TECH XV - relier les données sans perdre le fil

Pas à pas participant · séance proposée du 12 novembre 2026

Thomas Fesq · Formatome · Révision R3 du 10 novembre 2026

Le staff fictif a reçu des observations de touches et quelques notes complémentaires. Lundi, vous avez vu comment un assistant pouvait préparer une synthèse et comment revenir au fichier pour la vérifier. Cette fois, votre mission est de raccorder les notes aux bonnes actions. Une remarque supplémentaire ne doit pas faire apparaître une touche supplémentaire.

Vous pouvez suivre toute la séance en regardant la démonstration. Les manipulations personnelles sont facultatives et peuvent être reprises ensuite. Si vous travaillez sur papier, le raisonnement et les petits tableaux constituent une pratique complète. Vous n’avez pas besoin de transmettre un fichier de votre club, de créer un compte ou d’installer un outil.

À la fin, vous devez pouvoir montrer la table enrichie, retrouver les notes non raccordées et revenir à une version vérifiée si un essai a produit un mauvais résultat.

Les fichiers et les mots utiles

Le fichier de départ s’appelle **Cas-fictif-09-novembre-enrichi.csv**. Il comporte 23 lignes de données et six colonnes : identifiant d’événement, match, phase, zone, issue et repère vidéo. La copie de travail contrôlée, **Evenements-corriges.csv**, comporte 22 événements. Le formateur fournit aussi cinq notes fictives et un petit exemple GPS ; ce sont des données d’exercice, jamais des observations de joueurs réels.

Un **CSV** est un tableau enregistré sous forme de texte. Dans ce cas, le point-virgule sépare les colonnes. **UTF-8** désigne l’encodage qui permet de conserver correctement les accents. Le **grain** explique ce que représente une ligne : ici, un événement de match ; dans un export GPS, ce pourrait être un segment de séance. Une **clé composée** rassemble plusieurs champs pour retrouver le bon objet, par exemple le match et l’identifiant de l’événement.

Une **jointure** rapproche des tables à partir de leur clé. La **jointure gauche** part de la table placée à gauche et conserve ses éléments même lorsqu’aucune correspondance n’existe à droite. Une **note sans événement**, parfois appelée note orpheline, ne trouve aucune clé correspondante dans la table des événements. Une relation **un-à-plusieurs** apparaît lorsqu’un événement rencontre plusieurs notes : la fusion peut alors répéter l’événement sur plusieurs lignes.

Le **journal de versions** garde la trace d’une entrée, d’une modification, d’un contrôle et de la décision de conserver ou d’abandonner le résultat. Un **retour arrière** consiste à repartir d’une version vérifiée, tout en gardant la trace de l’essai qui a posé problème. Vous trouverez les autres termes dans le glossaire évolutif du module.

