer, Avancé

Outils de données: Convertir Remplissage instantané, Supprimer les doublons, Validation des données, Consolider Relations, Gérer le modèle de données, Analyse scénario

longueur

➤ LE FICHIER DE SAISIE: « Carnet de Pêche »

➤ CONTRAINDRE LA SAISIE

Exemple sur une variable texte:

espèce	
Bathylutichthys taranetzi	➡ Ok
Bathylutichthis taranetzi	➡ Faute de frappe
Bathilutichthis taranetzi	➡ Faute de frappe

➤ LE FICHIER DE SAISIE: « Carnet de Pêche »

➤ CONTRAINDRE LA SAISIE

Données Révision Affichage Aide Dites-nous ce que vous voulez faire

Trier et filtrer Trier Filtrer Réappliquer Avancé

Outils de données Convertir Remplissage Supprimer Validation des Consolidier Relations Gérer le modèle Analyser scénario

G H I J K L

ma liste d'espèce:

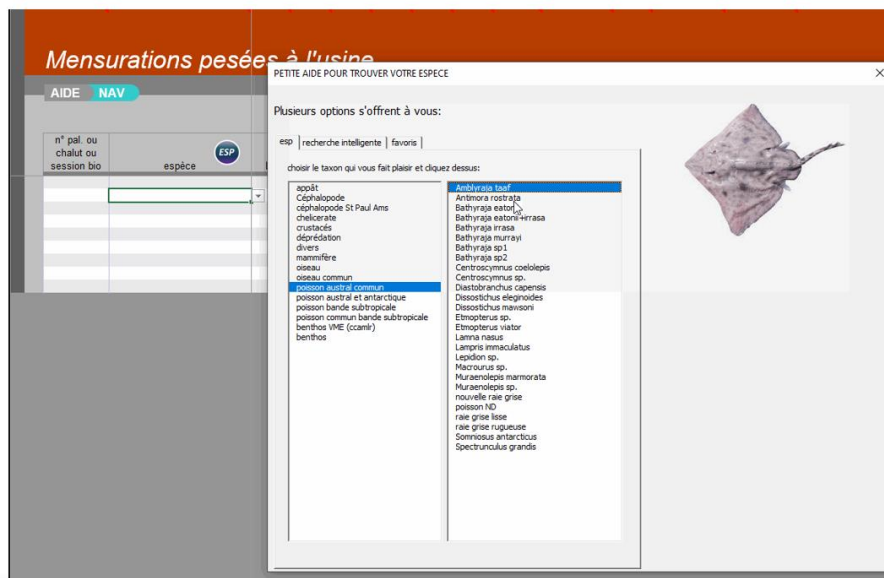
- Bathylagichthys australis
- Bathylagus antarcticus
- Bathylagus gracilis
- Bathylagus niger
- Bathylagus tenuis
- Bathylutichthys taranetzi
- Bathyonus pectoralis
- Bathypterois oddi

espèce

Bathylagichthys australis
Bathylagus antarcticus
Bathylagus gracilis
Bathylagus niger
Bathylagus tenuis
Bathylutichthys taranetzi
Bathyonus pectoralis
Bathypterois oddi

LE FICHIER DE SAISIE: « Carnet de Pêche »

FACILITER LA SAISIE



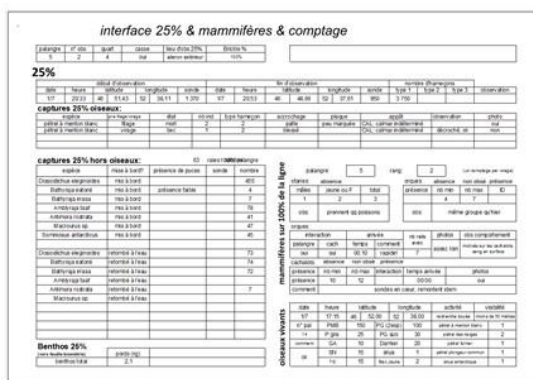
Exemple sur le choix d'une espèce:

- Plus rapide que le menu déroulant
- Permet de vérifier sur l'image

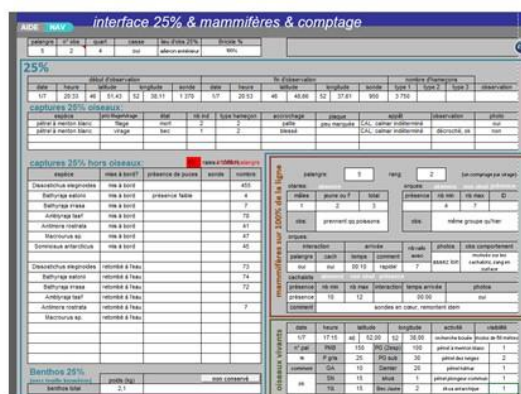
LE FICHIER DE SAISIE: « Carnet de Pêche »

FACILITER LA SAISIE

Feuille papier A4 pour noter au crayon



INTERFACE DU CARNET DE PECHE EXCEL:



Saisie à l'identique

LE FICHIER DE SAISIE: « Carnet de Pêche »

FACILITER LA SAISIE

INTERFACE DU CARNET DE PECHE EXCEL:

Une répartition dans toutes les tables en un clic

Observation des 25%

AIDE	NAV	date observation	latitude	longitude	no d'observation
n° palangre	n° site	date heure	deg	min	deg
1	1	ven 220422 07 40	49	25 55	67
2	1	ven 220422 11 45	49	29 53	67
3	1	ven 220422 13 45	49	29 25	67
4	1	ven 220422 15 23	49	29 36	67
5	1	sam 230422 11 37	49	36 40	67

Oiseaux observés pendant les 25%

AIDE	NAV	CP	no au large	no en mer	no en port
n° palangre	n° site	espèce	nombre	nombre	nombre

