



Analyser un corpus de données avec Excel

Mardi 27 janvier 2026

Illustrations crédits Saint Oma





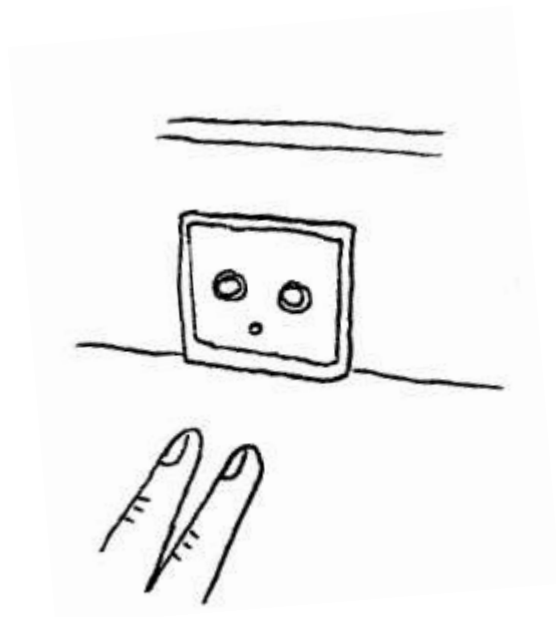
Objectifs et plan de séance

1. Bases et bonnes pratiques

TP : Structurer ses données de recherche

2. Nettoyer et Analyser ses données

TP : Analyse et contrôle des données



1. Bases et bonnes pratiques



Navigation dans Excel

Fichier Accueil Insertion Dessin Mise en page Formules Données Révision Affichage Automatiser Aide Acrobat

Raccourcis claviers utiles

Ctrl + A : tout sélectionner.

Ctrl + flèches : aller vite vers la prochaine cellule remplie

Ctrl + SHIFT : Sélectionner une plage.

Ctrl + SHIFT + flèches : sélection rapide.

Ctrl + F/H : rechercher/remplacer.

F2 : édition de la cellule.

Ctrl + S : enregistre le document

Ctrl + E : remplissage automatique

Ctrl + B : recopie les valeurs de la cellule du dessus

Ctrl + Shift + V : Coller les valeurs (et pas la formule)

Accueil pour :

Saisir et modifier les données, la mise en forme des cellules (police, couleur, bordures), Copier / coller et trier et filtrer.

Insertion pour:

insérer des tableaux, créer des graphiques, ajouter des images et des formes

Données pour:

Trier et filtrer, pour la validation des données, Supprimer les doublons et importer des données



Gestion des feuilles

TITRE

Ben liebt Anna / Peter Härtling bearbeitet von Iris Felter Illustrationen Niels Roland

Die Entdeckung der Currywurst / Uwe Timm

Eigentlich möchte Frau Blum den Milchmann kennenlernen : 21 Geschichten / Peter Bichsel

Gardinella y el gran viaje / Daniela Godel

Gutenachtgeschichten für Celeste : Ein sehr gruseliges Bilderbuch

Am Beispiel meines Bruders / Uwe Timm bearbeitet von Angela Vitt Illustrationen

Traumnovelle

Ausser sich : roman / Sasha Marianna Salzmann

Brigitta

Corpus delicti : ein Prozess / Juli Zeh

Der Mann im Strom : Roman / Siegfried Lenz

Die Ermordung einer Butterblume : ausgewählte Erzählungen / Alfred Döblin m

Der Zug war pünktlich : Erzählung / Heinrich Böll

Hotel Savoy : Ein Roman / Joseph Roth

Der Kleine Grenzverkehr / Erich Kästner Bearbeitet von Gisela Betke Nielsen

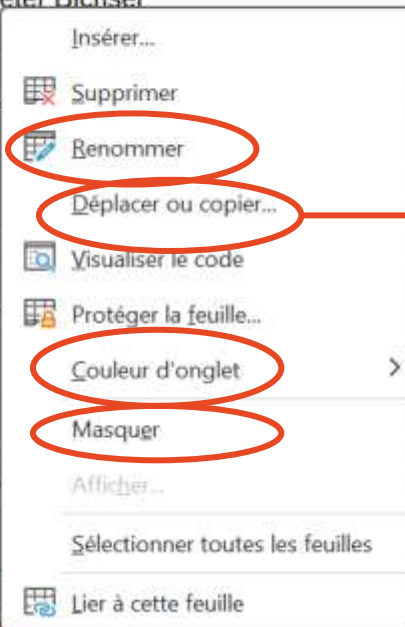
Der Steppenwolf : Mit Texten und Entwürfen zur Entstehung des Romans / Her

Der Vorleser : Roman / Bernhard Schlink

Deutsch für alle : Das endgültige Lehrbuch

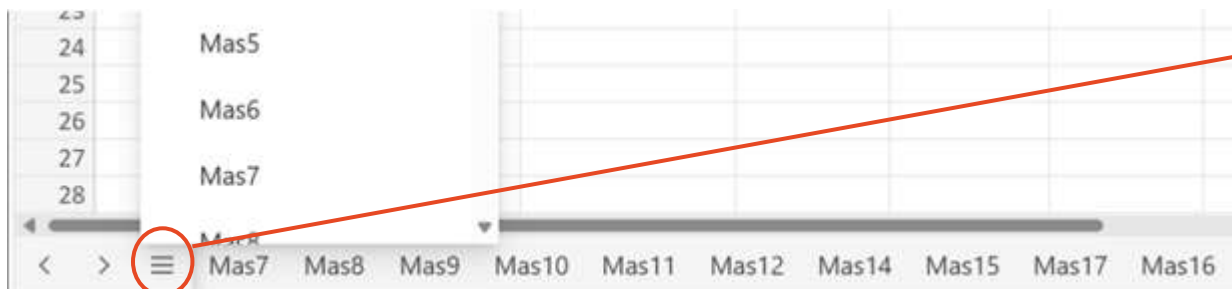
Die Brücke vom Goldenen Horn : Roman / Emine Sevgi Özdamar

