

CANEVAS DE DISCUSSION

L'ontologie processuelle de la donnée en ses termes

David Bories

Canevas de discussion — L'ontologie processuelle de la donnée en ses termes

Conçu et dirigé par David Bories — Albi, France.
Rédigé en utilisant l'intelligence artificielle générative.

Octobre 2026.

0.2.0-poche

contact@artisanat-de-la-donnee.fr

Table des matières

Canevas de discussion	1
<i>L'ontologie processuelle de la donnée en ses termes</i>	1
Préambule — Mode d'emploi	1
Temps I — Fabrication : la donnée naît	2
Temps II — Circulation : la donnée voyage	4
Temps III — Réception : la donnée est lue	6
Notions transversales	7
Questions vives — points de controverse	8

Canevas de discussion

L'ontologie processuelle de la donnée en ses termes

David Bories — Travail en cours, 2026

Préambule — Mode d'emploi

Ce document n'est pas fait pour être lu d'un trait, ni à voix haute. C'est un **canevas** : une trame de termes sur laquelle on brode oralement. Chaque entrée nomme une notion de l'ontologie processuelle, la cadre en quelques phrases, et ouvre une amorce de discussion. Le développement, lui, se fait à la parole : le canevas met sous la main le vocabulaire exact ; à celui qui parle de le déployer.

Il se distingue en cela du *manifeste* — qui remplace les termes par des images pour donner envie — et du *volume d'architecture* — qui énonce les thèses, les relie et renvoie aux tomes pour leur démonstration. Ici, on cadre et on ouvre. Le texte reste volontairement bref pour laisser à l'oral toute sa place.

Une règle gouverne l'usage des termes en discussion : la distinction du **fond** et de la **forme**. Le corpus n'affirme que ce qui est vrai de la donnée indépendamment des techniques du moment (le fond) ; il s'interdit de prescrire les mises en œuvre concrètes (la forme), qui se périment avec les outils. En débat, on défend donc le concept, jamais une implémentation particulière : les exemples contemporains illustrent, ils ne fondent rien.

Les termes sont rangés selon les trois temps de la donnée — **fabrication, circulation, réception** — suivis des notions transversales et d'une série de questions vives.

Temps I — Fabrication : la donnée naît

Flux du réel. Le monde en écoulement, indifférent, qui excède toujours les découpages qu'on y opère et ne s'arrête jamais pour se laisser compter ; sa continuité et son inépuisabilité sont des cas importants, non des axiomes. Il n'est pas fait de données : il est fait d'événements qui passent. La donnée n'y préexiste pas ; elle naît d'un acte qui la constitue. À *débattre* : si le réel n'attend pas qu'on le mesure, en quoi cela change-t-il l'idée que « les données sont là, il suffit de les recueillir » ?

Prélèvement et acte d'inscription. La donnée naît d'un geste : quelqu'un perçoit, distingue, interprète, vise un fragment du flux, et l'inscrit sur un support — cinq facultés. Avant ce geste, pas de donnée ; après, une donnée constituée, fixée sur son support. Le geste est intentionnel et situé : il retient une chose et en laisse filer mille. Le prélèvement est la voie **observationnelle** de cette constitution ; une donnée peut aussi naître d'un acte qui institue (**institutionnelle** : le vote compté, le solde arrêté) ou d'autres données (**dérivée** : une moyenne, une estimation). À *débattre* : qui prélève, avec quelle intention — et que laisse-t-on hors champ à chaque inscription ?

Flux d'observation, le canal. Non pas un prélèvement isolé, mais une série réglée de prélèvements, cadencée selon une intention, qui restitue par petites touches le mouvement d'un fragment du réel. C'est le canal qu'on creuse dans le fleuve : infiniment plus pauvre que lui, mais maîtrisable — on le suit, le rejoue, le relie, le confie, le comble. À *débattre* : passer du « seau » (la donnée isolée) au « canal » (le flux réglé) — qu'est-ce que cela déplace dans la manière de travailler la donnée ?