CP virage des palangres

AIDE	NAV	CP	date	heure	nombre	nombre	nombre	observation
n° palangre	RANG	date heure	nombre	nombre	nombre	nombre	observation	
1	1	ven 220422 07 31	0	0	0	0		
2	1	ven 220422 11 31	0	0	0	0		
3	1	ven 220422 13 47	0	0	0	0		
4	1	sam 230422 01 07	0	0	0	0		
5	1	sam 230422 11 37	0	0	0	0		

Comptage des oiseaux vivants

AIDE	NAV	CP	référence du comptage	espèce	nombre	commentaire
1	1	palangre à moulin blanc	4			
1	1	palangre gris	2			
1	1	grand albatros	80			
1	1	albatros à ailes noires	40			
1	1	albatros à tête grise	100			
1	1	palangre gris deux espèces conf	4			
1	1	palangre de Wilson	100			
1	1	palangre à moulin blanc	10			

Obs oiseaux vivants

AIDE	NAV	CP	date	heure	latitude	longitude	no de comptage	no de site	n° pal
1	1	ven 220422 07 31	49	25 55	67	11 31	1	1	
2	1	ven 220422 11 31	49	29 53	67	11 31	1	1	
3	1	ven 220422 13 47	49	29 25	67	11 31	1	1	
4	1	sam 230422 01 07	49	29 36	67	11 31	1	1	
5	1	sam 230422 11 37	49	36 40	67	11 31	1	1	

Observation des captures (hors oiseaux) pendant les 25%

AIDE	NAV	CP	n° palangre	n° site	espèce	no de capture	no de site	n° pal
1	1	ven 220422 07 31	49	25 55	67	11 31	1	1
2	1	ven 220422 11 31	49	29 53	67	11 31	1	1
3	1	ven 220422 13 47	49	29 25	67	11 31	1	1
4	1	sam 230422 01 07	49	29 36	67	11 31	1	1
5	1	sam 230422 11 37	49	36 40	67	11 31	1	1

LE FICHIER DE SAISIE: « Carnet de Pêche »

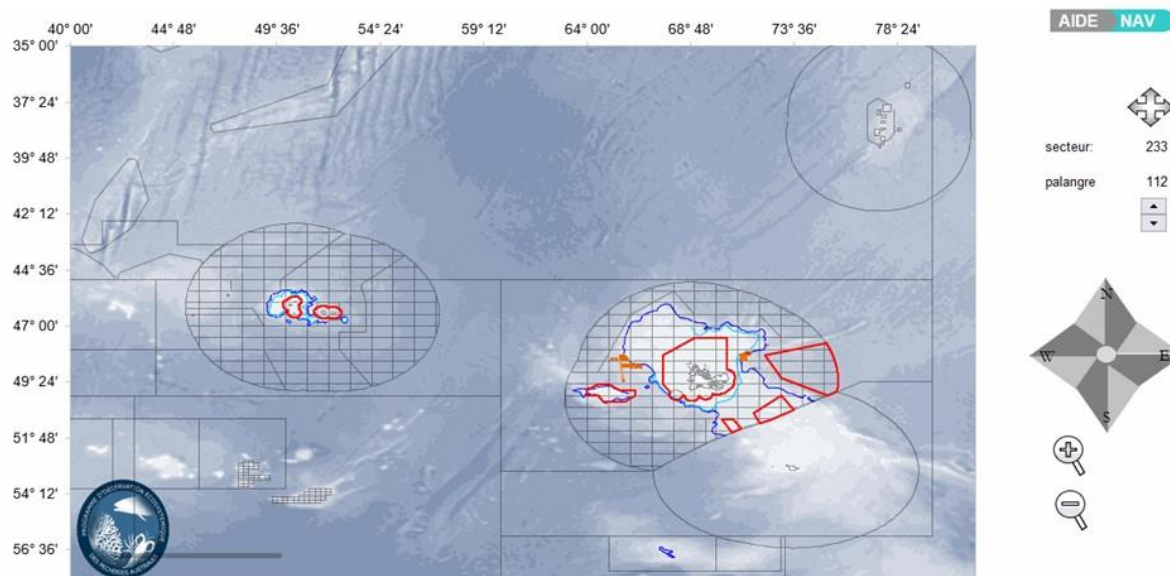
Ce serait sympa de visualiser les stations sur une carte...

- Pour vérifier les positions
- Pour s'organiser
- Pour mettre dans un rapport en fin de campagne

Exporter vers R ? Qgis ?...

LE FICHIER DE SAISIE: « Carnet de Pêche »

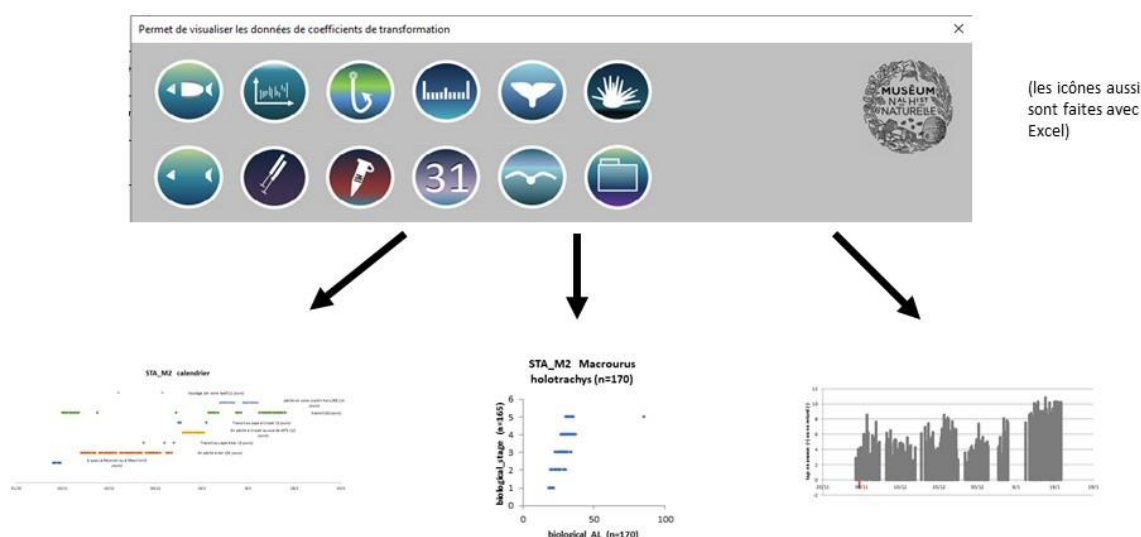
VISUALISER SES STATIONS EN RESTANT SUR EXCEL



zoom avant / arrière, zoom sur une station, déplacement, fond de carte en image (données fictives)