> liste_cotes_ECLA_en_traitement liste_cotes_ECLA_en traitem (2



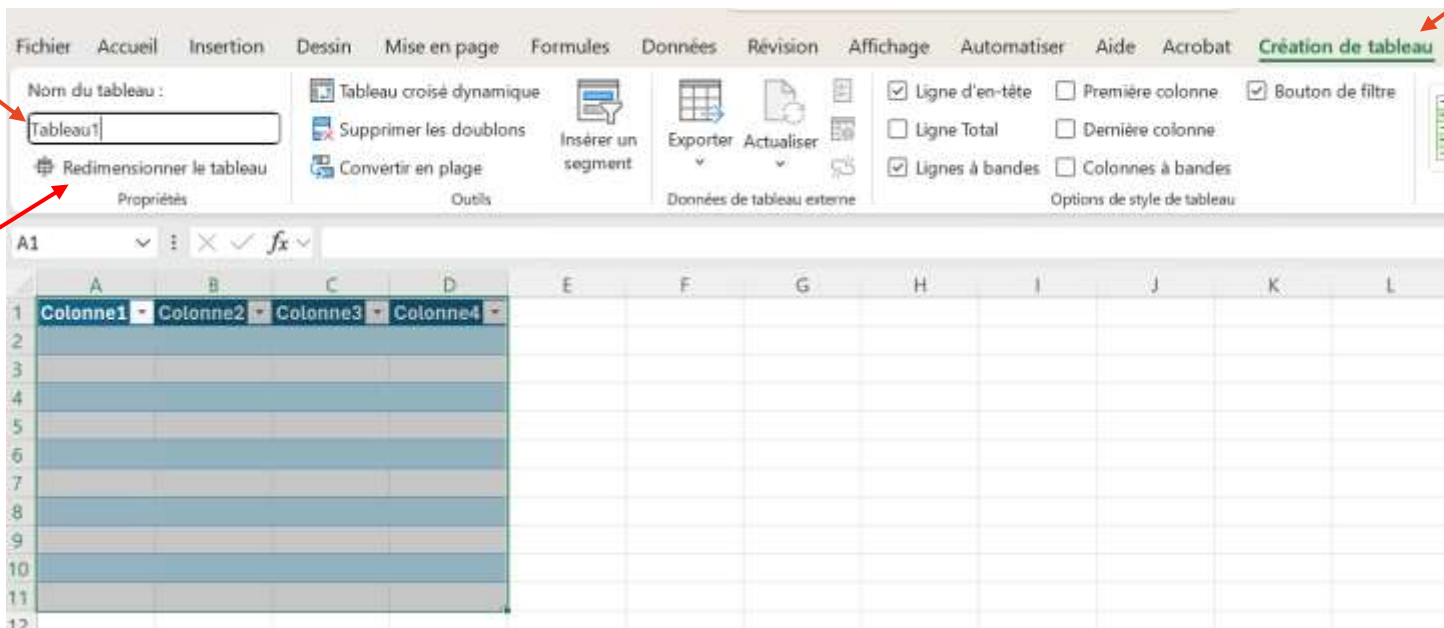
Excel permet de choisir
l'emplacement de la feuille
copiée

Excel affiche **la liste complète
des feuilles**





Création et gestion des tableaux



Cliquer sur **Insertion** → **Tableau**, puis définir la plage de données.

Sélectionner directement une plage de cellules → **Créer un tableau**.

Attention aux données formatées!

1 colonne = 1 type de donnée

Pas de cellules vides dans un tableau

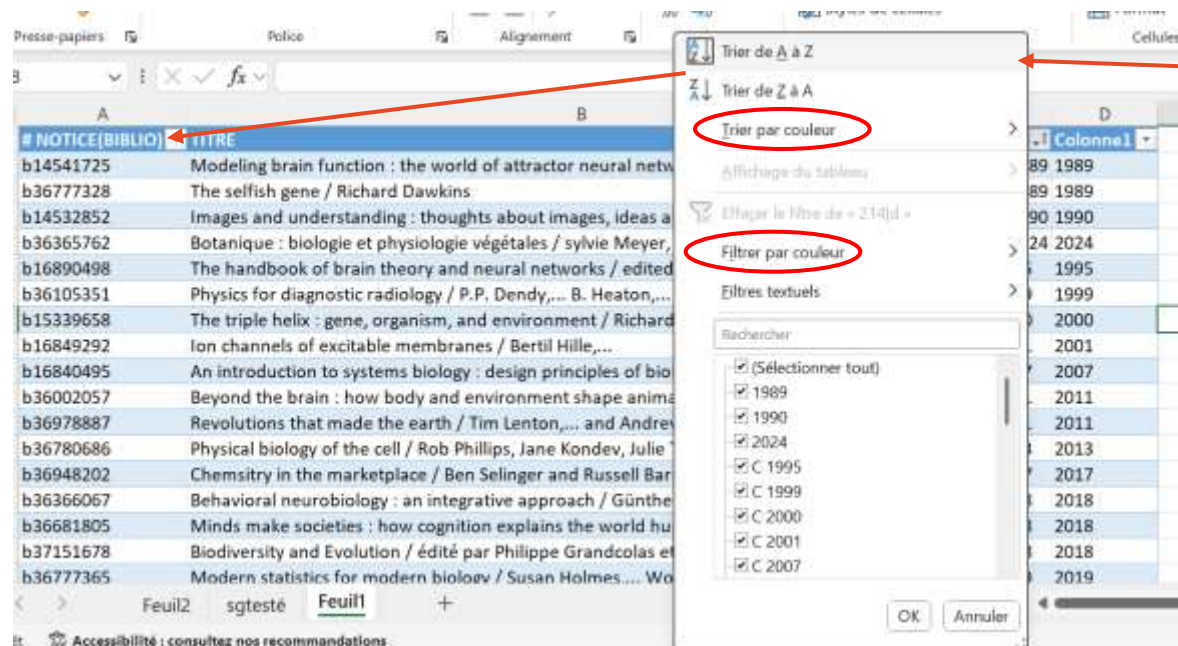
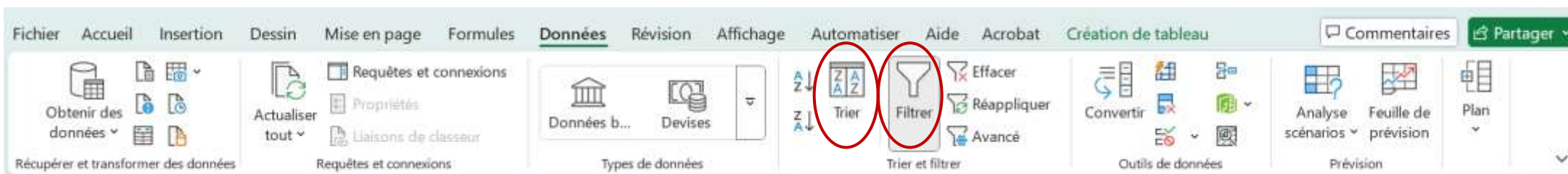
Formats corrects (nombre, texte, date)