1 - ouvrir le fichier sans le transformer

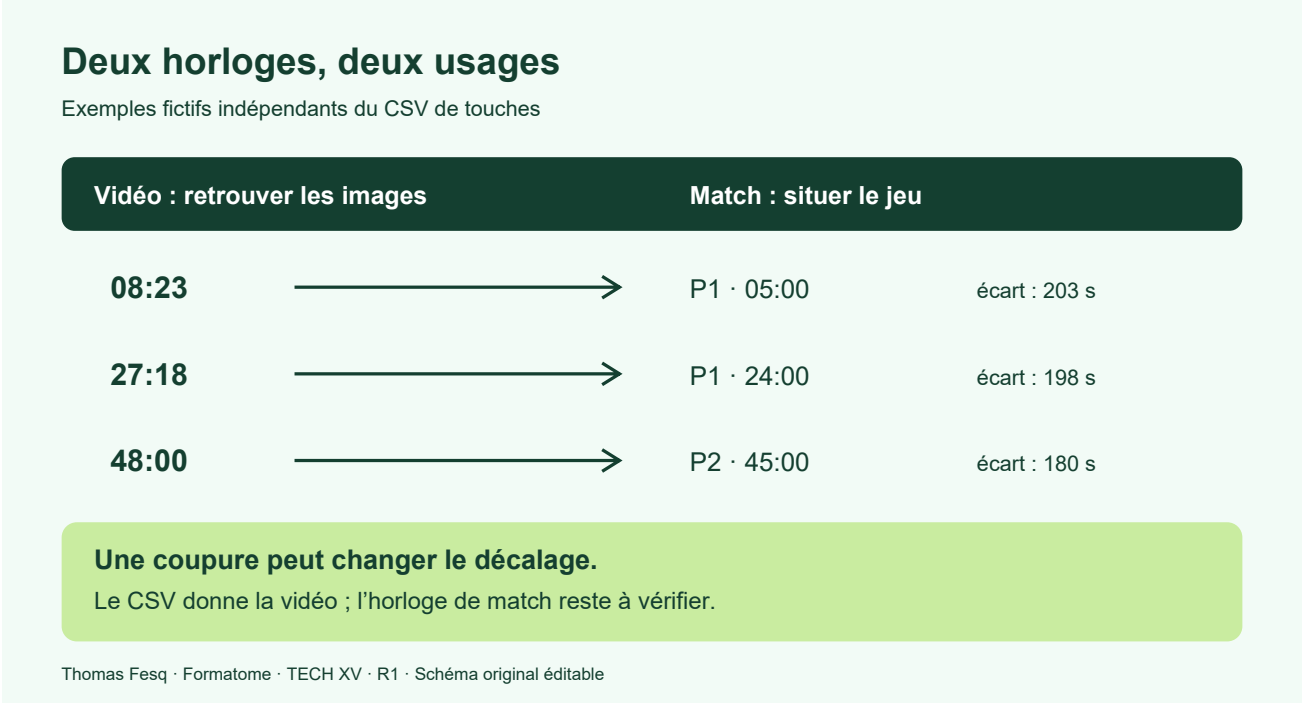
Gardez le fichier original à part et travaillez sur une copie. Dans votre tableur, utilisez l'entrée d'import d'un fichier texte ou CSV, puis choisissez le séparateur point-virgule et l'encodage UTF-8. Avant de charger le tableau, définissez la colonne timecode_video comme **texte**. Les identifiants restent eux aussi du texte.

Dans Excel, l'import se trouve généralement dans **Données → À partir d'un fichier texte/CSV**. Si vous passez par **Transformer les données**, examinez l'étape automatique de changement de type : corrigez-la ou retirez-la pour que le repère vidéo ne devienne pas une heure de la journée. Les libellés peuvent varier selon la version du tableur ; le contrôle de la valeur reste le même.

Regardez E02. Vous devez lire **27:18** et les mots « gagnée » ou « perdue » doivent garder leurs accents. Ici, 27:18 signifie 27 minutes et 18 secondes de vidéo. 61:33 signifie 61 minutes et 33 secondes, pas une heure de rendez-vous.

Le résultat que vous devez voir : six colonnes séparées, 23 lignes de données et des repères vidéo fidèles au fichier. Si vous voyez 27:18:00 ou des accents illisibles, arrêtez avant le calcul. Reprenez la copie du CSV original et réimportez-la avec les bons choix. Changer seulement l'apparence d'une cellule déjà transformée peut laisser une valeur incorrecte. Si l'import reste indisponible, utilisez la table imprimée ou lisez le CSV comme du texte.

2 - distinguer les deux horloges



Le repère vidéo permet de retrouver des images. L'horloge du match permet de situer le jeu. Ces deux repères peuvent différer si l'enregistrement commence avant le coup d'envoi, coupe une pause ou saute une séquence. Une période doit aussi être précisée lorsqu'un affichage recommence à zéro.

