Un motif, huit contrôles — le digest portable

Extrait du projet « Domaine du Mons » pour servir ailleurs. Rien de ce qui suit ne dépend du lieu. Septembre 2026.

Famille	*Généricité*	*Statut*
Le digest	GÉNÉRIQUE	EXPLORÉ — production du corpus, non validée par une décision
Écrit le	*Licence*	
18 septembre 2026	CC BY-SA 4.0 (documentation) / MIT (code)	

Un seul motif de défaillance, six fois

Six enquêtes indépendantes — juridique, climatique, hydrologique, associative, cartographique, topographique — ont produit six erreurs qui n'ont qu'une forme :

> **Un résultat propre, plausible, et faux — sans que rien dans la sortie ne le signale.**

Le mode de défaillance ne se voit pas parce qu'il ne casse rien. La requête est juste, le code tourne, le chiffre s'affiche, et il ressemble à un chiffre qu'on connaît. C'est précisément cette ressemblance qui le fait passer.

	l'erreur	**ce qui l'a révélée**
Le panorama	Une recherche par les cas trouve les solutions adoptées, pas les solutions disponibles. La recommandation juridique s'est inversée entre deux passes.	Une seconde passe d'un autre angle
Le seuil emprunté	Un seuil repris d'une source sans recalage sur la donnée locale décalait un stade phénologique d'un mois.	Le résultat était inexplicable sur le terrain

L'indicateur en butée	Une colonne officielle donnait la **même valeur à deux états différents**. Deux faux épisodes dans un tableau publié.	Une valeur impossible : huit écarts-types
Le registre trop étroit	Compter des associations dans un fichier qui n'immatricule que celles ayant une existence financière : 1 952 au lieu de 8 340.	La recherche d'un autre champ
Le code d'erreur mal lu	Un 504 pris pour « votre requête est mauvaise » alors qu'il dit « l'instance est chargée ». Une journée perdue à chercher un contournement.	La patience, le lendemain
L'inventaire aberrant	Un algorithme correct sur une donnée correcte : 66 % du volume inventorié était un artefact d'ouvrage d'art.	Ramener le total à une grandeur familière

Les huit contrôles qui les attrapent

Ils coûtent des minutes. Chacun a été écrit après avoir manqué.

1. **Demander au registre son critère d'entrée — et sa date.** La question n'est pas « cette source est-elle à jour ? » mais **« qu'est-ce qui fait entrer une ligne dans ce fichier ? »**, avant la première requête. Un registre ne se trompe jamais sur ce qu'il contient ; il se tait sur ce qu'il ne contient pas. Une source renvoie une liste complète — complète de ce qu'elle contient.

Préférer le registre dont l'**objet même est de dénombrer** la population cherchée, même s'il est plus pénible d'accès qu'une API commode portant une population voisine.

Et **un catalogue de sources se date, sinon il ment** : trois services publics ou parapublics sur lesquels ce corpus s'appuyait ont fermé ou changé d'adresse dans la même année.

2. **Deux sources sur tout dénombrement.** Un chiffre unique n'est pas un chiffre vérifié, aussi propre soit sa provenance. Quand l'écart est notable, **il est lui-même une information** : le rapport entre deux comptages mesure quelque chose, et cette mesure n'existait pas avant de trouver l'erreur.

3. **Le test de saturation.** Trier la colonne et chercher des **valeurs identiques, à la précision publiée, sur des entrées différentes**. La question n'est pas « la valeur est-

elle grande ? » mais « deux entrées différentes donnent-elles la même sortie ? ». Une valeur impossible est un cadeau ; une butée discrète ne se signale pas.

Et contrôler toute grandeur dérivée **contre la grandeur brute dont elle dérive**.

4. **La conversion en grandeur familière.** « 3,5 millions de m³ » ne dit rien ; « quatre mètres de profondeur moyenne sur un plateau » se réfute en une seconde. Publier avec le total **la part portée par ses trois plus gros éléments** : une somme dominée par deux objets décrit ces objets, pas la population.