Toujours une ligne d'en-tête

À vérifier **avant** : tableaux, calculs, tris, graphiques



Création et gestion des tableaux



Deux façons de trier

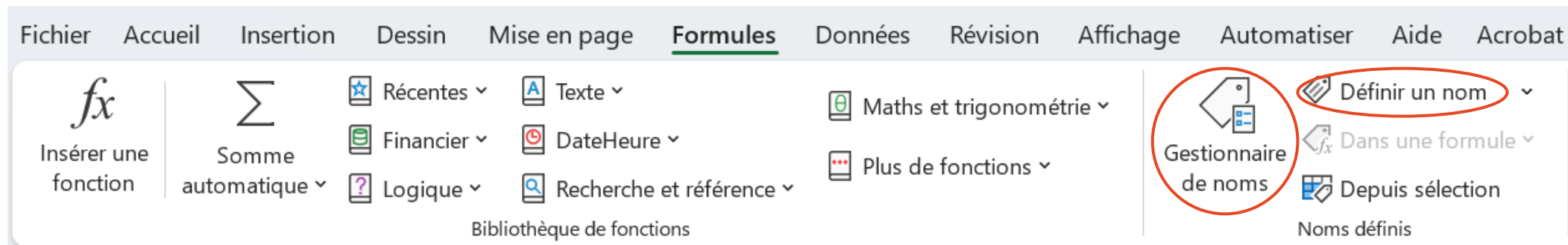
Trier ≠ Filtrer :

- Le tri modifie l'ordre des données
- Le filtre masque les données non pertinentes sans les supprimer.

Possibilité de trier et de filtrer par couleur



Création et gestion des tableaux



Les **plages nommées** sont des **zones de cellules** (ou une cellule unique) à laquelle on attribue un **nom personnalisé**. Elles sont essentielles pour faire des tableaux dans Excel.

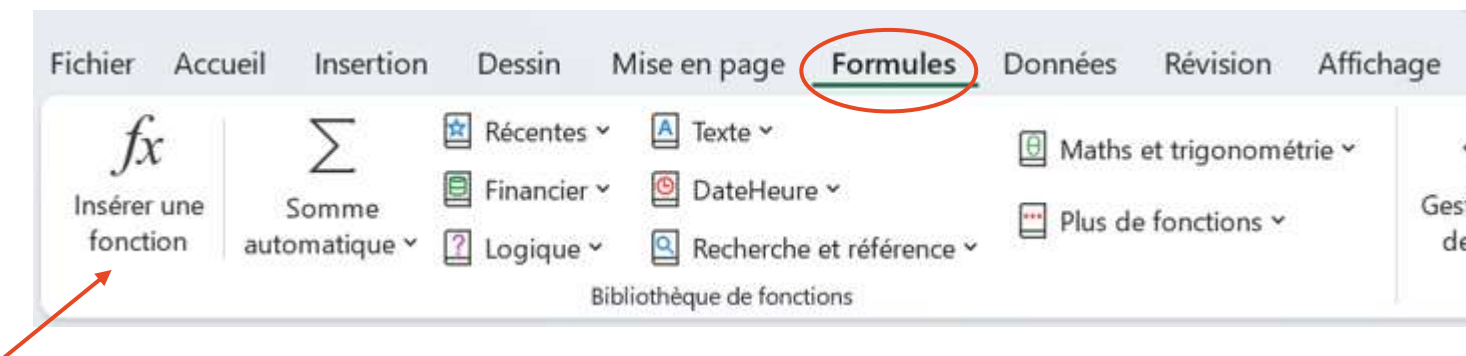
Une plage nommée permet de :

- **Simplifier la lecture des formules** (ex. : `=SOMME(Notes)` au lieu de `=SOMME(B2:B10)`)
- **Réduire les erreurs** dans les formules
- **Rendre un fichier plus clair et plus professionnel**
- **Faciliter la maintenance et la modification des calculs**
- **Réutiliser facilement une même plage de cellule**

Il est également possible de nommer des listes!



Formules utiles



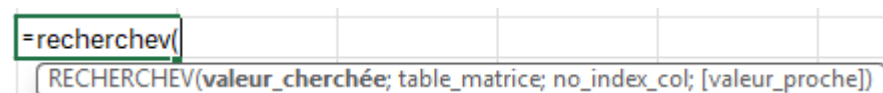
Formules : Insère et sélectionne des fonctions, rend les calculs plus clairs et vérifier la fiabilité des calculs

La fonction **=CONCAT()** permet de combiner du texte provenant de plusieurs cellules.

La formule **=NB.SI()** sert à compter le nombre de cellules dans une plage qui respectent un certain critère.



La fonction **=RECHERCHEV()** permet de chercher une valeur dans une colonne et de renvoyer la valeur correspondante d'une autre colonne.





L'identifiant unique

A2 ✕ ✓ f_x = "DOC-" & LIGNE() - 1

	A	B	
1	ID	# NOTICE(BIBLIO)	TITRE
2	DOC-1	b14541725	Modeling brain functi
3		b36777328	The selfish gene / Rich
4		b14532852	Images and understar
5		b36365762	Botanique : biologie e
6		b16890498	The handbook of brai

Aussi appelé clé est une valeur qui **ne se répète pas** et qui sert à **identifier une ligne de données de façon certaine**.