LE FICHIER DE SAISIE: « Carnet de Pêche »

VISUALISER LES DONNEES FACILEMENT



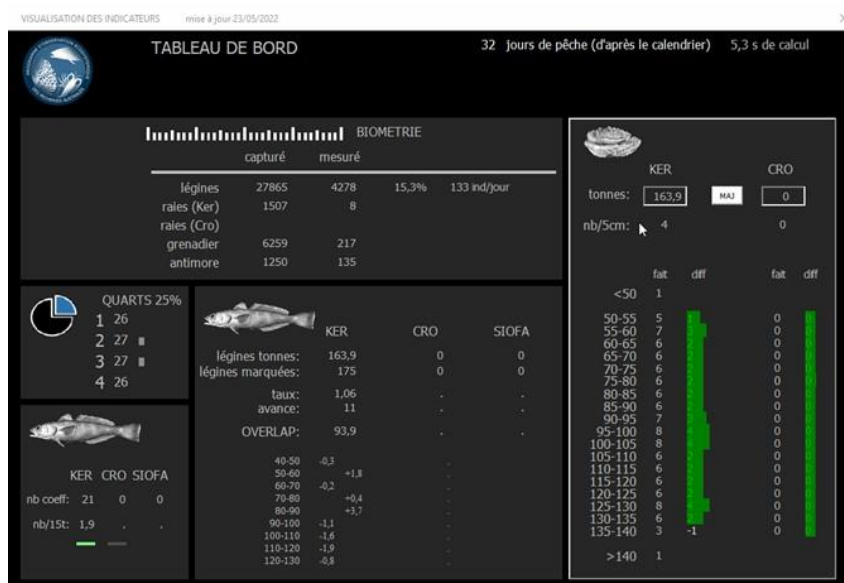
(les icônes aussi sont faites avec Excel)

150 graphiques générés à la demande en quelques minutes.

➤ LE FICHIER DE SAISIE: « Carnet de Pêche »

➤ VISUALISER LES DONNEES PRINCIPALES RAPIDEMENT

5 secondes de calcul pour visualiser tous les indicateurs clés



- Biométrie réalisée par espèce
- Échantillonnage stratifié de pièces osseuses
- Marquage
- Permet à l'agent sur le terrain de s'organiser plus facilement

➤ L'add-in de détection des anomalies

➤ La « table de vérification »

Exemple 1:

Variable 1	signe	valeur	Action déclenchée
Longueur totale	>	250	Le poisson est trop grand

Exemple 2:

Variable 1	signe	valeur	Action déclenchée
Vitesse du navire	>	11	Le bateau va trop vite

La table comporte plus de 960 lignes composées chacune de 1 à 5 conditions. Le programme VBA va « lire » cette table et vérifier toutes les conditions pour tout le jeu de données.

