

LE FLEUVE ET LE CANAL

*La littératie de la donnée
à l'épreuve du flux*

David Bories

Le fleuve et le canal — La littératie de la donnée à l'épreuve du flux

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

Juillet 2026.
5.2.0-poche

contact@artisanat-de-la-donnee.fr

Table des matières

Introduction	1
<i>Pour une approche processuelle de la donnée</i>	1
Préface méthodologique	3
<i>Le fond, la forme, et le statut de cet essai</i>	3
Pourquoi cette préface	3
L'analogie de 1950	3
Le statut des illustrations	4
Le statut du « flux du réel »	5
La position du lecteur et la responsabilité de la forme	5
Le statut de cet essai et le rôle de l'IA	5
La règle de cet essai en une phrase	6
De l'entaille à l'algorithme	7
<i>Une histoire de la donnée dans l'humanité</i>	7
Le geste	7
De l'objet au signe	8
L'écriture au service des entrepôts	9
L'institution hérite du geste	9
La série révèle la loi	10
La machine prend le relais	11
Le déluge	11
La métacognition	12
Repères	13
Lire, écrire, comprendre le monde par la donnée	15
<i>La littératie de la donnée : ce qu'elle est, d'où elle vient, où elle va</i>	15
Ce document et son temps	15
Une longue généalogie	15
Ce que le champ a établi	16
Ce que le champ n'a pas réglé	17
Le point aveugle	18
Repères	18
Le flux et l'annotation	19

<i>Pour une ontologie processuelle de la donnée</i>	19
Préambule — Ce que ce document ajoute	19
Partie I — Les deux ontologies de la donnée	19
Partie II — Le flux comme primat ontologique	22
Partie III — La métaphore de la sauvegarde	23
Partie IV — L'acte d'annotation	24
Partie V — Le corpus comme condition d'intelligibilité	26
Partie VI — La péremption temporelle	27
Partie VII — Du flux d'observation : le maillon manquant	29
Partie VIII — Conséquences plus larges	30
Conclusion — Le filet et le fleuve	32
Repères	33
Les deux temps de la donnée	35
<i>L'acte, le support, et le courant qui fait vieillir ce qui ne bouge pas</i>	35
Préambule — Un paradoxe à résoudre	35
Partie I — Les deux opérations qui font une donnée	36
Partie II — Les deux temps	38
Partie III — Ce que la distinction éclaire	41
Partie IV — Fraîcheur, vraisemblance et prédiction	45
Conclusion — Le point fixe et le courant	48
Repères	49
La trace et le lecteur	51
<i>Ce qui rend la donnée possible — Aux fondations de l'essai</i>	51
Préambule — Une question jamais posée	51
Partie I — La lisibilité du réel	52
Partie II — La trace n'est pas la donnée	55
Partie III — Le lecteur que le flux a produit	57
Partie IV — Ce que cette fondation légitime	59
Conclusion — La vague et l'océan	61
Repères	62
Creuser le canal	63
<i>Le flux d'observation : comment l'humain reconstruit le réel par la donnée</i>	63
Préambule — Le maillon manquant	63
Partie I — Qu'est-ce qu'un flux d'observation ?	64
Partie II — La mise en mouvement	67
Partie III — L'optimum observationnel	69
Partie IV — La manipulation du flux d'observation	71
Partie V — La dynamique : flux qui engendrent des flux	72
Partie VI — La boucle : quand l'observation transforme le réel	75
Partie VII — Le saut systémique	76
Partie VIII — Ce que le flux d'observation apporte à la littérature	79
Conclusion — L'art du canal	82
Repères	83

La vague et l'océan	85
<i>L'humain dans le flux — Conclusion de l'essai</i>	85
Préambule — Refermer l'arche	85
Partie I — La vague	86
Partie II — Le torrent	88
Partie III — La lucidité	89
Conclusion générale de l'essai — Ce que sait la vague	93
Repères conceptuels	94
Bibliographie raisonnée	97
<i>Appareil de références de l'essai</i>	97
Ce qui ne changera pas	97
Ce qui évolue et comment le suivre	101
Pour approfondir par soi-même : termes de recherche durables	102
Sources archéologiques et historiques citées dans l'essai	103
Note sur la « fraîcheur » et le « data decay »	104
Note sur le critère de Nyquist-Shannon	104
Note sur Van Dijck et la critique du <i>dataïsme</i>	105

Introduction

Pour une approche processuelle de la donnée

Nous ne vivons pas entourés de données.

Nous vivons entourés de réel qui s'écoule — et nous croyons voir des données là où il n'y a, tant que nous n'avons rien inscrit, que le monde en train de passer. C'est de ce déplacement de regard que part le présent essai. Il défend une idée simple et, peut-être, dérangeante : la donnée n'est pas un objet que l'on cueille dans un champ déjà constitué. Elle est un prélèvement opéré sur le flux continu et inépuisable du réel — un instantané qui, dès qu'il est pris, commence à s'éloigner du monde qu'il prétend décrire.

De ce déplacement découle tout le reste : une donnée se périme ; sa valeur dépend du corpus qui l'entoure et de la fidélité de l'acte qui l'a produite ; l'humain, loin de simplement « avoir » des données, creuse des canaux dans le fleuve du réel pour en capter ce dont il a besoin.

Les sept chapitres qui composent ce recueil déroulent cette idée, depuis les premières entailles préhistoriques jusqu'au torrent numérique contemporain, et tentent de montrer en quoi elle complète — et déplace — la manière dont la littérature de la donnée est aujourd'hui pensée et enseignée. Le lecteur remarquera que l'essai traverse deux fois certains paysages — ainsi le flux d'observation, esquissé au chapitre de l'annotation, puis traversé en entier dans *Creuser le canal* — : c'est la géologie assumée d'un texte qui a grandi par strates, et la seconde traversée suppose la première.