Elle permet:

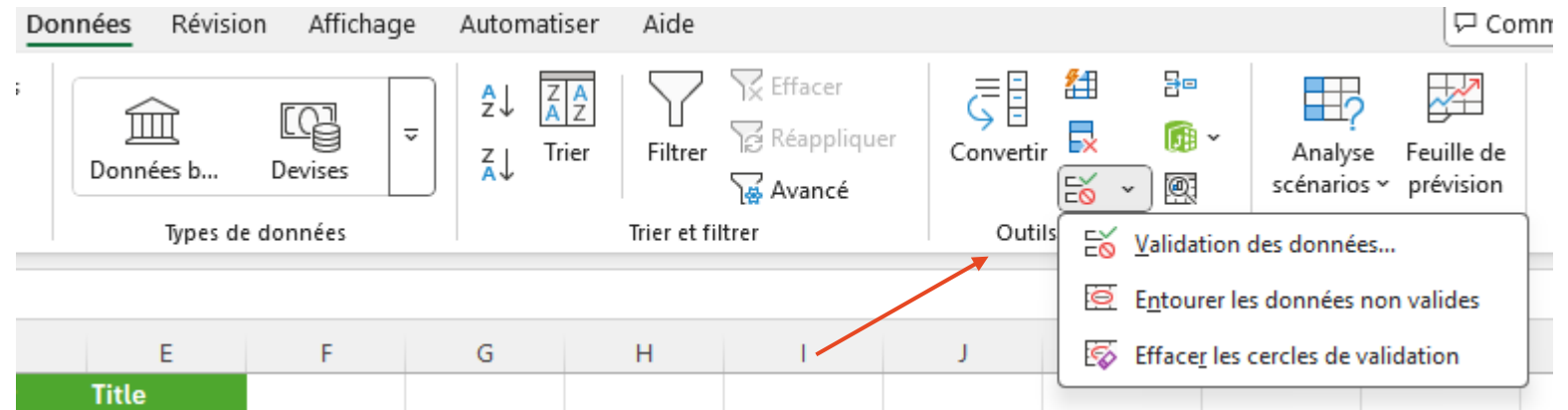
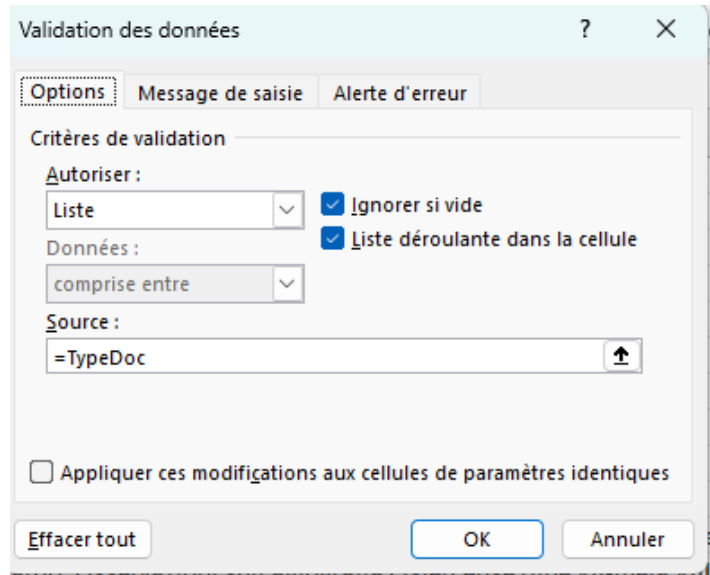
- D'éviter les doublons
- Facilite la recherche
- De travailler avec les tableaux croisés dynamique

Tirez vers le bas et Excel remplit automatiquement toutes les lignes.

Fonctionne aussi avec les jours de la semaine, les mois...



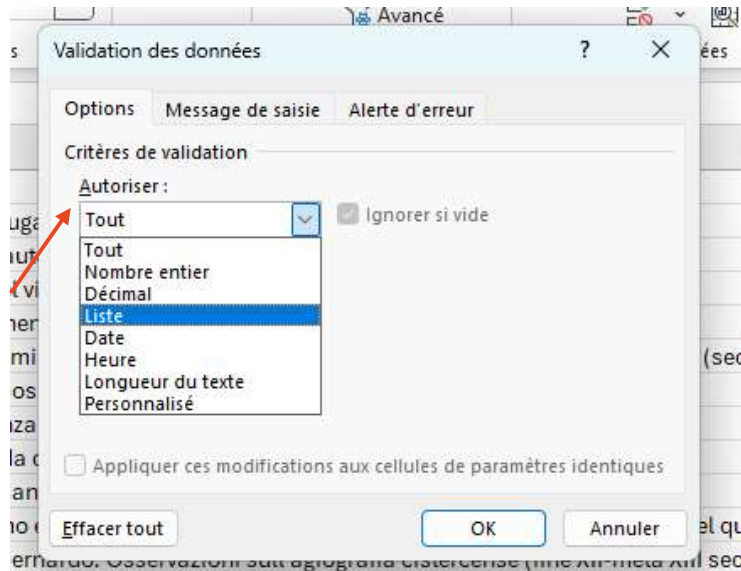
Validation des données



La **validation des données** sert à **contrôler ce que l'utilisateur peut saisir dans une cellule**.

Elle permet de:

- Éviter les erreurs de saisie
- Forcer un format
- Gagner du temps et standardiser les données
- Contrôler des données externes





Travaux pratiques : structurer ses données de recherche

1. Structure du classeur

Nouveau classeur Excel, créer les Feuilles : **Suivi**, **Personnes**, **Lieux**, **Listes**, une **Couleur** différente par feuille

2. Tableaux et identifiants

Créer un tableau par feuille avec en-têtes :

Suivi : IdDoc, TypeDoc, Auteur, Lieu, État

Personnes : Civilité, Nom, Prénom, Profession, Naissance, Décès, Âge

Lieux : IdLieu, NomLieu, Adresse, Pays, Site web

Nommer les tableaux : TabSuivi, TabPersonnes, TabLieux, TabListes

Identifiants uniques : IdDoc (Suivi), IdPers (Personnes), IdLieu (Lieux)



3. Calculs, listes et contrôles

Calculer l'**âge** (naissance / décès ou aujourd'hui), Créer **Nom + Prénom** Créer les listes :Civilité (M., Mme)Type doc (livre, article, archive, lettre, autre), État (à voir, dépouillé, à creuser), Validation des données :

Civilité (Personnes), Type doc, État (Suivi), Auteur (Personnes), Lieu (Lieux)