Repérez **Deux-horloges.csv** dans le dossier Donnees. Dans H01, **08:23 de vidéo** correspond à **05:00 de match en période 1**. Pour H01, H02 et H03, soustrayez temps_match_cumule_s de temps_video_s, sur papier ou dans une colonne d'essai de votre copie. Comparez avec decalage_s : **503 − 300 = 203 s ; 1 638 − 1 440 = 198 s ; 2 880 − 2 700 = 180 s**. Les coupures fictives font varier le décalage ; n'appliquez

donc pas un décalage unique aux 22 événements.

Les repères 08:23 et 27:18 coïncident avec ceux d'E01 et d'E02 dans le CSV de touches. Cette coïncidence appartient aux données fictives : H01 et H02 ne sont pas des annotations de ces événements et ne prouvent aucune correspondance entre les tables.

Le résultat que vous devez voir : deux colonnes aux noms explicites et une période lorsqu'elle est connue. Dans le CSV de touches, seule la vidéo est renseignée. Vous pouvez convertir 27:18 en 1 638 secondes de vidéo ; vous ne pouvez pas inventer une minute de match ou un alignement GPS. Si une correspondance manque, écrivez « à vérifier » et gardez le repère disponible.

3 - comprendre ce qui a été conservé

Comparez E03 et E04 : le match, la phase, la zone, l'issue et le repère vidéo concordent. Dans ce cas pédagogique, E04 est une répétition confirmée d'E03 et la copie de travail conserve E03. Le simple fait que deux lignes aient le même repère à la seconde ne suffit pas, dans tous les exports, à confirmer un doublon : il peut exister deux actions proches ou des versions contradictoires.

E05 pose une autre question. Son événement existe, mais son issue n'a pas été renseignée. La copie corrigée affiche « à vérifier ». Cela ne signifie ni « gagnée » ni « perdue ». Gardez l'événement et notez séparément ce que vous ignorez.

Le résultat que vous devez voir : 23 lignes reçues, une répétition retirée de la table de travail et 22 événements conservés. Parmi eux, 14 sont gagnés, 7 perdus et 1 reste d'issue inconnue. Le taux de touches gagnées **parmi les issues connues** utilise 21 événements : **14/21 = 66,7 %**. Si vous obtenez un autre nombre, revenez aux ID et au classement avant de changer une formule.

Sur un autre export, si deux lignes de même clé se contredisent, conservez les deux versions dans un espace « à arbitrer » avec le motif. Ne choisissez pas la réponse la plus plausible pour faire disparaître l'alerte.