Conçu et dirigé par **David Bories**.

Rédigé en utilisant l'intelligence artificielle générative.

Mai 2026

Préface méthodologique

Le fond, la forme, et le statut de cet essai

« Toutes les opinions ne se valent pas, mais toutes méritent d'être pesées. »

— adapté d'un adage attribué à Voltaire

Pourquoi cette préface

Cet essai déploie, en sept documents, une réflexion sur la donnée. Il n'est ni un manuel pratique, ni un manifeste politique, ni un traité académique. Il est un essai de conceptualisation, qui cherche ce qui demeure vrai de la donnée indépendamment des techniques qui la produisent à un moment donné, et qui s'interdit pour cela de prescrire la forme concrète que prennent les choses dans le temps présent du lecteur.

Cette posture mérite d'être exposée d'emblée, car elle commande tout ce qui suit : ce que l'essai dit, ce qu'il s'interdit, et ce qu'il laisse à d'autres. Une fois cette préface lue, chaque document peut être abordé sans rappel ; la règle vaut, partout, sauf indication contraire.

L'analogie de 1950

Imaginons un instant que nous sommes en 1950. L'automobile se démocratise. Il devient évident qu'elle va transformer l'organisation de la société. Pour penser cette transformation, on peut établir des faits qui relèvent du **fond** — vrais indépendamment de l'évolution future du produit : une automobile roule, donc il faudra un revêtement au sol ; une automobile consomme, donc il faudra des points de ravitaillement répartis sur le territoire ; une automobile se conduit, donc il faudra apprendre à la

conduire. Ces énoncés ne seront pas démentis par l'histoire, quelle que soit l'évolution technique du véhicule.

Mais on ne peut pas, en 1950, fixer la **forme** : la largeur exacte des voies, les limitations de vitesse, le nombre de carburants disponibles, l'électrification future, le détail du code de la route. Ces éléments dépendent d'une évolution technique et sociale imprévisible. Les figer en 1950 serait une erreur.

Cet essai applique strictement cette distinction au concept de donnée et de flux d'observation. **Il énonce le fond** : ce qui est vrai de la donnée indépendamment des technologies qui la produisent. **Il s'interdit de prescrire la forme** : les architectures concrètes, les protocoles, les seuils chiffrés, les institutions, les solutions actionnables. Le lecteur qui chercherait dans cet essai des recettes opérationnelles ne les y trouvera pas — non parce qu'elles seraient illégitimes, mais parce qu'elles relèvent d'un autre travail.

Cette retenue n'est pas un agnosticisme. Elle est la condition pour produire un texte qui ne se périmé pas en même temps que la génération technique qui l'a vu naître.

Le statut des illustrations

L'essai mobilise un certain nombre d'exemples concrets : la météorologie, le cinéma, la photographie et la compression d'image, la reforestation d'un sol aride, le relevé de température, la gestion d'un rayon de commerce, le calcul d'une trajectoire, les cernes de l'arbre, les strates géologiques, l'empreinte dans la boue, la vague et l'océan, le fleuve et le canal, le torrent.

Ces illustrations relèvent de la **forme**. Elles sont datées, révisables, et ne constituent pas des affirmations empiriques sourcées. Elles servent à rendre intuitif un concept abstrait — pas à le démontrer. Le lecteur est invité à les lire comme on lirait, dans un texte de 1950, « une automobile peut atteindre 100 km/h » : vrai à la date d'écriture, accessoire au raisonnement, destiné à être dépassé. Seul le concept qu'elles servent engage l'essai.

Le même statut vaut pour les images directrices — le fleuve, le canal, la vague, l'océan, le torrent. Ces images sont destinées à rendre le concept mémorable. Elles relèvent de la forme du propos ; seul le fond qu'elles servent engage l'essai. Une image qui cesse d'éclairer, ou qu'un lecteur d'une autre culture saurait formuler autrement, peut être remplacée sans perte.

Le statut du « flux du réel »

L'essai mobilise constamment la notion de « flux du réel ». Cette notion appelle une précaution.

L'essai n'engage pas une thèse métaphysique sur la nature ultime du réel. Il ne tranche pas le débat entre un réalisme processuel (le réel *est* fondamentalement écoulement) et une position épistémique plus modeste (notre rapport au réel prend la forme d'un écoulement pour des êtres situés comme nous). Pour la portée du propos, ces deux positions reviennent au même : que cette caractéristique appartienne au réel lui-même ou à notre condition de le percevoir, l'essai tient dans les deux cas.

De la même manière, quand l'essai emploie des formules qui semblent prêter une activité aux données — « la donnée se périme », « l'annotation entre dans la boucle du réel », « la trace dit » — ce sont des raccourcis d'écriture. Ce qui agit, dans tous les cas, ce sont des humains, des institutions, des dispositifs techniques — c'est-à-dire des systèmes sociotechniques. La donnée elle-même n'agit pas ; elle est ce sur quoi on agit, ou ce par quoi on agit. Le raccourci est commode ; il ne doit pas être pris à la lettre.

La position du lecteur et la responsabilité de la forme

L'essai vise un lecteur singulier : celui qui prendra le concept entre ses mains pour en faire quelque chose dans son propre contexte. Cette responsabilité de la forme — déterminer, dans tel cas concret, où se situe l'optimum observationnel, comment cadrer un flux d'observation, à quel rythme rafraîchir un prélèvement — n'est pas un défaut du concept. C'est le concept qui reconnaît la responsabilité humaine pour ce qu'elle est : irréductible à une règle générale, toujours à reprendre dans la situation.