Dans **Personnes** : compter les documents associés à chaque personne



2. Nettoyer et analyser les données

Sources de données et intégration dans Excel

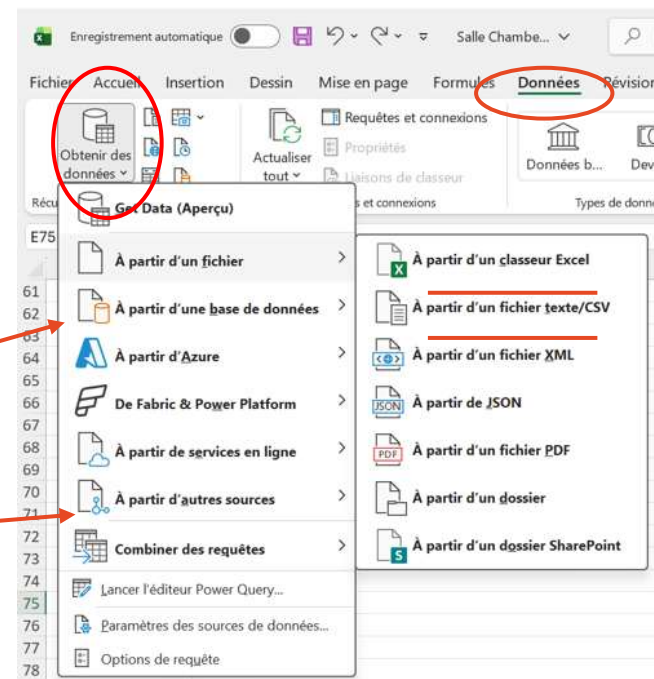
Sources possibles de données

- **Fichiers texte** : .txt (données séparées par tabulation ou espaces)
- **Fichiers CSV** : .csv (valeurs séparées par virgule ou point-virgule)
- **Autres fichiers Excel** : .xls / .xlsx
- **requêtes API** : récupérer des données en ligne
- **Bases de données** : SQL, Access, etc.



SQL, Access

Requêtes API





Nettoyer ses données

Avant de créer un tableau croisé dynamique, il est essentiel d'avoir des données **propres, cohérentes et sans erreurs**. Voici quelques méthodes:

Prévenir les erreurs avant la saisie

Validation des données :

- Listes déroulantes : choisir parmi des valeurs prédéfinies (vu en diapo 10)
- Nombres uniquement : empêche la saisie de texte
- Dates valides : évite les erreurs de date
- Cellules obligatoires : interdire les cellules vides

Contraintes de saisie :

- Valeurs minimales / maximales : encadrer les nombres
- Longueur du texte limitée : limiter le nombre de caractères

L'objectif est d'empêcher les fautes et les incohérences dès le départ.





Nettoyer ses données

Repérer les erreurs en utilisant :

La mise en forme conditionnelle :

- Doublons
- Cellules vides
- Valeurs anormales

Le tri des données (A → Z / Z → A / couleurs) :

- Regrouper des valeurs similaires
- Détecter les incohérences (ex. : *France* / *france*)

Les filtres automatiques :

- Filtrer pour afficher les cellules vides
- Isoler certaines valeurs

L'objectif est de comprendre où se trouvent les erreurs dans vos données.





Nettoyer ses données

Corriger les erreurs :

Supprimer /corriger les doublons

Rechercher et remplacer (Ctrl + H) : harmoniser les écritures (ex. : *cécile* / *Cecile* / *Cécile*)

MAJUSCULE / MINUSCULE / NOMPROPRE : uniformiser les écritures

Texte en colonnes : séparer ou rassembler nom / prénom, nettoyer dates, codes...

Formule SUPPRESPEACE() : supprimer les espaces invisibles

Conversion texte → nombre : corriger les valeurs mal formatées
(type de données identifié)

Fonctions logiques : SI()

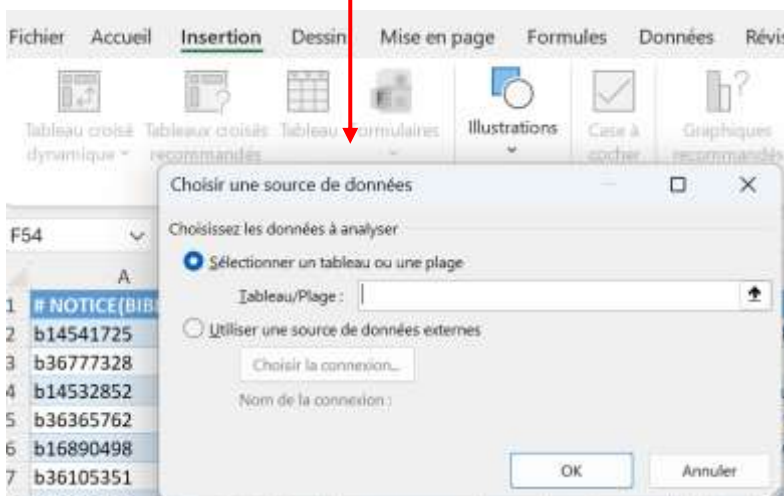
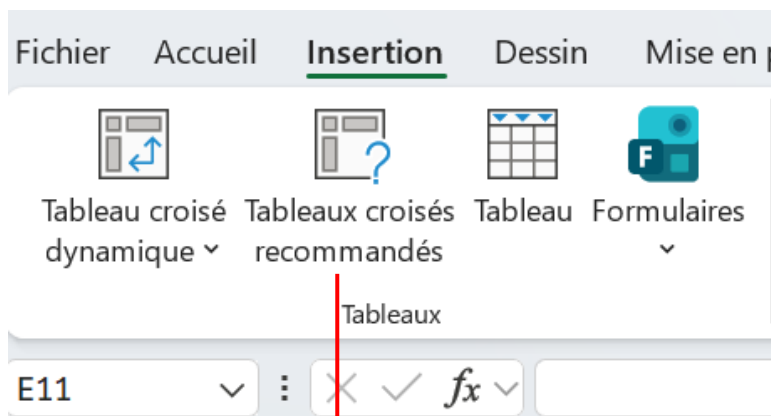
Masquer temporairement lignes ou colonnes :
se concentrer sur certaines données sans les supprimer.

	Standard
123	Aucun format particulier
12	Nombre
	Horvitz, Maria Eugenia
	Monétaire
	Horvitz, Maria Eugenia
	Comptabilité
	Horvitz, Maria Eugenia
	Date courte
	Horvitz, Maria Eugenia
	Date longue
	Horvitz, Maria Eugenia
	Heure
	Horvitz, Maria Eugenia
	Pourcentage
	Horvitz, Maria Eugenia
	Fraction
	Horvitz, Maria Eugenia
	Scientifique
	Horvitz, Maria Eugenia
	Texte
	Horvitz, Maria Eugenia
	Autres formats numériques...