Intention et critères. Un canal naît d'une intention, d'où découlent les critères qui le règlent : *quoi* prélever, à quelle *cadence*, avec quelle *fidélité*, avec quel *corpus* de contexte, sur quel *périmètre*. Ces critères sont la signature de l'intention. Un canal ne ment pas sur le réel ; il ne montre que ce que ses critères admettent. À *débattre* : devant un chiffre, faut-il d'abord demander « que dit-il ? » ou « quelle intention l'a fait naître ? »

Les deux temps de la donnée. La donnée naît de deux gestes : un **acte**, qui la fixe à un instant pour toujours ; un **support**, qui la fait durer. D'où la résolution d'un paradoxe : la donnée est à la fois figée (son contenu ne change pas) et vieillissante (l'écart au réel se creuse). Elle a deux temps — l'un éternel, l'autre qui dérive — qui ne mesurent pas la même chose. À *débattre* : comment une chose figée peut-elle vieillir ? Et que vieillit-il, au juste, si ce n'est pas le contenu ?

Fraîcheur et péremption. La péremption n'est pas une altération de la donnée, mais la relation qui se creuse entre une donnée immobile et un réel mobile. Sa vitesse n'appartient pas à la donnée : elle appartient au fragment de monde prélevé. La péremption n'est donc pas une loi uniforme : la **validité** d'une donnée se juge **pour un usage** — la même inscription peut être périmée pour l'alerte et exacte pour l'histoire. La fraîcheur ne se

rattrape pas — on ne rajeunit pas une donnée, on en prélève une nouvelle. À *débattre* : si la valeur d'une donnée se juge pour un usage, qu'est-ce que cela fait à l'idée de « patrimoine de données » ou de donnée comme « nouveau pétrole » ?

Sources stables. Cas où la péremption est si lente, *relativement à l'usage*, qu'elle est imperceptible : grammaire d'une langue, constante physique, fait historique établi. Stabilité non comme absence de mouvement, mais comme rapport entre deux vitesses. Solidité réelle — à condition de la tenir pour une longue durée, jamais pour une éternité. À *débattre* : sur quelles « sources stables » une société bâtit-elle, et que risque-t-elle à confondre une source stable avec une source qui se périme vite ?

Optimum observationnel. Plus n'est pas mieux. Entre le **sous-prélèvement** (on passe sous le seuil, on perd le signal) et la **sur-ingénierie** (on prélève au-delà du nécessaire, on ne gagne rien sinon du coût), il existe une zone où le canal est creusé au mieux. L'optimum est le réglage conjoint de la **cadence**, du **corpus** et de la **probité** qui maximise l'écart favorable entre bénéfice et coût, pour l'usage visé. Il n'existe pas d'optimum universel : la question — *le réel observé peut-il changer significativement entre deux prélèvements ?* — se repose à chaque fois. À *débattre* : contre le réflexe « toujours plus de données », où se situe le « juste assez » — et qui en juge ?

Saut systémique. Au-delà d'un seuil, l'accumulation d'améliorations cesse d'améliorer et se met à transformer : le régime change de nature. Des images assez rapprochées ne donnent pas une suite de photos plus dense — elles donnent le cinéma. Une transmission assez accélérée ne donne pas un courrier plus prompt — elle donne l'instantanéité. Le saut se mesure à ce qu'il rend possible (et qui n'existait pas) et à ce qu'il rend caduc (et qui semblait acquis). Le concept pointe la possibilité ; il ne prédit ni quand ni où (forme). À *débattre* : quels « sauts systémiques » avons-nous déjà traversés sans les nommer — et qu'ont-ils rendu caduc ?

Données consolidées et flux dérivés. D'un flux d'observation on peut extraire des données consolidées — moyennes, tendances, statuts — à haute valeur d'usage ; et un flux peut en engendrer d'autres, de second ordre, plus denses. Les flux engendrent des flux. On les stocke, rejoue, épure, enrichit, fusionne, transmet, supprime — autant de prises gagnées sur un réel qui, lui, se perd à chaque instant. À *débattre* : à mesure qu'on consolide et qu'on dérive, que perd-on du prélèvement d'origine — et le sait-on encore ?