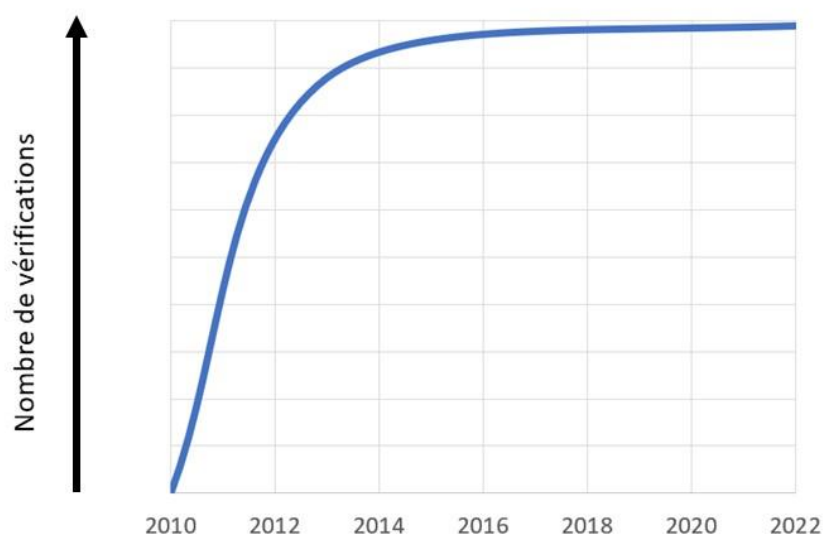
L'add-in de détection des anomalies

TABLE DES VERIFICATIONS (cette table contient toutes les vérifications possibles pour le carnet de pêche 13/05/2022 14:51)											
qu'est ce qui ne va pas?				qu'est ce qu'il faut faire				1		2	
rubrique	feuille	verif	niveau teste, si	action, si	action, en	champ	opérateur	valeur	champ	opérateur	valeur
25_pourcent	25_captures	1071	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_captures_LL_number	> 0	tally_captures_tanded	*	*
25_pourcent	25_captures	1072	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_captures_LL_number	> 0	tally_captures_number	*	*
25_pourcent	25_obs	1073	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_LL_number	*	*
25_pourcent	25_obs	1074	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_obs_number	*	*
25_pourcent	25_obs	1075	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_quota	*	*
25_pourcent	25_obs	1076	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_start_lat_deg	*	*
25_pourcent	25_obs	1077	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_start_lat_min	*	*
25_pourcent	25_obs	1078	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_start_long_deg	*	*
25_pourcent	25_obs	1079	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_start_long_min	*	*
25_pourcent	25_obs	1080	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_start_depth	*	*
25_pourcent	25_obs	1081	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_end_date_time	*	*
25_pourcent	25_obs	1082	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_end_lat_deg	*	*
25_pourcent	25_obs	1083	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_end_lat_min	*	*
25_pourcent	25_obs	1084	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_end_long_deg	*	*
25_pourcent	25_obs	1085	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_end_long_min	*	*
25_pourcent	25_obs	1086	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_end_depth	*	*
25_pourcent	25_obs	1087	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_hook_type	*	*
25_pourcent	25_obs	1088	3	ce champ doit être rempli	remplir le champ, voir avec	palangre	tally_obs_start_date_time	> 0	tally_obs_obs_location	*	*
25_pourcent	25_obs	1089	3	ce champ doit être rempli	remplir le champ, voir avec le MNBH si possible	palangre	ship_name	(*) CK2	ship_name	(*)	CK2
25_pourcent	25_obs	1125	3	les rangs d'observation ne se suivent pas		palangre	tally_obs_LL_number	> 0	tally_obs_obs_number	> 0	
25_pourcent	25_obs	1175	3	les obs 25% doivent être comprises dans une	corriger l'horaire, sauf si c'est juste que vous n'avez	palangre	CAL_age_setting_25_obs	> 2.5	tally_obs_start_date_time	> 0	
25_pourcent	25_obs	1176	3	les obs 25% doivent être comprises dans une	corriger l'horaire, sauf si c'est juste que vous n'avez	palangre	CAL_age_setting_25_obs	> 2.5	tally_obs_end_date_time	> 0	
25_pourcent	25_obs	1182	3	le nombre d'hampons observés ne peut être	corriger (sauf si c'est trop récent et que vous n'avez	palangre	CAL_age_setting_25_obs	> 2.5	tally_obs_hook_type	> 0	
25_pourcent	25_captures	1193	2	cette espèce n'est pas courante dans cette zone	envoyer une photo au MNBH pour confirmation,	biogéogr	tally_captures_species	*	Elmopterus violator	CAL_setting_area_of_tally	(*) ZEEKER
25_pourcent	25_captures	1194	2	cette espèce n'est pas courante dans cette zone	envoyer une photo au MNBH pour confirmation,	biogéogr	tally_captures_species	*	Elmopterus sp.	CAL_setting_area_of_tally	(*) ZEEKER
25_pourcent	25_captures	1195	2	cette espèce n'est pas courante dans cette zone	envoyer une photo au MNBH pour confirmation,	biogéogr	tally_captures_species	*	Elmopterus violator	CAL_setting_area_of_tally	(*) ZEEKER
25_pourcent	25_captures	1196	2	cette espèce n'est pas courante dans cette zone	envoyer une photo au MNBH pour confirmation,	biogéogr	tally_captures_species	*	Bathyrhynchus	CAL_setting_area_of_tally	(*) ZEEKER
25_pourcent	25_captures	1197	2	cette espèce n'est pas courante dans cette zone	envoyer une photo au MNBH pour confirmation,	biogéogr	tally_captures_species	*	Bathyrhynchus	CAL_setting_area_of_tally	(*) ZEEKER
25_pourcent	25_captures	1198	2	cette espèce n'est pas courante dans cette zone	envoyer une photo au MNBH pour confirmation,	biogéogr	tally_captures_species	*	Bathyrhynchus	CAL_setting_area_of_tally	(*) ZEEKER
25_pourcent	25_captures	1199	2	cette espèce n'est pas courante dans cette zone	envoyer une photo au MNBH pour confirmation,	biogéogr	tally_captures_species	*	Ambrystus	CAL_setting_area_of_tally	(*) ZEEKER
25_pourcent	25_captures	1223	2	la combinaison palmi oblongue n'est pas	vérifier que les données n'ont pas été rajoutées	palangre	tally_captures_LL_number	> 0	CAL_tally_capture_occurrence_combo	> 1	
25_pourcent	25_obs	1236	2	normalement vous avez passé un minimum de	corriger l'horaire ou si l'obs est si courte voir avec le	palangre	tally_obs_obs_number	> 0	CAL_temps_obs_25	> 0	
25_pourcent	25_obs	1237	2	l'observation est négative	corriger les horaires	palangre	tally_obs_obs_number	> 0	CAL_temps_obs_25	> 0	
25_pourcent	25_obs	1215	3	le bateau va généralement à moins de	revol les dates, heures et positions avec le bord et	palangre	tally_obs_LL_number	> 0	CAL_speed_tally_next	contient	vitesse élevée
25_pourcent	25_obs	1216	3	les obs 25% sont	revol vos horaires d'obs	palangre	tally_obs_LL_number	> 0	CAL_temps_obs_25	> 0	300

L'add-in de détection des anomalies

date	ref de la vérif	feuille	ligne	colonne	qu'est ce qui ne va pas?	qu'est ce qu'il faut faire?
Vérification unique du numéro 1152						
23/05/22 11:07 FILA	----- checking 1152 -----	CP_filage	52		vitesse de filage en KT: 49.68 > 11	le filage ne peut dépasser une certaine vitesse cal positions vérifier avec le capitaine et corriger ou confirmer le r valeur reste confirmée
temps de calcul: 5,82 s						

➔ L'add-in de détection des anomalies



Pratiquement plus aucune erreur ne passe à travers la vérification

➔ Les images: des données à part entière

Des milliers d'images par an.

Mise en place d'un standard de renommage
(saison / navire / auteur / station / espèce...)

 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00610	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00653
 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00611	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00654
 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00612	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00655
 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00613	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00656
 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00614	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00657
 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00615	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00658
 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00616	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00659
 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00617	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00660
 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00618	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00661
 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00619	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00662
 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00620	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00663
 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00621	 2021 ID2 21 03 CAQI P_061 SPW ZEE KER 00664

Un outils excel pour faciliter ce renommage

➤ Les images: des données à part entière

Carnet de pêche:

Explorateur windows:

PHOTOS à renommer

PHOTOS renommées

DCIM₀ (T)DCM₂ (2)

DCIM_03

Le dernier est vide.