(a



Analyse des données : Les tableaux croisés dynamiques (TCD)



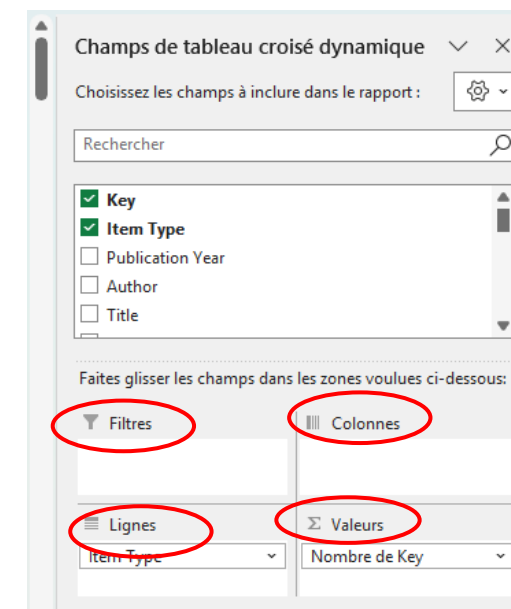
Un **tableau croisé dynamique** permet de :

- résumer rapidement** un grand volume de données
- De **croiser des données** comme le type de support, l'année de publication, les pics de publications ou des collaborateurs.
- De produire des **statistiques descriptives**
- D'explorer les données **sans modifier le tableau source**.

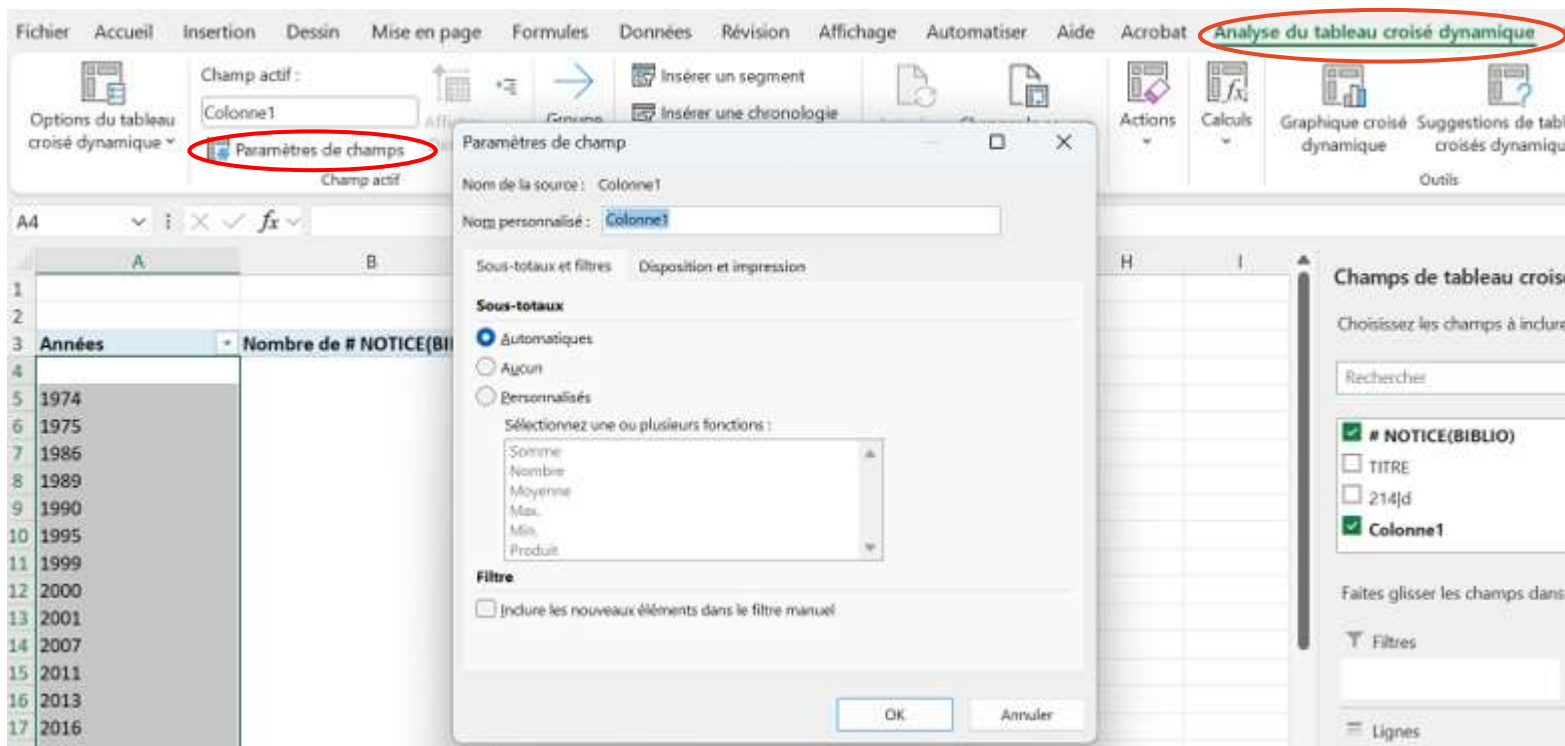
Au choix:

Tableau croisé dynamique classique : création manuelle, offrant une grande flexibilité dans l'organisation et l'analyse des données.

Tableau croisé dynamique recommandé : suggestions automatiques proposées par Excel, permettant une création rapide de tableaux synthétiques, mais avec moins de flexibilité.



Les tableaux croisés dynamiques (TCD)



"Paramètres des champs de valeurs..." permet de:

- Choisir le type de calcul appliqué (Somme, Moyenne, Nombre, Max, Min...)
- Modifier la façon dont les données sont résumées (par ex, passer d'un total à une moyenne)
- Personnaliser l'affichage, comme afficher les résultats en pourcentage du total
- Modifier le nom affiché du champ dans le tableau croisé dynamique



Les tableaux croisés dynamiques (TCD)

The screenshot displays the Microsoft Excel interface. The top ribbon is 'Analyse du tableau croisé dynamique', with the 'Insertion' tab highlighted. The 'Graphiques recommandés' group is circled in red. The bottom ribbon is 'Création', with the 'Analyse du tableau croisé dynamique' tab highlighted. The 'Champs de tableau croisé dynamique' task pane is open on the right, showing a list of fields to include in the report. The 'Key' and 'Publication Year' fields are checked.