La boucle, ou l'observation non neutre. On observe pour décider, on décide pour agir, et l'action transforme le réel — qui appelle de nouvelles observations. Observer pour agir, c'est entrer dans le mouvement du réel et le modifier. La donnée n'est jamais un simple miroir : elle participe de ce qu'elle prétend décrire. À *débattre* : où voit-on une observation transformer ce qu'elle observe — et qui assume cette transformation ?

Temps II — Circulation : la donnée voyage

Acte d'échange. Le deuxième geste fondateur, symétrique de l'acte d'inscription : mettre une donnée en mouvement vers un destinataire. Il étend la disponibilité de la donnée et sa possibilité d'usage ; il n'accomplit ni la donnée ni sa lecture. Échanger n'est jamais un transfert neutre ; le geste appelle quelqu'un et engage une responsabilité. *À débattre* : qu'engage-t-on, exactement, au moment où l'on décide de transmettre une donnée ?

Support de transport. Toute donnée en mouvement est, à chaque instant, inscrite sur un support physique : il n'y a pas de transport immatériel. Le support remplit trois fonctions minimales — déplacer, transmettre, rendre accessible. *À débattre* : que change le fait de rappeler qu'aucune donnée ne circule « hors-sol », sans matière ni énergie ?

Double navigation. En se déplaçant, la donnée accomplit deux trajets à la fois : une **navigation spatiale** (le support change de lieu — voulue, maîtrisée) et une **navigation temporelle** (le réel s'éloigne pendant le trajet — subie). Transporter, c'est toujours, en plus, laisser vieillir. *À débattre* : à quel moment le « voyage dans le temps » d'une donnée transportée devient-il un problème pratique ?

Transport comme rapport entre deux déplacements. Le transport n'est accompli que si deux mouvements se rejoignent : celui de la donnée vers le lecteur, et celui du lecteur vers la donnée. Son critère d'accomplissement : la **part du lecteur** rendue praticable — qu'il ait le temps, le moyen, le code d'en faire quelque chose. *À débattre* : déposer une donnée « à portée » de quelqu'un qui ne peut pas la déchiffrer — est-ce l'avoir transmise ?

Fidélité du transport. Une donnée peut arriver corrompue, tronquée, ou intacte mais détachée du contexte qui la rendait lisible. La fidélité n'est jamais garantie d'avance : le support a-t-il tenu, le format survécu, le sens transmis avec ce qu'il faut pour le comprendre ? *À débattre* : qu'est-ce qui se perd le plus souvent en route — la donnée elle-même, ou ce qui permettait de la lire ?

Multipliation par copie. Transmettre une donnée, le plus souvent, ce n'est pas la déplacer, c'est la copier : l'original reste, une réplique apparaît. La donnée ne se multiplie pas — il y a toujours un seul fait —, ses **porteurs** se multiplient. Elle se partage comme un fait, sans déposséder qui la transmet. D'où une diffusion potentiellement illimitée, aux deux visages : communs ouverts d'un côté, fuites irrattrapables de l'autre — la même loi vue de deux côtés. *À débattre* : la générosité (le commun) et la fuite procèdent de la même propriété — comment penser l'une sans nier l'autre ?

Une réponse praticable : la borne est dans le geste, pas dans la donnée. On choisit ce qu'on tend ; le pli reste clos pour qui le porte ; rien ne s'inscrit chez l'autre sans son geste d'accueil. La générosité et la fuite procèdent

de la même propriété — mais la générosité est un geste borné trois fois, et la fuite est ce qui reste quand une des trois bornes manque.

Propagation. L'agrégat des transports : la donnée déposée partout à la fois. Rendre commun, ce n'est pas multiplier la donnée, mais abaisser le coût de la part du lecteur jusqu'à ce que le plus grand nombre puisse l'accomplir — ce que fit, historiquement, la bibliothèque publique. À *débattre* : « rendre commun », est-ce ouvrir l'accès à la donnée, ou rendre praticable la lecture pour le plus grand nombre ?

Le commun du canal. Le commun n'est pas un lieu où la donnée s'accumule : c'est ce qui abaisse le coût de son voyage entre lecteurs situés, sans en multiplier les porteurs au-delà de ceux qui la reçoivent. Un commun qui concentre a déjà cessé d'être un commun : il est devenu une source — avec le pouvoir d'une source. À *débattre* : un commun peut-il rester un protocole, ou tout protocole finit-il par sécréter son dépôt ?