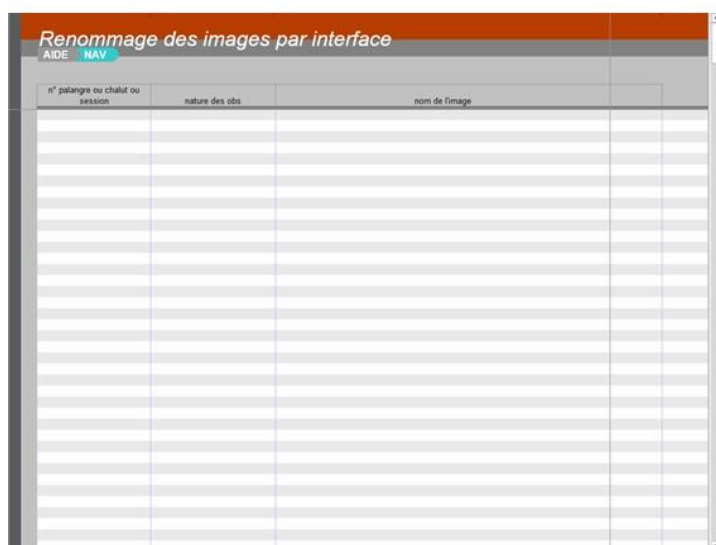
➤ Les images: des données à part entière

Carnet de pêche:

Explorateur windows:

PHOTOS à renommer

PHOTOS renommées

DCIM₁(T)

DCIM (2)



DCIM_ (3)

Le dossier est vide.

➤ Les images: des données à part entière

Carnet de pêche:

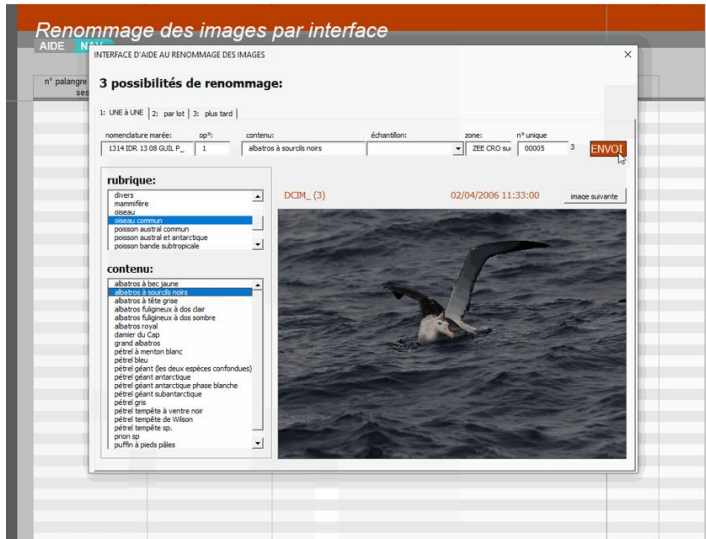
Explorateur windows:

PHOTOS

à renommer

PHOTOS

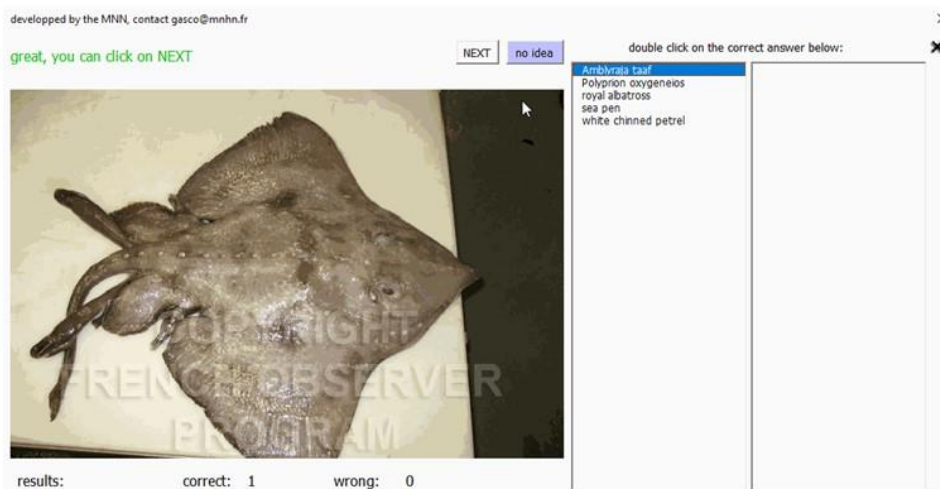
renommées



➤ LA QUALITE EN AMONT DU TERRAIN

Les agents doivent pouvoir identifier toutes les espèces courantes sur le terrain

Entraînement à l'identification des espèces courantes avant de partir:



<https://www.ccamlr.org/fr/node/92048>

➤ LA QUALITE EN AVAL...



Vérification des identifications faites sur le terrain.

- Les agents font des photos sur le terrain et renomment la photo avec le nom de l'espèce.
- Vérification des identifications au retour
- Bilan personnalisé pour éviter les erreurs la fois suivante.

checking identification

image name: 2021 02 20 10 430 P_000 albatros à bec jaune 00006

checking name: 2021 ID 20 10 KERV P_000

image new name:

species reported: albatros à bec jaune

group: ID ok ID not ok

OBSERVER'S PICTURE:

REFERENCE PICTURE:

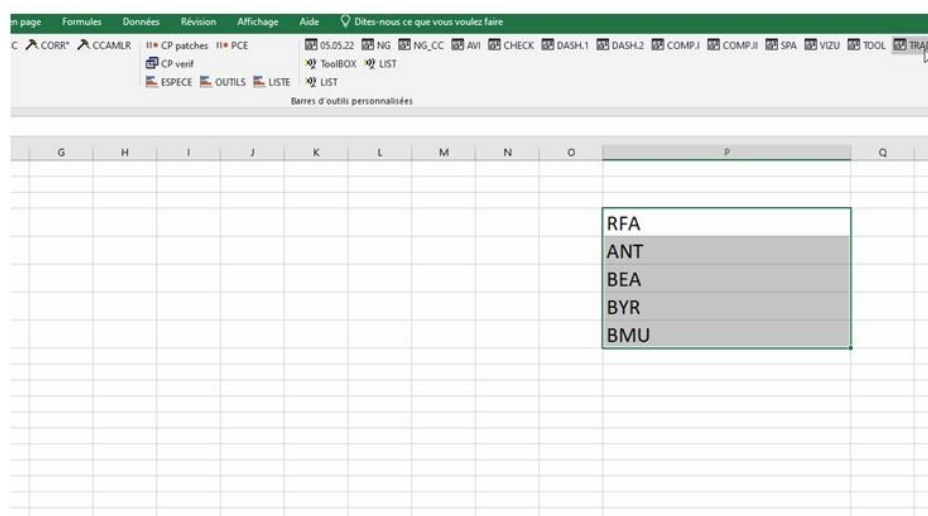
species name: albatros à bec jaune

group taxa comment

smart search

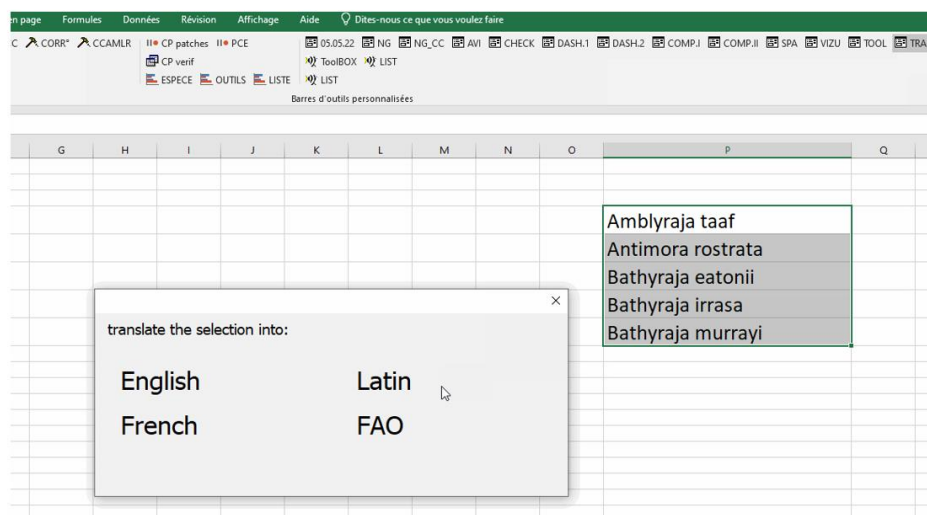
➤ AUTRES UTILISATIONS D'EXCEL

Un module de traduction automatisée sur une sélection de cellule



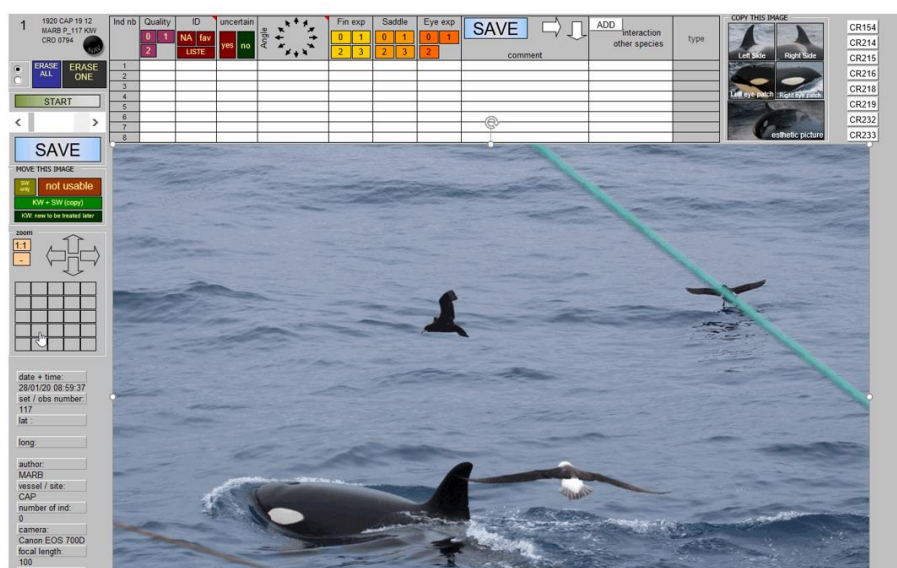
AUTRES UTILISATIONS D'EXCEL

Un module de traduction automatisée sur une sélection de cellule



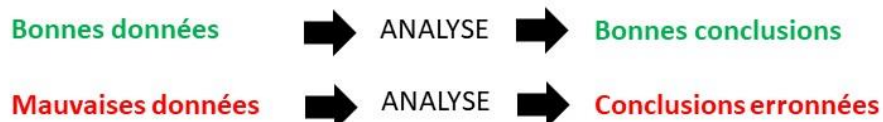
Acquisition de données à partir des images

- Au retour les photos d'orques et de cachalot sont passées dans une interface excel qui permet de saisir rapidement les infos (nom de l'individu, qualité de l'image etc) et de les stocker dans une BDD



CONCLUSION

- Excel est un outils adapté à la saisie sur le terrain (contrairement à R)
- Coût d'entrée peu important pour une amélioration importante
- Le langage VBA ouvre des possibilités importantes
- La qualité de la donnée est la base de tout!



MERCI POUR VOTRE ATTENTION



Acquisition de données en 1939, pour l'agrandissement du métro de New York, récupération de 4 millions de ticket de métro....