Le concept est une lampe : il éclaire la zone où il faut regarder, il avertit qu'il y a là des écueils. Il ne dessine pas la carte détaillée des récifs. Cette carte, c'est à chaque humain, dans chaque contexte, de la dresser.

Un mot enfin sur la destination. Cette traversée n'est pas close sur elle-même : la trilogie débouche, à son troisième tome, sur une éthique — celle du lecteur situé, de ses postures et de sa probité. Le présent tome n'en dit rien encore ; il en pose le sol.

Le statut de cet essai et le rôle de l'IA

Cet essai a été conçu et dirigé par David Bories, et rédigé en utilisant l'intelligence artificielle générative. Cette transparence n'est pas un effet

de mode : elle est une cohérence avec la thèse même de l'essai.

Selon la théorie qu'il développe (et notamment celle des trois âges de la donnée et de la distinction entre fraîcheur et vraisemblance), un grand modèle de langage n'est pas un lecteur direct du flux du réel : il est entraîné sur un corpus figé, et produit des vraisemblances calculées à partir de ce corpus. L'essai prend acte de cette nature : l'IA est ici un instrument puissant — un canal sophistiqué — mis en mouvement par une intention humaine. C'est l'humain qui a vu l'urgence du sujet, choisi l'angle, arbitré les phrases, retenu et rejeté. L'IA n'a pas créé la pensée ; elle a aidé à la formuler, à la tester, à la déplier. Que cette collaboration ait été féconde, en dépit du caractère figé de l'instrument, fait partie des questions que l'essai laisse ouvertes — et qu'un travail séparé, distinct de l'essai lui-même, pourra examiner plus frontalement.

La règle de cet essai en une phrase

Cet essai dit ce qu'il est vrai de dire de la donnée pour longtemps. Il laisse à d'autres, et au lecteur, le soin de déterminer ce qu'il faut en faire pour aujourd'hui.

Préface rédigée en mai 2026.

De l'entaille à l'algorithme

Une histoire de la donnée dans l'humanité

« *Ce qui n'est pas mesuré ne peut être gouverné.* »

— Adage administratif

Le geste

À un moment de son histoire, l'homme a accompli quelque chose d'inédit parmi les êtres vivants qui l'entouraient : il a choisi de ne pas laisser le monde passer sans laisser de trace. Non pas une trace involontaire — l'empreinte dans la boue, l'écorce griffée par le passage — mais une trace *voulue*. Un signe posé délibérément pour représenter quelque chose d'autre que lui-même, pour forcer le souvenir, pour que demain sache ce qu'aujourd'hui a vu.

Quand cela s'est-il produit pour la première fois ? La question est peut-être mal posée. Les traces que nous avons retrouvées ne sont pas les premières traces qui ont existé : elles sont les premières qui ont traversé le temps. La roche dure conserve ce que le bois, l'écorce, la peau ne conservent pas. Ce que nous connaissons de l'homme préhistorique est une sélection involontaire opérée par les matières et les millénaires, non un inventaire fidèle.

Parmi les gestes qui ont survécu, certains résistent à toute interprétation assurée. Des parois de grottes portent des représentations d'animaux, des mains en négatif, des formes géométriques répétées. Étaient-ce des annotations du réel — un relevé de chasse, un décompte de saisons, une cartographie du territoire — ou l'expression d'une intériorité que nous ne saurons jamais déchiffrer ? La distinction entre *annoter* et *exprimer* paraît nette depuis notre présent ; elle l'était peut-être moins pour celui qui peignait à la lueur d'une torche. Ce que ces représentations nous disent avec certitude, c'est que la capacité à fixer le monde sous forme de

signe est très ancienne, bien plus ancienne que l'écriture, peut-être aussi vieille que l'humanité consciente elle-même.

D'autres gestes sont plus lisibles. Des os portant des séries d'entailles régulières ont traversé des dizaines de millénaires dans des sédiments protecteurs. Le nombre d'encoches, leur organisation, leur régularité suggèrent un comptage : des jours, des lunes, des animaux, des membres d'un groupe. Ce qui importe ici n'est pas la raison exacte — elle reste discutée — mais l'acte lui-même : observer un phénomène, décider qu'il vaut la peine d'être retenu, lui associer un signe matériel, maintenir la correspondance dans la durée. Ce geste est le prototype de tout ce qui suivra. Il n'a rien de spectaculaire. Il est, en silence, fondateur.

La dissociation venait d'avoir lieu. Pour la première fois, un fragment du réel existait séparé du réel. L'événement pouvait continuer de couler ; sa trace, elle, restait. Depuis ce moment, l'homme allait produire, avec une ingéniosité et une constance extraordinaires, des supports toujours plus efficaces pour fixer ce que le temps autrement emporterait.

Ces traces, une fois fixées dans la matière, ne restent pas face au réel — elles en deviennent partie. La peinture sur la paroi a perdu depuis longtemps ce qu'elle signifiait pour celui qui la posa : son inscription, la fraîcheur de son sens, est périmée à tout jamais. Mais sa fixation, elle, demeure. Le réel l'a absorbée. Elle est devenue l'une des traces que le réel garde de lui-même.

De l'objet au signe

Compter sur ses doigts a une limite évidente : on ne peut pas les garder levés indéfiniment. Compter avec des cailloux en a une autre : passé un certain nombre, manipuler des cailloux devient aussi ingérable que ce que l'on cherchait à comptabiliser. L'homme a résolu ce problème d'une manière qui ressemble à une évidence rétrospective, mais qui représente un saut cognitif considérable : il a fabriqué lui-même les objets qui lui serviraient à compter.