« **Propager la donnée n'est pas propager la connaissance.** » La donnée arrive à portée ; l'information, la connaissance, la sagesse ne voyagent pas avec le signal. Elles exigent que le lecteur fasse sa part — déchiffrer, intégrer. Une société peut être traversée de données comme jamais sans en être plus savante. Aucun perfectionnement du transport n'accomplira la part du lecteur à sa place. À *débattre* : sommes-nous mieux informés parce que mieux abreuvés de données — ou est-ce une confusion ?

Empêchement du transport. Le rapport peut être empêché de deux manières, de fond différentes. L'empêchement **volontaire**, le secret : on ferme l'accès pour un intérêt propre. L'empêchement **involontaire**, l'incapacité : le destinataire ne peut pas faire sa part, faute de code, de moyen, d'accès. Remédier à l'incapacité sans forcer les secrets ouvre la place d'un artisanat de la donnée. À *débattre* : faut-il traiter de la même façon ce qui est caché par choix et ce qui reste illisible par manque de moyens ?

Torrent, sur-débit, auto-amplification. Multipliez canaux, copies, vitesses, destinataires : le ruisseau devient **torrent** — le régime où nous marquons le réel plus vite qu'aucun regard ne peut suivre. Vu du lecteur, c'est le **sur-débit** : il passe devant chacun infiniment plus que ce qu'il peut recevoir, toujours relatif à sa fenêtre finie. Le torrent grossit par **auto-amplification** : la circulation appelle la circulation, la boucle creuse son propre lit. À *débattre* : le sur-débit est-il un fait de la technique, ou de nos choix de prélever et de propager toujours plus ?

Sagesse collective. À l'échelle d'un collectif, s'abstenir ou infléchir un régime de circulation n'est jamais neutre — c'est une position, dès lors qu'elle est rendue visible et délibérée. Grille de lecture, non drapeau : elle décrit la structure d'un choix collectif (adopter ou refuser un régime), sans l'approuver. À *débattre* : décider collectivement de *ne pas* tout capter ni tout propager — est-ce une renonciation, ou une décision pleine ?

Temps III — Réception : la donnée est lue

Acte de lecture. Plus court que l'acte d'inscription : il mobilise percevoir, distinguer, **déchiffrer**, et s'arrête là. Il ne vise ni n'inscrit — ce serait produire une donnée nouvelle, non lire celle qui est là. Lire un thermomètre ne fabrique pas un second thermomètre. *À débattre* : où s'arrête la lecture et où commence une nouvelle inscription ?

Lecteur situé. Celui qui lit n'est jamais un regard de nulle part : il lit depuis un corps (ou, pour une machine, un entraînement), une attention finie, un corpus de connaissances qui lui permet de déchiffrer ceci et pas cela. La machine qui « lit » occupe une **position située de reprise ou de traitement**, avec sa propre fenêtre ; elle n'est **lecteur effectif** que si un sens s'actualise pour elle, ce que le corpus ne tranche pas — asymétrie à examiner, non frontière nette. Ce qui ne se délègue pas sans reste, c'est l'**imputabilité** : quelqu'un doit répondre de ce qui est retenu. *À débattre* : la même inscription ne livre pas la même chose à tous — qu'est-ce que cela implique pour l'idée de donnée « objective » ?

Fenêtre de lecture. La portion bornée de ce qui défile qu'un lecteur peut saisir, découpée par son corps, son attention, son savoir. La fenêtre n'est pas une prison mais la condition de toute lecture : c'est parce qu'elle découpe qu'elle montre. Les instruments (loupe, filtre, résumé) déplacent la fenêtre ; ils ne l'abolissent jamais, et chaque déplacement ouvre de nouveaux angles morts. *À débattre* : peut-on rêver d'une lecture « totale », ou la borne est-elle ce qui rend la lecture possible ?