Webinaire Collecte des données et Contrôle Qualité avec Excel

31 mai 2022

HUGON Florèn, GASCO Nicolas, NAVARRO Xavier et D'AMICO Frank.

Compte Rendu des Questions – Réponses

Frank D'AMICO – Quelques considérations sur la collecte des données

Commentaire : Échantillonnage probabiliste peu enseigné en écologie et sciences de l'environnement. Pas autant utilisé que ça devrait l'être. Alors qu'il ne faut pas négliger l'échantillonnage. Important car l'échantillonnage ne doit pas introduire un biais de sélection. S'il y a une sélection, alors il faut faire un échantillonnage plus important.

→ Le mieux, c'est faire de l'échantillonnage spatialement équilibré pour réduire ces forces de sélection. Utiliser des unités d'échantillonnage similaires avec le même poids probabiliste.

Question : Comment mesurer l'hétérogénéité ?

→ S'intéresser aux conditions d'échantillonnage, permet d'informer les modèles. Utiliser des covariables pertinentes – collecte de celles en lien avec l'écologie des espèces, qui caractérisent le site ou l'échantillonnage. Attention à ne pas utiliser trop de covariables.

Xavier NAVARRO – Introduction à la programmation sur Excel

Commentaire : Procédure de calcul fortement connectée à Excel. Excel est un endroit pour entreposer les données, les lire. Calcul dans un tableau ou ailleurs (mettre les données dans des bibliothèques, cf DLL). Fonctionnement Excel se rapproche de celui des Base De Données (BDD). Permet de concevoir la BDD, de faire quelques traitements pour ensuite améliorer la collecte des données.

Commentaire : VBA : Très pratique pour stocker la donnée, faire du prototypage, initier l'exploitation, maîtriser ce que l'on fait, bon environnement pour se lancer dans la programmation Mais code VBA n'est pas une véritable compilation. Si beaucoup de données, temps de calculs longs et besoin d'outils plus performants.

Libre Office vs Excel

→ Langage objet sur Libre Office, demande une compréhension plus importante.

Florèn HUGON – Excel : un outil adapté pour contrôler la qualité des données. Exemple sur des données de suivis d'espèces pour une modélisation de l'abondance

Commentaire : Intégrer le contrôle qualité au moment de la saisie, c'est mieux car retrouver la « vraie » donnée est beaucoup plus compliquée

→ Pas toujours possible de le faire si les données ont été récoltées en amont. Intéressant à mettre en place lorsque l'observateur rend compte des données en « temps réel »

Commentaire : Bancariser les fichiers Excel. Éviter d'y introduire soi même des erreurs en effectuant les contrôles. Davantage problématique sur les questionnaires papiers. Besoin d'éditer des règles de gestion (correspond au tableau de critère défini dans la présentation)

Commentaire : Le fichier de saisie Excel n'est pas toujours structuré pour l'analyse de données. Cela requière de savoir comment les données vont être analysées. Il faudrait écrire un **protocole de saisie des données**. Tableau de saisie qui doit être adapté à l'observateur terrain et à celui qui analyse les données. Compromis à trouver entre stockage et traitement / analyse.

Commentaire : Possibilité de **mettre un contrôle à la saisie**, en ne listant que certaines possibilités de saisie. Construire un **protocole de contrôle qualité**. Bien d'intégrer l'informaticien au moment de la création de la base de données ; pour mettre en place menu déroulant, format des dates, ... Permet d'anticiper les erreurs potentielles. Mais attention de ne pas trop priver de liberté l'observateur qui pourrait indiquer une valeur à laquelle nous n'aurions pas pensé et qui serait quand même correcte biologiquement.

Nicolas GASCO – Excel : un outil adapté à l'acquisition des données. Exemple sur des données de pêcheries pour implémentation dans une base de données

Commentaire : Pour des vérifications de saisies, une technique possible est de scanner les sources papiers et d'avoir deux écrans, un pour le fichier où sont les données saisies et l'autre pour afficher (de façon automatisée par excel) le scan associé. Si le nombre de données est très important on peut faire un choix aléatoire.

→ Les vérifications aléatoires sont faites au retour du navire par comparaison aléatoire avec la source papier. De plus le muséum reçoit chaque lundi la totalité des jeux de données des navires en mer pour vérification.

Question : peut-on adapter ce fichier pour nos saisies de captures de tortues ?

→ Le fichier étant très spécifique il serait plus simple de repartir d'un fichier vierge et reprendre les mêmes idées.

Question : peut-on saisir directement sur excel et lier ça à un GPS ?

→ La partie papier est assez pratique pour récolter les données à différents endroits sur le bateau, ensuite il y a un temps de saisie en cabine. Pour l'usine on utilise des planches PVC car trop humide. Pour les GPS en cabine le signal ne passe pas, la plupart des positions sont de toute façon récoltées par le capitaine dans un autre fichier excel (non présenté ici) qui est synchronisé avec le fichier de l'observateur.

Commentaire : la saisie par l'observateur sans délai est toujours meilleure car ça permet de déceler les erreurs directement sur le terrain.

→ Tout ce qui est carnet de terrain électronique est difficile à mettre en place sans avoir un informaticien dédié.

Question : comment est organisé/construit le programme ?

→ Le travail s'étant étalé sur plusieurs années les plus anciens sont moins bien construits, les plus récents utilisent beaucoup de fonctions développées spécifiquement. Les outils sont développés dans des addins (macro complémentaires) pour ne pas alourdir le fichier central (le carnet de pêche).

Question : depuis combien de temps est développé l'outil ?

→ Le fichier a commencé en 1997, rapidement une version assez standardisée était disponible ensuite étant tout seul à programmer et sans cursus informatique. Les outils et améliorations se sont faites au fil de l'eau, les nouveaux systèmes de vérifications sont assez récents.

Question : quel poids pour ce fichier et les images du bouton espèce ?

→ Le CP vide fait 5 Mo vide car je limite la mise en forme des cellules à une dizaine de lignes, au démarrage l'utilisateur clique sur un bouton et la mise en forme est étendue pour la saisie pendant trois mois et le fichier passe à 12 Mo sans données, la mise en forme consomme beaucoup de place.