Non plus des cailloux ramassés au hasard, mais des formes façonnées à la main, standardisées, dont chacune représentait une catégorie précise de ce que l'on dénombrait. Un objet pour une mesure de grain. Un autre pour un mouton. Un troisième pour une jarre d'huile. Puis, passé un nouveau seuil, un objet en représente plusieurs — la quantité se condense, le nombre d'objets diminue sans que la réalité qu'ils représentent soit réduite. L'abstraction est en marche.

Le pas suivant est plus discret, mais peut-être plus décisif encore. Ces objets, on les enferme dans des enveloppes d'argile pour les protéger, pour les transporter, pour certifier une transaction. Puis on grave à la surface de

l'enveloppe l'empreinte de ce qu'elle contient — pour éviter de la briser à chaque vérification. Et bientôt, l'empreinte suffit. L'objet à l'intérieur n'est plus nécessaire. Le signe extérieur dit tout ce que l'objet disait. C'est le moment où la représentation se détache de la chose représentée. Le signe n'imité plus l'objet — il le *désigne*. Ce déplacement discret est l'acte de naissance de l'écriture.

L'écriture au service des entrepôts

On imagine volontiers l'écriture comme une invention de poètes ou de philosophes — un outil né du désir de transmettre une pensée, de mettre en récit une épopée. Les traces les plus anciennes que nous ayons déchiffrées racontent autre chose.

Les premières tablettes portent des listes. Des rations distribuées, des têtes de bétail, des quantités de grain reçues et expédiées. L'écriture, dans les formes que nous pouvons lire, est née dans les entrepôts, pas dans les temples. Elle a été conçue pour gérer des données avant d'être utilisée pour transmettre des idées. Les scribes qui l'ont développée étaient des agents administratifs, non des lettrés au sens que nous donnerions au mot.

Cela ne signifie pas que l'expression symbolique n'existait pas avant — les parois peintes sont là pour le rappeler, et elles précèdent de beaucoup les premières tablettes. Cela signifie que le premier système assez rigoureux, assez partagé, assez stable pour traverser les siècles et demeurer lisible était un système de gestion. La pensée abstraite, la narration, la poésie ont ensuite trouvé dans ce système un véhicule qu'elles n'avaient pas inventé.

Ce renversement mérite d'être retenu : la donnée n'est pas un usage dérivé de l'écriture. Dans les traces que nous avons retrouvées, c'est l'écriture qui est une réponse à un besoin de données. Elle a servi le réel avant de servir l'imagination.

L'institution hérite du geste

Le geste individuel — noter pour soi, compter pour soi — n'est resté individuel qu'un temps. À mesure que les sociétés se complexifiaient, des structures ont pris en charge l'annotation du monde à une échelle que nul individu ne pouvait atteindre seul.

L'administration recense les hommes, les terres, les troupeaux — pour taxer, pour mobiliser, pour planifier. Les institutions religieuses tiennent les registres des naissances, des unions, des morts, assurant la mémoire des individus là où l'État n'avait pas encore les moyens de le faire. Le marchand invente des écritures comptables capables de suivre les flux d'un commerce qui traverse des continents. Chacun, à sa manière,

organise le même geste fondateur : observer le réel, en fixer une représentation, conserver cette représentation pour pouvoir s'y référer.

Dans ce mouvement, les bibliothèques occupent une place singulière. Elles ne sont pas seulement un lieu de conservation : elles sont la première infrastructure organisée de stockage *et* de mise à disposition de données accumulées par l'humanité. D'abord réservées à une élite restreinte — les foyers du pouvoir, du culte et du savoir des civilisations qui en eurent les moyens —, elles se sont progressivement ouvertes, jusqu'à devenir, avec la bibliothèque publique moderne, un équipement accessible à tous. En cela, elles incarnent l'un des mouvements lents mais continus de l'histoire de la donnée : ce qui était privilège devient, avec le temps, commun.

Aujourd'hui, l'ensemble du savoir accumulé dans les bibliothèques du monde — constitué sur des millénaires, entreposé sur des rayonnages qui s'étendraient sur des centaines de kilomètres — peut être contenu dans une infrastructure numérique dont l'emprise au sol ne représente qu'une fraction des bâtiments physiques. Ce n'est pas seulement un changement de volume : c'est un changement de la nature de l'accès. Les bibliothèques n'ont pourtant pas disparu. Elles perdurent, parfois réinventées, parfois en difficulté — comme une certaine presse écrite qui résiste. Non par inertie, mais parce qu'elles répondent à quelque chose que le datacenter ne remplace pas entièrement : la médiation humaine, le rapport physique au document, la curation du savoir par des personnes qui en ont fait leur métier. Le numérique n'a pas effacé le papier. Il l'a déplacé.

La donnée a cessé d'être un aide-mémoire individuel. Elle est devenue un instrument de gouvernement, d'instruction et de commerce collectifs. Celui qui organise les annotations — et l'accès à ces annotations — contrôle une part de la réalité commune.

La série révèle la loi

Une observation isolée informe peu. Elle dit ce qui s'est passé une fois, à un endroit, dans des conditions particulières. Répétée, comparée, organisée dans le temps, la même observation commence à dire autre chose : ce qui se passe *régulièrement*, ce qui suit un rythme, ce qui obéit à une loi.