5. **Vérifier qu'un total égale la somme de ses parties.** Le contrôle le plus bête est celui qui trouve. Il a ramené un parc de logements de 292 à 146.

6. **Écrire les hypothèses tacites du modèle sur le monde.** Un calcul d'écoulement de surface suppose que l'eau ne circule que par la surface ; toute buse viole cette hypothèse sans que rien ne le signale. Les hypothèses non écrites ne sont jamais testées. Croiser avec une source indépendante, **même approximative**.

7. **Distinguer les codes d'erreur qui parlent de vous de ceux qui parlent de l'autre.** 400, 404, 422 appellent une correction. 429, 502, 503, 504 appellent une attente qui double. Réessayer vite est la seule réaction qui aggrave la panne qu'on constate. Et mesurer la liaison avant de l'accuser. **Un quota n'est pas une panne** : c'est une information sur le volume demandé, souvent corrigible par le cadrage plutôt que par le code.

8. **Éprouver les deux chemins.** Un dossier écrit puis relu n'est pas le même objet : en mémoire une année est un entier, relue depuis un fichier c'est une chaîne. Le défaut n'apparaît que sur l'un des deux chemins, et une relecture ne le voit jamais — seul un essai de bout en bout, dans les deux sens, l'attrape.

Et, avant tout cela : vérifier la mesure avant de croire ce qu'elle dit. Quand un contrôle annonce qu'une chose ne marche pas, suspecter d'abord le contrôle. Deux correctifs ont été déployés sur un faux diagnostic parce qu'un harnais réutilisait la même adresse et mesurait l'état précédent. Une sonde de surveillance a signalé une source morte qui allait très bien, parce qu'elle lisait mal une réponse. **Un outil qui crie au loup finit par ne plus être lu** : lui déclarer les réponses attendues, et le faire réessayer avant de conclure.

Rendre les corrections cumulatives

Les huit contrôles ci-dessus trouvent des erreurs. Sans ce qui suit, ils les trouvent **à chaque fois**, sur chaque nouveau territoire, chaque nouveau jeu de données. Quatre gestes, le jour même, transforment un travail répété en travail qui progresse.

Une correction sans test n'est pas une correction. Elle est corrigée pour ce cas-là. Extraire du brut fautif le fragment minimal qui portait l'erreur, l'enregistrer tel quel, écrire l'essai avec sa date, son objet, ce que la sortie affichait et ce qui l'a démasquée. Ces essais-là valent plus que les autres : **ils décrivent des erreurs qui se sont produites, pas des erreurs imaginées.**

Ce qui ne se rejoue pas n'est pas reproductible. Garder chaque réponse brute telle quelle, avec son empreinte, et une ligne de provenance — adresse, producteur, licence, date. L'épreuve tient en une commande : refaire tout le travail hors ligne, depuis le cache. La première collecte prend des minutes, le rejeu quelques secondes. C'est ce qui distingue un travail vérifiable d'un travail à croire sur parole — et ce qui permet à une source fermée de ne pas emporter l'analyse avec elle.

Sonder les sources, et comparer au sondage précédent. Un état daté qui distingue trois choses que rien d'autre ne distingue : un refus temporaire (quota), un refus durable (la source a changé), un injoignable (souvent la liaison). La comparaison est ce qui a de la valeur : elle affiche d'elle-même ce qui a bougé depuis la dernière fois.

Tenir un journal de ce sur quoi l'outil a réellement tourné. Une ligne par cas : date, échelle, ce qui a été collecté, ce qui a manqué, ce qui a été signalé. C'est la seule façon de savoir ce que « ça marche » veut dire — sur combien de cas *réellement différents*, et non sur combien de fois sur le même.

Ce qui se mesure d'une fois à l'autre : part de ce qui se collecte sans intervention (doit monter) · manques par cas (doit baisser) · corrections trouvées après coup (doit baisser, et chacune a son test) · délai entre collecte et livraison (doit baisser sans que les corrections montent). Un cinquième compte davantage et ne se mesure pas : **le résultat a-t-il servi à décider quelque chose.**

Quatre règles d'écriture

Séparer la matière du commentaire. Tant que les deux sont mêlés, le commentaire précède le contrôle. Une fois séparés, le texte ne peut citer qu'un chiffre déjà contrôlé, et la matière n'a plus à être jolie.