Champs de tableau croisé dynamique

Choisissez les champs à inclure dans le rapport :

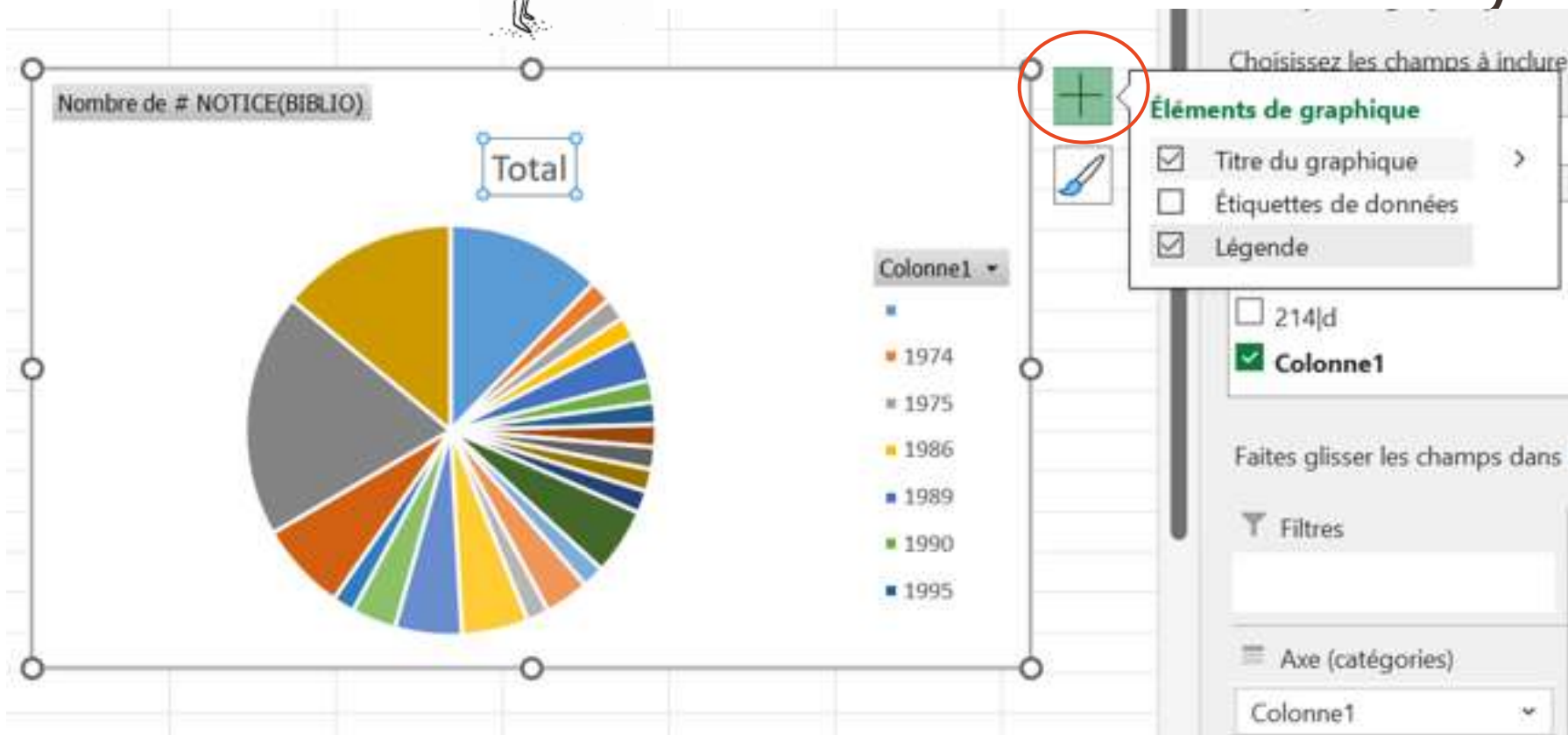
Rechercher

- ☒ Key
- ☐ Item Type
- ☒ Publication Year
- ☐ Author
- ☐ Title
- ☐ Publication Title
- ☐ ISBN
- ☐ ...

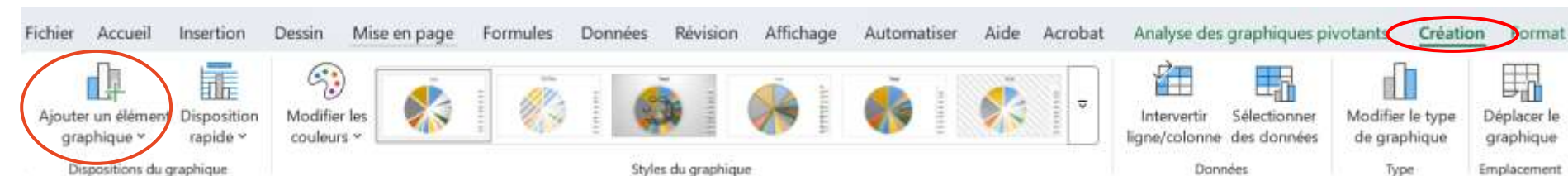
	1957	1961	1962	1963	1964	1965	1970	1972	1973	1974	1975	1976	1977	1978	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2006	2007	2008	2009	2010
1	1	2	1	3	1	1	1	1	1	2	3	2	1	2	3	2	5	5	20	37	37	38	61	66	29	25	42	40	24	13	19	13	8	8	11	8	5	7	8	5	4	4	5	

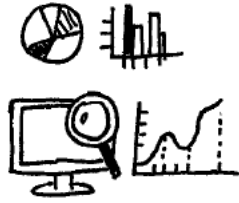


Les tableaux croisés dynamiques (TCD)



Vous pouvez choisir
le style de votre
graphique ainsi que
ses couleurs



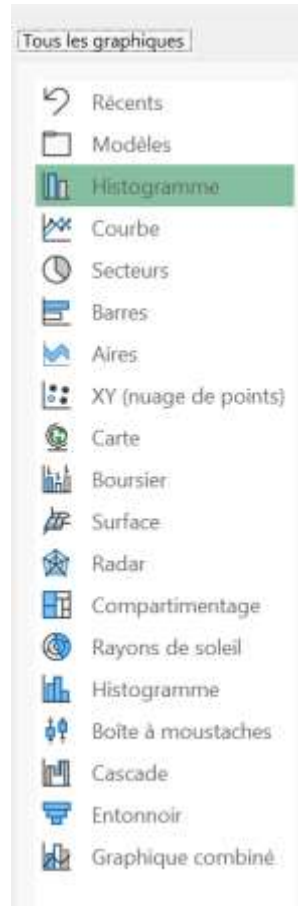


Les tableaux croisés dynamiques (TDC)

Visualisation

Les différents types de graphiques doivent être choisis en fonction des besoins : toutes les visualisations ne sont pas adaptées à tous les types de données.