La trace, ou la donnée. Sans lecture, la donnée demeure **constituée** : c'est l'acte d'inscription qui l'a fait exister ; la lecture, elle, l'**actualise**. Une donnée que nul ne lit n'est pas une trace qui attend : c'est une donnée faite, à qui il manque un lecteur — non actualisée, ou muette pour ce lecteur-ci. Elle ne redevient **trace** — une marque sur un support, privée de l'interprétation qui en faisait une donnée — qu'au cas-limite de la donnée morte. *À débattre* : d'une montagne de données stockées et jamais lues, que vaut ce qui demeure constitué sans être jamais actualisé ?

Donnée muette / donnée morte. La donnée **muette** a une interprétation, mais ce lecteur-ci ne la possède pas : état relatif et réversible (muette pour moi, lisible pour un autre, ou pour moi demain). La donnée **morte** est l'idéal-limite où plus aucun lecteur, nulle part, ne peut l'interpréter : l'interprétation est universellement perdue, il ne reste que l'inscription — c'est là seulement que la donnée retombe en trace. Le torrent fabrique en masse de la donnée non actualisée, un peu de mutité durable, et à la longue très peu de mort. *À débattre* : comment distinguer, dans ce qu'on n'a pas lu, ce qui est muet (récupérable) de ce qui est mort (perdu) ?

Les trois chemins. Trois façons de ne pas actualiser : une donnée n'est pas lue pour trois raisons, qui n'appellent pas la même réponse.

L'incapacité : on ne *peut* pas lire (code, moyen, accès manquants) — remédiable. Le **désintérêt** : on *pourrait* lire mais on ne le fait pas — affaire d'attention, non de moyen. La **péremption rapide** : la donnée n'a pas *attendu* — sa fraîcheur était trop brève ; périmée pour cet usage, elle demeure constituée. D'où l'énoncé central : **toutes les données non lues ne sont pas des échecs**. À *débattre* : la tâche du lecteur débordé n'est pas de tout lire, mais de discerner ce qui vaut d'être lu — selon quels critères ?

Trois postures. Devant le sur-débit, le lecteur dispose de trois gestes, combinables, chacun avec son coût. **Cadrer** : choisir délibérément ce qu'on regarde et ce qu'on laisse passer. **Déléguer** : confier sa lecture à un autre — humain ou machine — en héritant de sa fenêtre et de ses bornes. **Accepter** : consentir avec lucidité à ce qu'une part échappe. À *débattre* : quand on délègue sa lecture à une machine, de quoi hérite-t-on sans toujours le voir ?

Probité du lecteur. L'éthique minimale de qui sait ce qu'il lit, ce qu'il ne lit pas, et ce qu'il a délégué de sa lecture. (À ne pas confondre avec la *probité* comme variable de l'optimum — la fiabilité du prélèvement, au temps 1.) À *débattre* : qu'est-ce qu'une lecture honnête à l'ère du sur-débit, quand nul ne peut tout lire ?

Le référent. Quand le déchiffrement d'un lecteur bute sur la limite de son corpus, il se tourne vers un corpus plus riche : le non-connaisseur vers le connaisseur. Rapport asymétrique mais réciproque dans son principe — chacun est le connaisseur de quelqu'un et le non-connaisseur d'un autre — et qui n'est pas une hiérarchie de valeur entre personnes. À *débattre* : à qui se fie-t-on pour lire ce qu'on ne sait pas lire — et comment ce rapport peut-il aider sans déposséder ?

Notions transversales

Donnée observée / donnée vraisemblable. L'observée est prélevée du flux du réel ; la vraisemblable est produite par modélisation — elle ressemble à une observation sans en être une. La distinction commande tout le rapport aux contenus générés (statistiques, simulations, sorties d'IA). À *débattre* : à quoi reconnaît-on, et avec quelles conséquences, qu'une donnée a été observée plutôt que produite ?