4 - lire les notes avant de les raccorder

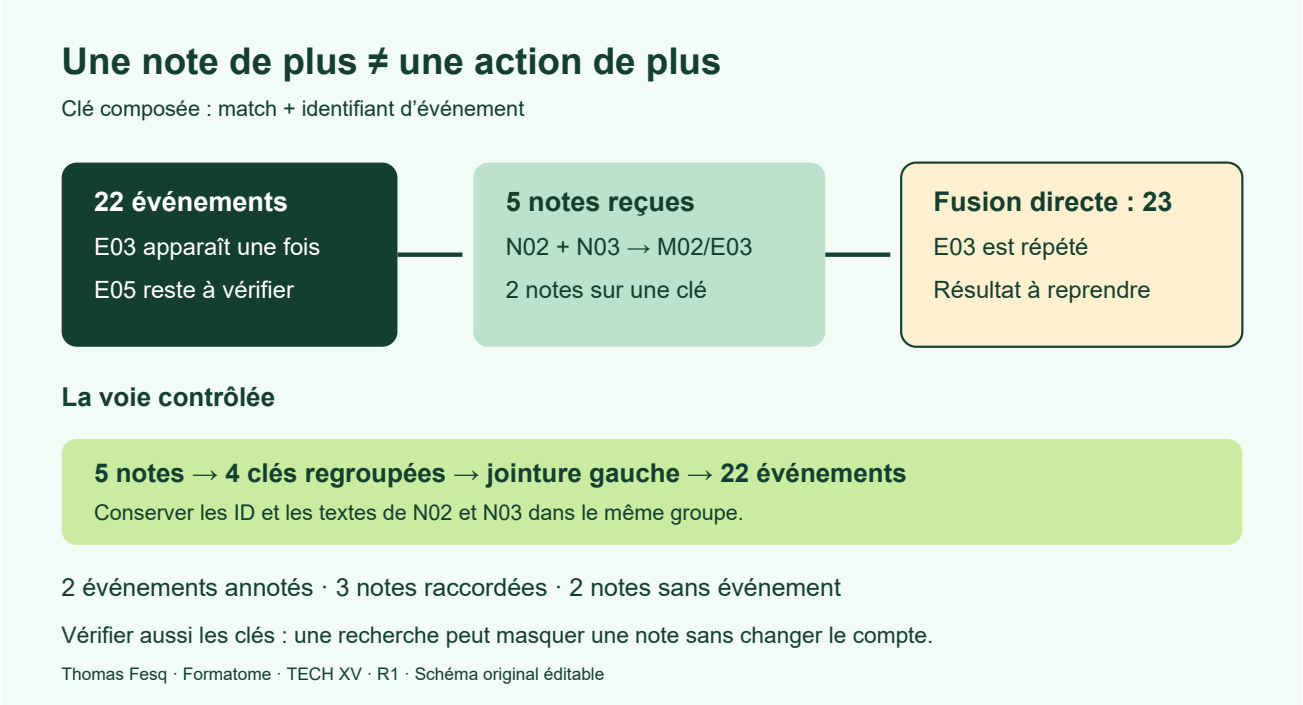
La table des notes contient cinq lignes. Chaque note possède son propre ID, un match, un identifiant d'événement et un texte. N02 et N03 concernent toutes deux M02/E03 ; elles apportent deux commentaires différents. Elles ne doivent pas être supprimées comme si l'une avait été copiée par erreur.

Pour retrouver un événement, utilisez **le match et l'identifiant ensemble**. Le petit signe |, si vous formez une clé visible, sert seulement à séparer les deux parties : M02|E03. Vous pouvez aussi sélectionner directement les deux colonnes dans l'outil de rapprochement. Leur ordre doit être le même de part et d'autre et leurs valeurs doivent être du texte.

Regardez N05 : elle vise M02/E01. E01 existe dans M01 ; ce n'est pas une correspondance. Une recherche sur le seul identifiant pourrait attacher la note au mauvais match.

Le résultat que vous devez voir : cinq notes, quatre clés différentes et deux notes sur la clé M02/E03. Si vous ne savez pas ce qu'une ligne de droite représente ou si sa clé est unique, ne lancez pas encore la fusion. Comptez les notes par clé et expliquez le cas qui se répète.

5 - ajouter les commentaires sans ajouter d’actions



Avant la fusion, écrivez votre attente : **la table finale doit toujours représenter 22 événements**. Une jointure gauche protège les événements sans note, mais elle peut créer plusieurs lignes lorsqu’un événement trouve plusieurs réponses à droite. Ici, une fusion directe des cinq notes produirait 23 lignes : E03 apparaîtrait deux fois.

La solution montrée en cours consiste à regrouper les notes par match et identifiant. Pour M02/E03, gardez les deux textes et leurs ID dans la même cellule, en indiquant qu’il y a deux notes. La table regroupée contient alors quatre lignes, une par clé. Une autre solution consiste à conserver les notes dans une table séparée ; le choix dépend de la restitution attendue.

Geste facultatif si Power Query est déjà disponible. Travaillez sur une copie de la requête des notes, avec match, id_evenement, id_note et note en texte.

1. Choisissez **Regrouper par** → **Avancé**, ajoutez les deux regroupements match et id_evenement, puis l’opération **Toutes les lignes**, nommée lignes.