L'accumulation patiente d'observations comparables — les positions des astres nuit après nuit sur des décennies, les naissances et les morts dans une ville sur plusieurs générations, les prix d'un marché sur des années, les éruptions volcaniques et les marées sur des côtes habitées, les variations du climat et des saisons sur des terres cultivées — finit par révéler des régularités que nulle observation ponctuelle ne pouvait soupçonner. Et une régularité, une fois identifiée, permet de faire quelque chose que la simple annotation ne permettait pas : *prévoir*.

Prévoir ne signifie pas tout voir. L'homme qui observe le réel n'en perçoit

toujours qu'une portion — limitée par ses instruments, par ses angles d'observation, par ce qu'il a choisi ou su mesurer. Sa prédiction est partielle, incertaine, révisable. Mais elle suffit à produire une décision : agir sur une portion du réel, modifier ce qui sans intervention suivrait son cours. Cette action sur le réel le transforme. Et la transformation produit de nouvelles observations, qui appellent de nouvelles annotations, qui alimentent de nouvelles prédictions. Un cycle se referme — et un autre s'ouvre.

La donnée a changé de temporalité. Elle regardait en arrière pour ne pas oublier ; elle regarde maintenant en avant pour anticiper. Et l'homme qui prédit pour agir ne se contente plus d'observer le monde : il commence, par ses décisions, à le modifier en retour. L'annotation entre dans la boucle du réel.

La machine prend le relais

Compter, classer, trier, comparer : ces opérations ont en commun d'être répétitives. Elles n'exigent pas de jugement — elles exigent de la méthode et de la constance. Toute l'histoire des techniques porte la trace d'un mouvement constant : confier à l'outil ce qui est systématique pour libérer l'attention vers ce qui demande une décision.

Les premiers outils mécaniques d'aide au calcul sont apparus bien avant l'ère industrielle. Puis la mécanisation a pris une ampleur nouvelle : des machines capables non seulement de calculer, mais de stocker des instructions, d'exécuter des opérations en séquence, de conserver des résultats. Le passage de la machine à calculer à la machine à traiter l'information est un saut qualitatif aussi significatif que le passage du jeton d'argile à l'écriture.

Ce saut a obéi à la même logique : confier à l'outil ce qui est systématique, se réserver ce qui ne l'est pas. Les machines ont d'abord pris en charge le calcul brut, puis la gestion de données en volumes croissants, puis — progressivement — des formes de reconnaissance, de classement et de décision que l'on croyait réservées à l'intelligence humaine. À chaque étape, la frontière entre ce que la machine fait et ce que l'homme conserve s'est déplacée.

Le déluge

Il arrive un point où les chiffres ne parlent plus. Dire que l'humanité produit chaque jour une quantité de données numériques représentant des millions de fois le contenu d'une grande bibliothèque ne dit rien que

l'on puisse vraiment saisir. Ce qui dit quelque chose, c'est ce que cela exige pour exister.

Stocker, traiter et distribuer ce flux continu d'annotations nécessite une infrastructure physique que l'on peut mesurer en termes concrets : des bâtiments dont la surface cumulée à l'échelle mondiale représente l'équivalent de plusieurs grandes agglomérations ; une consommation d'électricité comparable à celle de pays entiers ; des besoins en eau pour le refroidissement des équipements qui se comptent comme la consommation annuelle de métropoles. Et cette infrastructure continue de croître — en quelques années récentes, l'humanité a produit autant de données numériques qu'elle n'en avait accumulé depuis les décennies entières précédentes.

Ce volume n'est pas homogène. Une partie de ces données a vocation à être conservée, archivée, analysée à terme. Une autre est éphémère par nature : les capteurs embarqués dans les machines, les véhicules, les réseaux industriels — et plus largement les usages numériques comme la visioconférence ou la téléphonie — produisent en permanence des observations qui n'ont de valeur qu'à l'instant où elles sont produites — pour une décision qui doit être prise dans la seconde — et qui seront effacées l'instant d'après. La question de ce qu'il faut garder et de ce qu'il faut laisser passer est devenue une décision à part entière, une nouvelle forme du geste fondateur.

L'homme n'a pas cessé d'annoter le monde. Mais il annote désormais largement par l'intermédiaire de systèmes numériques d'une puissance et d'une rapidité sans précédent. Cette médiation lui donne une capacité inédite : observer, dans leur ensemble et en temps quasi réel, les annotations qu'il produit sur le monde. Il n'observe plus seulement le monde directement — il observe aussi l'image que ses propres annotations en dessinent. Et ces deux regards ne se confondent pas toujours.

La métacognition

L'homme a longtemps observé le monde réel qui l'entourait pour l'annoter. Il observe maintenant, de plus en plus, la réplique du monde que ses annotations ont construite.

Cette réplique n'est pas le monde. Elle en est une projection partielle, filtrée par les critères selon lesquels les annotations ont été produites — les phénomènes choisis pour être observés, les instruments utilisés, les seuils de mesure appliqués, les décisions prises sur ce qui mérite d'être capté. Elle dit ce que l'on a décidé de regarder, dans les formes où l'on a décidé de le regarder. Ce qu'elle ne dit pas est au moins aussi vaste que ce qu'elle dit.

Mais cette réplique a une propriété que le monde réel n'a pas : elle est

navigable, comparable à elle-même dans le temps, interrogeable à volonté. On peut y chercher des régularités, y tester des hypothèses, y prendre des décisions. Et à la vitesse à laquelle les systèmes numériques la mettent à jour, elle donne l'impression d'une synchronicité avec le monde réel — impression qui est en grande partie une illusion. Le réel continue de s'écouler, et la réplique court après lui.