Le bouton espèce fait appel à des images stockées dans un dossier lui-même dans le même dossier que le carnet de pêche, elles ne sont pas stockées dans le fichier lui-même sinon ce serait beaucoup trop lourd. Afficher une image dans une fenêtre n'est pas très compliqué, c'est juste que ça m'a pris 10 ans pour voir l'idée (on se dit toujours « mais pourquoi je n'y ai pas pensé avant ?? »).

Question : On utilise des tablettes en mer, les fichiers sont très longs à traiter pour pouvoir les charger en base.

→ (Nicolas) Excel sur tablette sous android pas trop de solution. On utilise en revanche des GPS tracker à Saint Paul Amsterdam (autre pêcherie du programme d'observation) et avec quelques clics bouton on récupère nos données. L'interfaçage entre des fichiers et des chargements dans une base de données : sûrement possible d'automatiser des choses.

→ (Xabi) C'est un problème d'automatisation, un informaticien doit pouvoir faire ce travail. Un autre outils puissant : Google form, mais il faut un accès internet.

Échanges libres

Commentaire : Excel et VBA très très pratique pour prototyper, je ne connais pas de meilleur environnement et comme c'est un langage qui n'est pas compilé, on a immédiatement accès au déroulement du code et du contenu des variables.

Commentaire : Excel souffre de son image : « ça fait pas sérieux ». VBA mauvaise presse alors que sur le terrain si on a un problème l'utilisateur va toujours arriver à se débrouiller, quand on développe des applications web le coût de développement est très élevé et les versions changent souvent, alors qu'Excel est beaucoup plus stable, on a pas eu de problèmes de versions de bibliothèques etc. Les utilisateurs sont contents. La diffusion reste difficile, quand on parle à des informaticiens ça fait pas sérieux, ils ont l'impression qu'on fait des petits jouets. Quelques problèmes pour les mises à jours, les utilisateurs n'utilisent pas forcément la dernière version et c'est difficile à suivre.

L'informaticien lui-même a tendance à dénigrer VBA, va faire surtout du C++, ne va pas forcément avoir envie de collaborer par exemple pour améliorer un fichier comme le Carnet de Pêche.

Ce qui est important c'est de savoir ce qu'on a à dire et si on peut l'exprimer dans un langage. Excel est un outil extrêmement puissant. Pour les figures, R est plus performant.

Commentaire : il y a beaucoup de langages de programmation, le propos est surtout d'avoir **une courbe de progression et d'apprentissage qui correspond à son besoin**, si on ne fait pas du R tous les jours on oublie très rapidement, le VBA c'est très différent mais l'objectif n'est pas le même. Il faut utiliser les outils pour ce qu'ils font le mieux. On peut prototyper en VBA et passer en Power Basic (en 32 bits uniquement) pour avoir une vitesse d'exécution comparable à C ou Fortran. Dans Excel, on peut compléter l'environnement et ajouter des fonctionnalités supplémentaires, faire du profilage. On peut avoir un environnement très performant en vitesse de développement. Quand on est biologiste il est

plus facile de se lancer dans le VBA et aller dans le code pour déboguer sans demander à un informaticien.

Commentaire : On a du mal à traiter les données à cause de la qualité. Adapter son travail à un autre sujet sans doute difficile et effectivement plus facile de repartir d'une feuille blanche. Avant l'informaticien faisait un peu de tout, maintenant le domaine est tellement large qu'on peut pas tout faire. En tout cas, Excel est utilisé dans pas mal d'entreprises pour commencer et même continuer. Il y a d'autres outils mais qui nécessitent plus de compétences en informatique.

Commentaire : Excel est un outil plus « démocratisé » *versus* R shiny plus adapté à de la visualisation que de l'acquisition/saisie. R shiny a quand même beaucoup évolué et on peut faire de la saisie dans des formulaires comme excel, besoin de bien s'y connaître en R pour développer.

L'intérêt fort de la collaboration étroite informaticien écologue c'est que l'informaticien réponde vraiment aux besoins de l'écologue. Le programme informatique doit rendre service. Pour mettre en place cette collaboration, il faut que l'écologue montre des exemples de choses à faire à l'informaticien. Exprimer le besoin informatique par le métier (connaissance biologique sur les variables par ex.)

Documentation

Ci-dessous, des informations relayées par Marie Grosdidier, Ingénieure de Recherche Epidémiologiste, Équipe Opérationnelle de la Plateforme ESV - Unité BioSP (marie.grosdidier@inrae.fr)

Les 3 Plateformes d'épidémiosurveillance en France (santé végétale, santé animale et chaîne alimentaire) ont créé un guide sur la qualité des données de surveillance : <https://wiki.esa.inrae.fr/books/guide-pratique-sur-la-qualite-des-donnees-de-surveillance>

La surveillance ici correspond à la surveillance des organismes nuisibles en santé végétale, des agents pathogènes en santé animale et des contaminants de la chaîne alimentaire.

La surveillance en santé végétale par exemple se rapproche beaucoup de la collecte de données naturalistes étant données que nous collectons des données sur l'organisme nuisible et sur son environnement (plantes hôtes, climat ...).

Bilan des retours pour une éventuelle formation / seconde édition du webinaire (5 retours e-mail)

Objectif : apprendre à écrire un programme « propre » pour réussir seul à formater les données

Création de macros/VBA pour la mise en forme et/ou le contrôle qualité des données

Connaissances en Langage Basic

Applications de contrôle qualité

Utiliser un exemple concret pour la formation (participant avec leurs fichiers input ? Exemple publié dans GitHub?), application de quelques lignes de code (copier-coller du code exemple, formule de base, ...)

Modules de formation ? Base manipulation des données / Validation des données / Module avancé macro complémentaire etc)

Seconde édition qui intéresseraient des personnes du réseau de 3 participants sur les 5 répondants.