2. Par **Ajouter une colonne** → **Colonne personnalisée**, créez groupe_trie avec Table.Sort([lignes], {{"id_note", Order.Ascending}}).

3. Ajoutez les trois colonnes ci-dessous. Les formules M gardent leurs noms anglais, même dans une interface française.

Nouvelle colonne	Formule personnalisée
ids_notes	Text.Combine([groupe_trie][id_note], " / ")
notes	Text.Combine([groupe_trie][note], " / ")
nombre_notes	Table.RowCount([groupe_trie])

Conservez les deux clés et ces trois sorties ; retirez ensuite les colonnes de tables imbriquées lignes et groupe_trie. Vérifiez quatre clés, cinq notes au total et, pour M02/E03, N02 puis N03, leurs deux textes et

nombre_notes = 2. Si le geste n’est pas disponible, filtrez M02/E03 et recopiez les deux ID et textes dans une cellule de votre table de regroupement ; cette reprise manuelle convient. La documentation Regrouper par, Text.Combine et Table.RowCount décrit ces opérations ; contrôlez leur résultat dans votre version.

Partez ensuite de la table des événements et choisissez un rapprochement **externe gauche** avec les notes regroupées. Sélectionnez les deux champs dans le même ordre, puis ajoutez seulement les textes, les ID des notes et leur nombre. Si votre tableur ne propose pas de jointure, retrouvez les clés exactes avec un filtre et recopiez les deux correspondances vérifiées. Une recherche exacte peut être utilisée après avoir vérifié qu’il ne reste qu’une ligne de notes par clé.

Le résultat que vous devez voir : 22 lignes et 22 clés d’événement distinctes, avec 2 événements annotés et 20 sans note. E03 porte bien les deux commentaires. E05 reste présente, avec son issue à vérifier. Si vous voyez 23 lignes, revenez à la version avant fusion et cherchez d’abord la clé répétée dans les notes ; ne supprimez pas une ligne au hasard dans le résultat.

6 - retrouver aussi les notes qui n’ont rien rejoint

La table des événements enrichis ne montre pas toutes les difficultés. Elle conserve les événements de gauche ; une note visant un événement inexistant n’y entre pas. Pour la retrouver, parcourez les notes dans l’autre sens et cherchez chaque clé dans les événements.

Gardez les notes sans correspondance dans une table dédiée avec leur ID, leur clé, leur texte et une suite à donner. Ici, N04 et N05 doivent rester visibles. « Demander la source et vérifier le match » est une suite acceptable ; créer un événement pour y accrocher la note ne l’est pas.

Le résultat que vous devez voir : chaque note est soit raccordée, soit conservée parmi les notes sans événement. Ne confondez pas « événement sans note » et « note sans événement » : la première expression part des événements, la seconde part des notes.