Pour maintenir cette réplique aussi proche que possible du réel, l'homme a délégué à la machine l'observation directe du monde — les capteurs, les algorithmes de traitement, les systèmes d'annotation automatique. Il se réserve, de plus en plus, l'observation de cette observation : décider ce que les chiffres signifient, ce que les tendances impliquent, ce qui doit être fait. L'acte n'est plus simplement cognitif — il est devenu méta-cognitif. L'homme n'observe plus seulement le monde : il observe ce que son observation du monde produit.

C'est dans cet espace — entre le flux du réel et la réplique qu'il en construit — que se joue aujourd'hui l'essentiel de ce que l'on appelle la maîtrise de la donnée. Non pas posséder des données, mais comprendre ce que l'on a choisi d'observer, pourquoi on a choisi de l'observer ainsi, et ce que cette observation ne montre pas. C'est l'objet de l'essai qui suit.

Repères

Pour l'ensemble des références mobilisées dans ce document — datations des artefacts, sources archéologiques, ouvrages historiques —, voir la *Bibliographie raisonnée* en fin d'essai, notamment les sections « Sources archéologiques et historiques citées dans l'essai » et « Histoire et anthropologie de la donnée ».

Conformément à la préface méthodologique, les exemples mobilisés dans ce document relèvent de la *forme* ; ils sont datés et révisables. Seul le mouvement de fond qu'ils éclairent — la donnée comme geste humain antique, sa fonction administrative première, son institutionnalisation, son industrialisation, sa numérisation — engage l'essai.

Lire, écrire, comprendre le monde par la donnée

La littératie de la donnée : ce qu'elle est, d'où elle vient, où elle va

« *La littératie des données est moins une question de compétence technique que de participation démocratique.*
»

— Catherine D'Ignazio

Ce document et son temps

Ce chapitre situe l'essai dans le champ qui l'accueille. Il en donne l'essentiel — la généalogie, les acquis, les tensions, et le point aveugle d'où partent les documents suivants. L'état détaillé du champ — cadres institutionnels, écosystème d'acteurs, chiffres, critiques — vit désormais dans une note complémentaire, *La littératie de la donnée — état du champ*, qui l'assume comme un repère daté, révisable à son rythme. Ici ne demeure que ce qui ne se périme pas. Le lecteur pressé peut encore traverser ce chapitre d'un trait : la traversée ontologique commence au suivant.

Une longue généalogie

Le mot *littératie*, calqué sur l'anglais *literacy*, dérive du latin *littera* — la lettre. Dans son sens le plus strict, il désigne la capacité à lire et à écrire ; pendant la plus grande partie de l'histoire humaine, cette capacité fut le privilège d'une élite, qui en tirait — Jack Goody l'a montré — une part essentielle de son pouvoir, avant de devenir, en un siècle, un bien commun : moins de six adultes sur dix savaient lire au milieu du XX^e siècle, plus

de huit sur dix au début du XXI^e. Ce mouvement n'est pas sans rappeler ce que l'essai examine ailleurs : ce qui était l'apanage de quelques-uns devient, avec le temps, bien commun.

Trois transformations ont, depuis, renouvelé le concept. La *littératie fonctionnelle* d'abord (Gray, 1956 ; UNESCO, 1978) : non plus une technique mais une compétence sociale — *contexte, autonomie, participation* —, marquée par Paulo Freire, pour qui apprendre à lire le monde précède apprendre à lire les mots. *L'éclatement* ensuite : à mesure que les sociétés se complexifient, une constellation de littératies spécialisées apparaît — médiatique, informationnelle (1974), statistique, numérique (1997). La *consolidation* enfin (UNESCO, 2003) : identifier, comprendre, interpréter, créer, communiquer et calculer, dans des contextes variés, en vue de participer pleinement à la vie sociale.

C'est de ce sol qu'émerge, au début des années 2000, la **littératie de la donnée** — avec un précurseur que la littérature du champ a effacé de sa propre généalogie : la *datalogie* de Peter Naur (1966), science de la nature et de l'usage des données, qui en portait déjà tout le programme — dissiper le mythe technique, former à la compréhension critique — trente ans avant le mot.

Ce que le champ a établi

L'acte de naissance le plus cité est l'article de Milo Schield (2004), qui fixe la tripartition fondatrice : la littératie informationnelle porte sur les *sources* ; la statistique sur le *raisonnement* ; la littératie de la donnée, entre les deux, sur la *matière brute* elle-même — sa sélection, sa transformation, sa présentation. La valeur d'une statistique, écrit Schield, dépend lourdement de la façon dont les données sous-jacentes ont été sélectionnées, converties et manipulées.

Le concept s'enracine ensuite dans la pédagogie — Mandinach et Gummer le définissent comme la capacité à transformer de l'information en connaissance actionnable ; Ridsdale et son équipe en donnent le cadre académique le plus repris (collecte, gestion, évaluation, application) — et se double d'un courant critique, au MIT, autour de Catherine D'Ignazio et Rahul Bhargava : une littératie *capacitante*, héritière de Freire, résumée en quatre verbes — *lire les données, travailler avec, analyser, argumenter avec* — et couronnée par une définition civique : le désir et la capacité de s'engager de manière constructive dans la société, *à travers et au sujet* des données. La formule compte : la donnée n'y est pas seulement un outil, elle est un objet social sur lequel réfléchir.

