

Données — Structuration et traitement

SNT · Seconde
Contrôle n°1 · 55 min · sur 20

Exercice 1 — Structurer

5 pts 13 min

Structurer des données

Réponds à chaque question.

LA RÈGLE DU COURS

Une organisation régulière rend le traitement automatique possible, et deux formats de fichier se partagent les usages.

1. Quand des données sont-elles dites structurées ? (1 pt)

Lorsqu'elles sont organisées selon un schéma connu, qui permet à un programme de les lire sans interpréter.

→ C'est cette régularité, et elle seule, qui rend un traitement automatique possible.

2. Comment une table est-elle organisée ? (1 pt)

Une ligne par enregistrement, une colonne par descripteur, et la même nature d'information dans toute une colonne.

→ C'est le modèle le plus courant de données structurées.

3. Qu'est-ce que le format CSV ? (1 pt)

Une table écrite en texte brut : une ligne par enregistrement, des valeurs séparées par un caractère, et une première ligne qui nomme les colonnes.

→ Le sigle signifie comma-separated values.

4. Qu'est-ce que le format JSON, et à quoi convient-il ? (1 pt)

Un format organisant les données en couples nom-valeur et acceptant des structures imbriquées ; il convient aux données arborescentes.

→ Le CSV, lui, convient aux tables plates.

5. Que sont les métadonnées, et pourquoi sont-elles indispensables ? (1 pt)

Les données sur les données — ce que mesure une colonne, dans quelle unité, à quelle date, par quel instrument, avec quelle précision ; sans elles aucune valeur n'est interprétable.

→ Un tableau sans métadonnées n'est qu'une suite de nombres.

Exercice 2 — La réidentification

4 pts 11 min

Analyser la réidentification

Réponds à chaque question.

LA RÈGLE DU COURS

Un fichier peut ne porter aucun nom et désigner pourtant une personne. La définition juridique en tient compte.

1. Pourquoi un fichier sans nom n'est-il pas anonyme ? (1 pt)

Parce que le croisement suffit à réidentifier.

→ Un fichier portant la date de naissance, le code postal et le sexe désigne souvent une seule personne dans une commune.

2. Que rend réellement la suppression du nom ? (1 pt)

Elle rend les données moins immédiatement nominatives, pas anonymes.

→ L'ajout d'une seconde source lève le dernier doute.

3. Qu'est-ce qu'une donnée personnelle ? (1 pt)

Une information se rapportant à une personne identifiée ou identifiable, directement ou par recoupement.

→ La définition vise l'identifiabilité, pas le seul état civil.

4. Cite trois exemples de données personnelles autres que le nom. (1 pt)

Une adresse IP, un identifiant de téléphone, une photographie — un historique d'achats convient aussi.

→ Toutes permettent de remonter à une personne.

Exercice 3 — Le RGPD

4 pts 11 min

Énoncer les principes du RGPD

Réponds à chaque question.

LA RÈGLE DU COURS

Un règlement européen encadre la collecte et l'usage, par quelques principes et quelques droits.

1. Que signifie le sigle RGPD, et depuis quand s'applique-t-il ? (1 pt)

Règlement général sur la protection des données ; il s'applique en Europe depuis 2018.

→ Il encadre ce que l'on a le droit de faire des données.

2. Qu'énoncent les principes de finalité et de minimisation ? (1 pt)

On ne collecte que pour un usage annoncé, et l'on ne collecte que le nécessaire.

→ Les deux limitent la collecte avant même qu'elle ait lieu.

3. Cite deux autres principes du règlement. (1 pt)

La durée limitée de conservation et le consentement libre et éclairé.

→ Ils portent sur ce qu'on fait des données une fois collectées.

4. Quels droits le règlement donne-t-il à la personne concernée ? (1 pt)

L'accès, la rectification, l'effacement et l'opposition.

→ Ce sont quatre droits qu'elle peut exercer auprès du responsable du traitement.

Exercice 4 — Traiter

4 pts 11 min

Traiter une table de données

Réponds à chaque question.

LA RÈGLE DU COURS

Quatre opérations suffisent à exploiter une table, et un très grand volume de données change la manière de chercher.

1. Quelles sont les quatre opérations élémentaires sur une table ? (1 pt)

Rechercher, filtrer, trier, calculer.

→ Elles se combinent pour produire n'importe quel traitement simple.

2. L'ordre dans lequel on les applique change-t-il le résultat ? (1 pt)

Non : il change le coût du traitement, jamais le résultat.

→ C'est ce qui permet de les réordonner pour aller plus vite.

3. Qu'est-ce que le big data, et en quoi le changement n'est-il pas seulement d'échelle ? (1 pt)

Le cas où le volume, la vitesse d'arrivée et la variété des données dépassent ce qu'un traitement classique absorbe ; on y cherche des corrélations sans hypothèse préalable.

→ La démarche change, pas seulement la taille.

4. Quelle vigilance cette démarche impose-t-elle ? (1 pt)

Une corrélation n'est pas une cause, et un très grand nombre de croisements en produit mécaniquement qui ne signifient rien.

→ Le volume augmente le nombre de coïncidences apparentes.

Repérer une erreur et la corriger

Chaque affirmation est fausse. Écris ce qu'il fallait dire.

LA RÈGLE DU COURS

Trois fautes reviennent. Un anonymat supposé, une corrélation promue en cause, et un calcul faussé par des cases vides.

1. « J'ai retiré la colonne des noms : mon fichier est anonyme. » (1 pt)

Le croisement avec une autre source suffit à réidentifier les personnes.

→ Date de naissance, code postal et sexe désignent souvent une seule personne.

2. « Ces deux courbes montent ensemble : l'une cause donc l'autre. » (1 pt)

Une corrélation n'est pas une cause.

→ Un très grand nombre de croisements produit mécaniquement des corrélations sans signification.

3. « Les cases vides de ma colonne valent zéro : ma moyenne est correcte. » (1 pt)

Une valeur manquante n'est pas une valeur nulle : la moyenne est faussée.

→ Il faut repérer les valeurs manquantes ou aberrantes avant tout calcul.

Barème — ce que le correcteur regarde

Les lignes chiffrées couvrent les 5 exercices et totalisent 20 points.

EXERCICE	ATTENDU	POINTS	POINT PERDU SI
Ex.1	Structurer pour exploiter	5 points — 1 par question	CSV et JSON employés indifféremment.
Ex.2	La réidentification	4 points — 1 par question	Un fichier sans nom déclaré anonyme.
Ex.3	Le RGPD	4 points — 1 par question	Un principe cité sans être défini.
Ex.4	Traiter une table	4 points — 1 par question	Une valeur manquante comptée pour zéro.
Ex.5	L'erreur repérée, et la réponse juste donnée	3 points — 1 par ligne	Une correction qui garde la corrélation pour une cause.

ET MAINTENANT ?**16 à 20**

Le chapitre est acquis. Entraîne-toi maintenant à exploiter un fichier en cinq temps : la ligne d'en-tête, les métadonnées, les valeurs manquantes ou aberrantes, les opérations dans l'ordre, et une conclusion qui dit ce que les données ne mesurent pas.

10 à 15

Refais l'exercice 2 en te demandant chaque fois quelle seconde source lèverait le dernier doute.

moins de 10

Reprends la fiche de cours et retiens ce que le nombre 37 ne dit pas : sans indication sur ce qu'il mesure, quand et comment, il n'autorise aucune conclusion.