7 - garder une trace et essayer un retour arrière

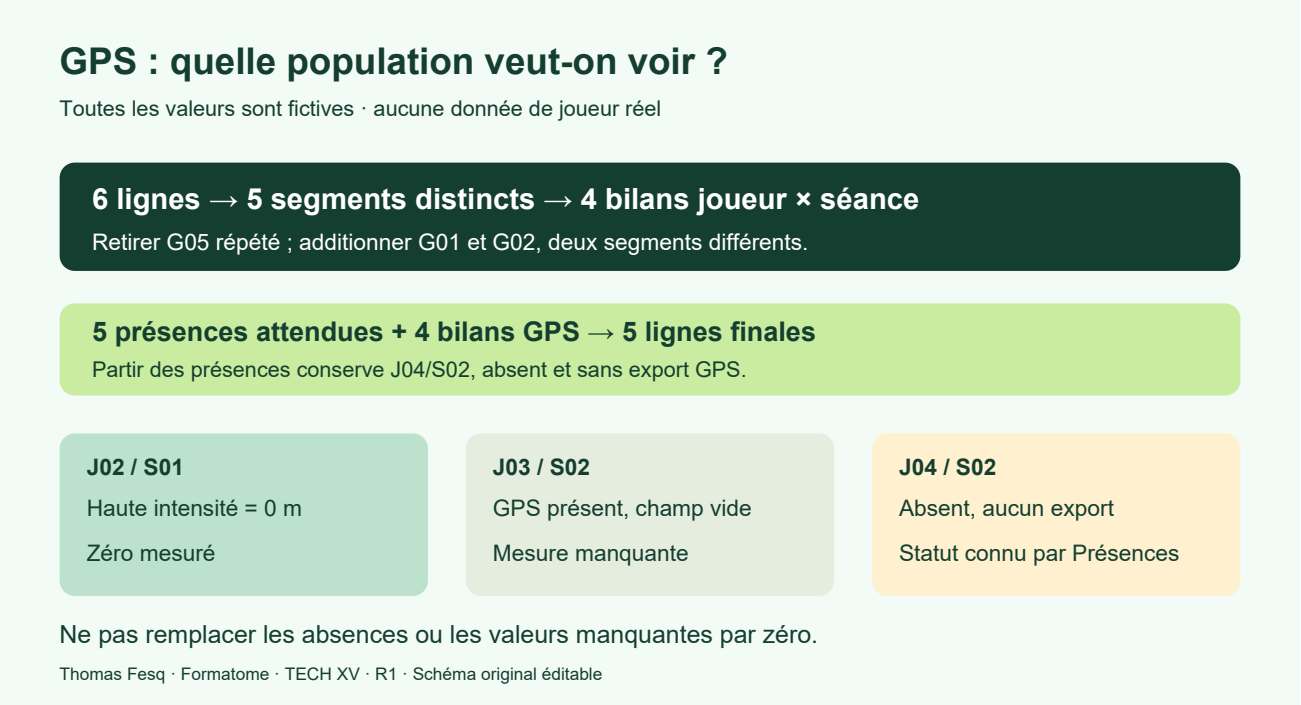
Donnez à chaque étape importante un nom de version simple. **Journal-versions-vide.csv** et l’onglet Journal utilisent dix colonnes : version, entree, grain, cle, modification, nombre_avant, nombre_apres, controle, decision, version_retour. Le grain dit ce que représente une ligne. Notez les exceptions dans controle et leur traitement dans decision ; version_retour indique la copie conservée à reprendre. Par exemple : **V01 reçu, 23 lignes** ; **V02 contrôlé, 22 événements** ; **ESSAI03 fusion directe, 23 lignes, résultat à reprendre à partir de V02**.

Faites maintenant le petit exercice de retour arrière : gardez ESSAI03 comme trace, ouvrez de nouveau V02 et vérifiez ses 22 événements ainsi que la présence d’E05. Marquez dans le journal « retour à V02, essai non retenu ». Corrigez ensuite la table des notes et produisez **V03 enrichi**, avec 22 événements. Le numéro de version ne remplace pas un contrôle : une version plus récente peut être moins fiable.

Le résultat que vous devez voir : vous savez quelle version utiliser et pourquoi. Vous pouvez rouvrir la version précédente, expliquer l’erreur et retrouver le résultat corrigé. Si vous ne retrouvez plus l’entrée ou si la seule copie a été écrasée, arrêtez les modifications et repartez de l’original conservé. Ne présentez pas un retour arrière comme réussi tant que vous n’avez pas retrouvé les valeurs et le nombre attendus.

En rejouant le même import, remplacez le résultat précédent au lieu de l’ajouter une seconde fois. Le nombre doit rester 22. On appelle cela une opération **idempotente** : la répéter sur la même entrée ne change pas son résultat. Si vous obtenez 44 lignes, vérifiez si le même lot a été ajouté deux fois.