Sous ces cadres, deux fondements théoriques. La pyramide DIKW d'abord (Ackoff) : la donnée, signe brut ; l'information, donnée mise en contexte ; la connaissance, information intégrée ; la sagesse, son application à bon escient. L'outil, malgré ses critiques, garde sa vertu pédagogique : ne pas confondre les niveaux, savoir se déplacer entre eux. Allons plus

loin encore : la donnée elle-même n'est pas un fait *donné* mais un fait *capté*. Le philosophe des sciences Kaplan le notait dès le milieu du XX^e siècle : ce n'est pas la nature qui fournit la donnée au chercheur, c'est le chercheur qui la capte dans la nature selon ses buts et ses cadres. Parler de données brutes est un oxymore ; toute donnée embarque les choix de ceux qui l'ont produite — les phénomènes retenus, les instruments, les seuils. Comprendre cela est le premier acte de la littératie — et c'est le point d'où partira tout le travail des documents suivants.

Trois traditions, enfin, se croisent dans le champ sans qu'aucune l'épuise : la *statistique* (le raisonnement probabiliste, la lecture critique des chiffres), la *documentaire* (les sources, les métadonnées, la chaîne de l'information), la *critique* (les rapports de pouvoir et les biais qui se nichent dans les données). Les cadres les plus aboutis articulent les trois.

Ce que le champ n'a pas réglé

La data literacy est un *objet frontière* : un responsable de données, un professeur documentaliste et un militant associatif n'entendent pas la même chose sous le même mot — et parlent pourtant du même objet. D'où le pluriel que la recherche tend à valider : des *data literacies*, qui coexistent sans converger. D'où, aussi, trois finalités en tension — professionnelle (la littératie comme capital productif), éducative (comme compétence formative), civique (comme condition de la démocratie) — entre lesquelles toute politique de formation arbitre, qu'elle le dise ou non. Et d'où les critiques qui traversent le champ : la récupération managériale qui dépolitise l'enjeu, la technique privilégiée sur l'éthique, les asymétries de pouvoir invisibilisées, le risque d'une littératie à deux vitesses, l'inflation d'un mot qui, à désigner tout, ne signifierait plus rien.

Le champ, enfin, se reconfigure à chaque transformation de la donnée elle-même — et la vague la plus récente, celle des modèles génératifs capables de lire, d'écrire et de raisonner avec les données, oblige à distinguer la littératie de la donnée, qui concerne les *matériaux*, de la littératie de l'IA, qui concerne les *systèmes* appris sur ces matériaux. À la question qui en découle — à quoi bon former des humains à ce que les machines font mieux ? — deux propositions résistent au temps : utiliser intelligemment ces machines exige précisément une littératie de la donnée, sans laquelle l'utilisateur est prisonnier de l'outil ; et les décisions collectives sur l'usage des données restent des décisions humaines — une société qui aurait délégué entièrement sa compréhension des données ne pourrait plus, à proprement parler, se gouverner.

L'inventaire détaillé — cadres institutionnels et leurs versions, acteurs et leurs logiques, chiffres et leurs sources, critiques et leurs auteurs — est tenu dans la note complémentaire *La littératie de la donnée — état du champ*.

Le point aveugle

À l'époque de l'imprimerie, savoir lire a cessé d'être un privilège pour devenir une condition de la participation civique. Le parallèle avec la donnée a ses limites, mais il indique une direction : la donnée fait de chacun de nous un objet enregistré ; la littératie est ce qui peut en faire un sujet qui lit et qui questionne — le lecteur critique de son propre monde, plutôt que son figurant statistique.

Les cadres examinés ici — Schield, Mandinach-Gummer, Ridsdale, et les référentiels institutionnels qui en descendent — ont un mérite considérable : ils ont formalisé des compétences qui n'existaient pas comme telles. Mais ils partagent une limite. **Aucun ne thématise la nature processuelle de la donnée.** Tous la traitent comme un objet à manipuler — à collecter, gérer, évaluer, appliquer — sans interroger ce que c'est, pour une donnée, que d'exister. Les documents qui suivent examinent ce que ce point aveugle change, et quelles compétences nouvelles il appelle.

Repères

L'état détaillé du champ — généalogie complète, cadres, acteurs, chiffres, critiques — est dans la note complémentaire *La littératie de la donnée — état du champ* (repère daté, révisé à son rythme). Pour les références — Schield, Mandinach et Gummer, Ridsdale, D'Ignazio, Naur, Ackoff, ainsi que les cadres institutionnels —, voir la *Bibliographie raisonnée* en fin d'essai, notamment les sections « Littératie de la donnée — textes fondateurs » et « Ce qui évolue et comment le suivre ».

Le flux et l’annotation

Pour une ontologie processuelle de la donnée

« Tout s’écoule. »

— Héraclite, fragment 12 (vers 500 av. J.-C.)

« *Un datum est ce qui est donné — pour une occasion ultérieure.*
»

— Alfred North Whitehead, *Process and Reality* (1929)

Préambule — Ce que ce document ajoute

Les deux documents qui précèdent ont posé deux bases. Le premier a retracé la trajectoire longue de la donnée comme geste humain. Le second a cartographié la littérature de la donnée et son point aveugle : aucun des cadres existants ne thématise la nature *processuelle* de la donnée ; tous la traitent comme un objet à manipuler.

Ce document construit le cadre philosophique que la donnée appelle : une ontologie de la donnée comme *processus*, qui en tire les conséquences temporelles, épistémiques et politiques. Ce cadre conduira ensuite à une notion que le document *Creuser le canal* développera : le flux d’observation, maillon que la théorie processuelle rend nécessaire et que l’usage pratique réclame.

Partie I — Les deux ontologies de la donnée

L’ontologie objectale : quatre présupposés tacites

L’ontologie dominante de la donnée repose sur quatre présupposés rarement examinés, mais omniprésents dans le langage courant, dans le

droit et dans la pratique technique.