Ne rien imputer. Une valeur absente vaut « nd » jusque dans le texte final. Une tendance ne s'annonce que si l'intervalle de confiance exclut zéro — sinon on écrit « pas de tendance détectable ». Une source qui échoue se déclare, elle ne se comble pas en silence.

Publier l'incertitude coûte moins cher que publier une certitude fausse. Une page qui dit « à l'étude » n'a jamais à être rétractée. Marquer dans le texte ce qui est vérifié et ce qui ne l'est pas, et conserver ces marques : sans elles, une information de seconde main devient indiscernable d'un texte contrôlé au bout de trois mois.

Publier le doute plutôt que de trancher en silence. Un tri silencieux n'est pas contestable ; un doute affiché l'est. C'est la différence entre un travail qu'on peut vous reprendre et un travail qu'il faut vous croire.

Trois outils prêts à l'emploi

La grille d'intérêt général

Trier un tissu associatif sur ce que les organisations **disent faire**, et non sur leur statut, leur code d'activité ou leur nom.

Cinq faits font entrer — un seul suffit, s'il est net : *ouverture* (mettre un lieu ou une ressource à disposition de non-membres), *transmission* (former, accompagner des personnes extérieures), *commun* (prendre soin d'un bien qui n'appartient à personne en particulier), *solidarité* (aider des personnes en difficulté qui ne sont pas les membres), *territoire* (soutenir une activité non marchande ou coopérative).

Trois motifs font sortir, et l'emportent sauf action déclarée vers l'extérieur *en plus* : *cercle*, *économique*, *politique*.

Trois verdicts : retenu, douteux, hors — et les douteux sont publiés.

Trois extensions, nées de l'application à un corpus de nature différente : qualifier **la nature de l'acteur** avant d'appliquer les faits, car pour un acteur marchand la transaction ne compte pas et il faut un fait en plus de la vente ; publier **la base de preuve** avec chaque verdict — objet déclaré, réseau tiers qui atteste une pratique, ou rien, une fiche sans objet et sans réseau ne pouvant jamais être retenue ; distinguer **hors périmètre** de hors intérêt général.

Écrire la grille **avant** la première lecture, et ne pas l'ajuster après coup. Sa valeur est dans les cas limites : sans eux, ce n'est qu'une déclaration d'intention.

Le diagnostic de territoire

Produire, à partir des données publiques françaises, un dossier de données daté, sourcé, contrôlé et rejouable — puis le rapport qui s'écrit dessus, puis la boucle qui fait que le territoire suivant coûtera moins cher.

Deux échelles, qui ne se confondent pas : un lieu (ferme collective, site habité) et une commune rurale. Une étude tient dans un fichier de configuration ; changer de territoire, c'est changer ce fichier. Le brut n'est jamais modifié. Les contrôles tournent à chaque collecte. Rien n'est imputé.

Contre-indication majeure : la chaîne **pré-localise, elle ne qualifie pas**. Zones humides, périmètres de captage, débroussaillement — le critère réglementaire reste pédologique, floristique ou administratif, et le terrain reste à faire.

La cartographie honnête

Cartes web sans requête sortante, à partir des données publiques françaises. Quatre modes de rendu, trente-cinq règles opposables, et trois refus codés en dur : une choroplèthe d'effectifs bruts, un rayon proportionnel à la valeur, des zones sans donnée coloriées dans la palette. Ils se contournent, en l'écrivant dans la fiche de la carte — donc à un endroit où le lecteur le verra aussi.

CC BY-SA 4.0 pour la documentation, MIT pour le code.

CoMons, Collectif du Mons — Un motif, huit contrôles — le digest portable. Fiche du Grain, CC BY-SA 4.0.
Version en ligne : comons.actitude.org/grain-digest.html