8 - reconnaître le même problème dans un export GPS



Le petit exemple GPS change le grain. Dans ces fichiers fictifs, S01 et S02 sont des identifiants de séances GPS ; ils ne désignent pas les cours des 9 et 12 novembre. Deux segments peuvent appartenir au même joueur et à la même séance. G01 et G02 sont les deux segments de J01/S01 : il faut les additionner pour produire un bilan de séance. G04 et G05 répètent, au contraire, le même segment fictif ; une seule version est conservée après contrôle.

Après ce contrôle, les cinq segments distincts donnent quatre bilans **joueur × séance**. Les notes de présence décrivent cinq couples attendus, dont J04/S02, absent et sans export GPS. Si vous voulez voir toute la population prévue, partez des présences et ajoutez les bilans GPS par la clé joueur + séance. Vous obtenez cinq lignes, dont une sans mesure GPS. Partir seulement du GPS ne ferait pas apparaître cette absence.

Observez la haute intensité : J02/S01 a un **0 mesuré** ; J03/S02 a un champ **manquant**. Le zéro indique que la mesure est connue et vaut zéro. La case vide indique que la valeur n'est pas connue dans l'export. Remplacer les deux par zéro ferait disparaître cette information.

Avant toute somme par joueur et séance, comptez les segments distincts du groupe, puis les valeurs numériques connues de la mesure, **zéro inclus**. La somme n'est complète que si le groupe contient au moins un segment et si les deux comptes sont égaux. Sinon, gardez « manquante » et le nombre de segments incomplets. Une simple SOMME.SI.ENS peut renvoyer **0 lorsque toutes les cellules à additionner sont vides** : ce résultat ne prouve pas un zéro mesuré. Pour J03/S02, un segment et aucune haute intensité connue donnent donc « manquante » ; pour J02/S01, un segment et une valeur connue égale à zéro donnent 0. La même règle conserve un total manquant si une partie seulement des segments est renseignée.

Le résultat que vous devez voir : une ligne par couple joueur × séance, des segments regroupés, la répétition documentée et trois états lisibles : valeur mesurée, mesure manquante, absence signalée par la présence. Si vous obtenez plus de cinq lignes après ajout aux présences, vérifiez le regroupement des segments et la clé. N'alignez pas une courbe GPS sur une vidéo sans avoir établi la correspondance des horloges.

Contrôle de huit minutes

Prenez une minute pour lire les quatre situations, puis quatre minutes pour préparer vos réponses. Vous pouvez réfléchir sans rendre de fichier. Pour chaque situation, donnez une décision et une raison ; les trois dernières minutes servent à la correction collective.

1. Les 22 événements deviennent 23 lignes après une jointure sur match et ID. Peut-on annoncer 23 événements ? Quel contrôle faire d’abord ?

2. N05 porte M02/E01. Peut-on la raccorder à E01 de M01 ? Pourquoi ?

3. E05 existe, mais son issue manque. Entre-t-elle dans le nombre d’événements ? Dans le dénominateur du taux sur issues connues ?

4. J02/S01 a 0 en haute intensité ; J03/S02 a une valeur vide. Peut-on écrire 0 pour les deux ?

Reprise facultative après la séance

Pour une reprise courte, refaites le rapprochement des cinq notes et gardez la table à 22 événements, les deux notes sans événement et le journal. Faites relire un ID et un compte par un collègue, ou relisez-les après avoir masqué votre première réponse. Si le résultat ne concorde pas, revenez à la dernière version vérifiée avant de recommencer la fusion.

Pour aller plus loin, consacrez 15 à 20 minutes au cas GPS. Expliquez en quatre phrases ce qui change lorsque vous partez des segments ou de la population prévue, puis comparez une valeur manquante et un zéro mesuré. Une table papier accompagnée de cette explication convient.

Vous avez terminé lorsque quelqu’un peut retrouver les sources, comprendre vos choix et refaire le contrôle sans une explication orale supplémentaire. Les derniers échanges de séance sont réservés aux questions métier ; les dates futures et les éventuels accès supplémentaires restent à confirmer par le formateur.