Inappropriabilité du flux. On peut posséder un fichier, un support, une inscription — un canal. On ne possède pas le flux du réel, qui s'écoule pour tous à la fois : ce que l'un y prélève n'ôte rien à ce que l'autre peut y prélever. Cette **non-rivalité** est **ontologique** : elle ne préjuge d'aucun régime juridique — support, base, secret, personne, accès se règlent chacun en son ordre. Leçon de modestie (nul n'a le fleuve) et de liberté (nul ne peut prétendre le posséder à notre place). À *débattre* : la vraie

question politique est-elle « qui possède les données ? » ou « qui décide de ce qu'on prélève, et qui garde l'accès au flux ? »

Fond / forme. Le corpus dit ce qui est vrai indépendamment des techniques (le fond) et s'interdit de prescrire les implémentations (la forme). Affirmer qu'un optimum existe est de fond ; dire où il se situe ici est de forme. Les exemples datés illustrent, ils ne fondent pas. *À débattre* : comment tenir un propos durable sur la donnée quand les techniques, elles, se périment vite ?

Fait partageable entre disciplines. Le corpus n'apporte ni méthode ni outils, mais une grammaire conceptuelle commune à des disciplines qui produisent, échangent et lisent des données sans partager les termes pour en penser le statut. *À débattre* : à quelles conditions une donnée devient-elle un fait partageable d'un domaine à l'autre ?

IA et LLM (statut illustratif). Dans le corpus, l'IA n'est pas un concept mais une illustration. Un grand modèle de langage est une *sauvegarde* : appris d'un monde figé à une date, il diverge du présent — d'où les tentatives de le rebrancher en direct sur le flux. On peut le lire comme un saut systémique du stockage, et comme une position située de reprise ou de traitement de plus — lecteur effectif ou non, question que le corpus ne tranche pas. *À débattre* : l'IA déplace-t-elle le geste de la connaissance plutôt qu'elle ne le remplace — et que comble-t-elle en silence ?

L'artisan de la donnée — un métier vacant. Cas particulier du référent : le référent du domaine propre de la donnée, pour les très petites structures exclues de l'économie industrielle de la donnée. Le corpus en établit la nécessité et en appelle la création — il ne le décrit pas comme existant. Posture : aider, rendre autonome, ne pas déposséder. *À débattre* : à supposer ce métier nécessaire, qu'est-ce qui le distinguerait d'un prestataire qui se rend indispensable ?

Questions vives — points de controverse

Une dizaine d'entrées pour lancer et tenir une discussion. Chacune mobilise ses termes.

1. La donnée est-elle une chose qu'on collecte, ou le résultat d'un acte qui la constitue — prélèvement, institution, dérivation ? (*flux du réel, prélèvement, acte d'inscription, trois familles de constitution*)
2. Si une donnée ne vaut que pour un usage, qu'est-ce qui justifie encore d'en accumuler ? (*deux temps, fraîcheur, péremption, validité pour un usage, sources stables*)
3. « Toujours plus de données » : progrès, ou erreur de réglage ? (*optimum observationnel, sous-prélèvement, sur-ingénierie*)
4. Quand une amélioration cesse-t-elle d'améliorer pour transformer ? (*saut systémique*)

5. Propager la donnée, est-ce propager la connaissance ? (*propagation, part du lecteur*)
6. Le sur-débit est-il subi, ou produit par nos choix ? (*torrent, sur-débit, auto-amplification*)
7. Une donnée que personne ne lit existe-t-elle encore comme donnée — et que lui manque-t-il ? (*donnée constituée / actualisée, donnée muette, donnée morte, trace*)
8. Toutes les données non lues sont-elles des échecs ? (*les trois chemins*)
9. Déléguer sa lecture à une machine : de quoi hérite-t-on, et qui en répond ? (*lecteur situé, position située de reprise ou de traitement, lecteur effectif, fenêtre, trois postures*)
10. La question politique de la donnée : qui possède, ou qui prélève et accède ? (*inappropriabilité du flux*)
11. Une donnée observée et une donnée vraisemblable ont-elles le même statut ? (*observée / vraisemblable, IA*)

Pour le développement complet de chaque notion, voir le corpus À l'épreuve du flux — une ontologie processuelle de la donnée (dix-sept volumes). Ce canevas n'en est que la table des termes, ouverte pour la discussion.

David Bories — Conçu et dirigé par l'auteur.

contact@artisanat-de-la-donnee.fr

Canevas de discussion

—

L'ontologie processuelle de la donnée en ses termes

—