- **Graphique en courbes (Line chart):** Évolution d'une variable en fonction du temps
- **Graphique en colonnes / barres:** Comparer des catégories ou conditions
- **Graphique en nuage de points (Scatter plot):** Étudier la relation entre deux variables quantitatives
- **Graphique en histogramme:** Visualiser une distribution, répartition des valeurs
- **Graphique en secteurs (camembert):** À utiliser seulement si : Peu de catégories (≤ 5) et des proportions claires





Travaux pratiques : Analyse et contrôle des données

1. Qualité et cohérence des données

Appliquer une mise en forme conditionnelle sur **l'âge** et sur **le nombre de documents**.

Utiliser une mise en forme conditionnelle pour repérer **les doublons** dans **Personnes** et dans les **Lieux**.

Remplir quelques lignes de données « au hasard » pour tester le classeur.

Créer les **identifiants selon un plan de nommage** et **vérifier l'absence de doublons** par mise en forme conditionnelle.

Masquer la feuille Listes une fois les contrôles effectués.

2. Bilan et visualisation

Créer une feuille **Bilan** contenant des statistiques simples.

Calculer **la répartition des documents par type**.

Créer un **graphique en camembert** sur les types de documents.

3. Analyse du contenu (TDC)

À partir de la table des contenus, **identifier les langues les plus utilisées**.

Repérer **les collaborateurs et les pics de publication**.





Pour aller plus loin

[Pix.fr](https://pix.fr) | Évaluer, développer et certifier vos compétences numériques | Pix

PIX est le **service public de certification des compétences numériques**.

La certification PIX est **reconnue par l'État** et inscrite dans le cadre officiel de l'Éducation nationale.

Elle est utilisée dans l'enseignement scolaire, l'enseignement supérieur, la formation professionnelle.

Elle permet d'**attester officiellement le niveau de compétences numériques**

Pix est **gratuit** en passant par le **cadre institutionnel**.

<https://www.automateexcel.com/formula-generator-ai-bot/>

Outil en ligne gratuit qui utilise l'**intelligence artificielle (IA)** pour **créer des formules Excel complexes** automatiquement et **répondre à des questions en langage naturel** (Décrivez l'action, l'outil crée la formule.). Il peut être utilisé pour **Excel ou Google Sheets**

<https://openclassrooms.com/fr/courses/7168336-maitrisez-les-fondamentaux-d-excel>

OpenClassrooms est une plateforme de formation en ligne qui propose des cours dans de nombreux domaines, y compris Excel. Ces cours comportent des vidéos, des explications détaillées et des exercices



Vos retours sur la séance, merci !

enquetes.ens.psl.eu/index.php/414889



Se former à la BnF

POUR LES DOCTORANTS : ATELIER “PRÉPARER SA THÈSE A LA BNF”

En 2h30, rendez-vous
personnalisé afin de repérer la
documentation et les sources
utiles à son sujet de thèse dans
les collections de la BnF.

Contact : doctorants@bnf.fr

SERVICE SINDBAD : SERVICE D'INFORMATION DES BIBLIOTHÉCAIRES À DISTANCE

Pour obtenir une réponse à une question,
identifier le bon interlocuteur à la BnF ...

Lien : <https://www.bnf.fr/fr/une-question-pensez-sindbad>



BNF DATALAB : ACCOMPAGNEMENT DE PROJETS DE RECHERCHE EN RELATION AVEC LES COLLECTIONS NUMÉRISÉES OU NÉES NUMÉRIQUES DE LA BNF

Repérage des documents en lien avec un sujet
donné : constitution de corpus, extraction, ...
Formations et activités scientifiques : journées
d'étude, ateliers de prise en main d'outils.

Lien : <https://www.bnf.fr/fr/postulez-au-bnf-datalab>



[Catalogue de la formation des usagers](#)



Se former à la BnF

Chercher dans les collections de la BnF
Séance du jeudi 05/02/26

Chercher dans les archives du web
Séance du jeudi 12/02/26





Bibliothèque[s] de l'ENS
ULM-JOURDAN

L'offre de formation des bibliothèques ENS

Pour toute demande
d'informations :
formations-bib-ens@ens.psl.eu

Rubrique Formations du site des bibliothèques :
<https://bib.ens.psl.eu/formations>

Science
ouverte, Hal

Zotero,
Tropy,

Archives
du web

Identité
numérique

Circuit de la
thèse

Veille
documentaire

Droit
d'auteur

Plan de
gestion de
données



L'offre de formation des bibliothèques ENS

Calendrier de la formation doctorale

- Organiser sa recherche en sciences humaines et sociales** - Lundi 19/01/26, 14H-16H
- Gérer sa bibliographie avec Zotero - Niveau 1** - Mardi 20/01/26, 14H-16H - Jeudi 22/01/26, 14H-16H
- Approfondir Zotero : fonctionnalités avancées - Niveau 2** - Lundi 26/01/26, 14H-16H
- Analyser un corpus de données avec Excel** - Mardi 27/01/26, 14H-16H
- S'initier à la science ouverte** - Lundi - 02/02/26, 14H-16H
- Boîte à outils pour la recherche en SHS** - Mardi 03/02/26, 14H-16H
- Organiser sa recherche dans les collections de la BnF** - Jeudi 05/02/26, 9h30-12h
- Organiser ses images avec le logiciel Tropy** - Lundi 09/02/26, 14H-16H
- Gérer ses « données de recherche » en sciences humaines et sociales** - Mardi 10/02/26, 14H-16H
- Les archives du web : une source pour la recherche (Bnf)** - Jeudi 12/02/26, 9h30-12h
- Gérer son identité numérique** - Lundi 16/02/26, 14H-16H
- Le circuit des thèses électroniques : dépôt, archivage et diffusion** - Mardi 17/02/26, 14H-16H

S'inscrire : plateforme ADUM (doctorants PSL) ou formations-bib-ens@ens.psl.eu (autres publics)