Premier présupposé : la donnée préexiste à sa collecte. Quand on dit “collecter des données”, on suppose qu’elles existent déjà, quelque part, et qu’on les déplace vers un nouvel emplacement. Le verbe *collecter* implique des objets qui attendent qu’on les ramasse. Cette image est trompeuse : il n’y a pas, dans le réel, de données qui attendent. Il y a des phénomènes qui ne deviennent données qu’en étant annotés.

Deuxième présupposé : la donnée est stable. Une fois inscrite, elle est censée demeurer identique à elle-même. La permanence du support est étendue au contenu. Cette stabilité est plus apparente que réelle : la donnée ne change pas, mais le réel dont elle est extraite continue de changer. Très vite, elle représente un état du réel qui n’existe plus.

Troisième présupposé : la donnée a une valeur intrinsèque. D’où la fortune de la métaphore du “nouveau pétrole”, d’où les *data lakes*, d’où l’idée qu’il vaut la peine de stocker massivement parce qu’on saura plus tard quoi en faire. Cette croyance masque que la valeur d’une donnée dépend toujours d’un contexte d’usage et d’une fraîcheur qu’aucun stockage ne garantit.

Quatrième présupposé : la donnée peut être possédée. “Mes données”, “les données de l’entreprise”, “les données souveraines” : autant d’expressions qui supposent qu’une donnée est un bien, susceptible d’appartenir, donc d’être vendu, acheté, dérobé. C’est sur cette présomption qu’est bâti le droit des données personnelles, mais aussi tout le marché de la *data*.

Ces quatre présupposés ne sont pas faux. Ils sont *vrais dans un cadre étroit* : celui où la temporalité du réel a été suspendue ou ignorée. Dès qu’on réintroduit le temps — la durée bergsonienne, le devenir héraclitéen, le processus whiteheadien — l’ontologie objectale révèle ses limites.

Cette position a ses lettres de noblesse : la philosophie de l’information contemporaine — Floridi en tête — définit la donnée comme une *différence*, un défaut d’uniformité dans le tissu du réel. C’est la version la plus forte de l’objectal : la donnée n’y est plus une chose massive, mais elle demeure un donné — quelque chose que le monde contiendrait avant tout geste. Le présent essai fait un pas de plus : la différence est bien dans le flux ; mais tant que nul ne l’a prélevée et inscrite, elle n’est pas une donnée — elle est du réel qui diffère.

L’ontologie processuelle : quatre renversements

L’ontologie processuelle inverse chacun de ces présupposés.

Premier renversement : la donnée n’existe pas avant son annotation. Ce qui existe, c’est le flux — un événement physique, un comportement, une transaction, un battement cardiaque. Cet événement, pris en lui-même, n’est ni une donnée ni une non-donnée : il *est*. La donnée naît du geste qui le détache du flux pour lui assigner une représentation symbolique. Ce

geste est constitutif. Sans lui, il y a un fleuve, mais aucun poisson tiré du courant .

Deuxième renversement : la donnée se périmé. Dès l’instant où elle est inscrite, le flux dont elle provient continue son cours tandis qu’elle, gravée dans son support, ne change pas. Cet écart entre la mobilité du flux et la fixité de la donnée se creuse à chaque seconde. La donnée vieillit non parce qu’elle change, mais parce que le monde a continué sans elle.

Troisième renversement : la valeur de la donnée n’est pas intrinsèque. Elle dépend de trois variables conjointes : la proximité de son inscription au présent du flux (sa *fraîcheur*), la densité du corpus auquel on peut la rattacher (son *voisinage référentiel*), et la qualité de l’opération d’annotation qui l’a produite (sa *probité*). Une donnée fraîche d’aujourd’hui peut être plus précieuse qu’un volume considérable stocké il y a dix ans, parce que l’une est encore proche de son flux, et l’autre s’en est définitivement détachée.

Quatrième renversement : la donnée n’appartient pas vraiment. On peut posséder le support, on peut posséder l’annotation — mais on ne possède pas le flux duquel elle a été tirée et qui continue, indifférent. Cette distinction entre la propriété du *prélèvement* et l’inappropriabilité du *flux* est centrale ; nous y reviendrons.

Pourquoi l’ontologie objectale a triomphé

Si l’ontologie processuelle est philosophiquement plus juste, comment se fait-il que ce soit l’ontologie objectale qui structure aujourd’hui la pensée de la donnée, le droit, l’économie et même la littérature ? Quatre raisons convergent.

Une raison philosophique. La métaphysique occidentale, depuis Aristote, privilégie la substance sur le processus. Une chose *est* d’abord, elle *change* ensuite. Cette préséance de l’être sur le devenir a été contestée — par Héraclite avant Parménide, puis épisodiquement par Leibniz, Bergson, James, Whitehead, Deleuze, Simondon — mais elle est restée majoritaire.

Une raison technique. L’informatique fondationnelle, depuis la machine de Turing jusqu’au modèle relationnel d’Edgar Codd (1970), pense en termes de registres, de mémoires, de tables : des structures statiques où la donnée occupe une *place*. Les architectures de traitement en continu (*stream processing*) restent minoritaires face au modèle dominant du traitement par lots — où les flux d’eau vive sont reçus en continu mais traités par blocs figés, comme si des ruisseaux étaient transformés en cubes de glace pour être comptés. Et pourtant, les flux d’observation systématiques précèdent l’informatique de plusieurs millénaires : les journaux astronomiques babyloniens (VIII^e siècle av. J.-C.) constituaient déjà des prélèvements cadencés sur des décennies.

Une raison économique. La donnée-objet est valorisable comptablement, peut figurer à l’actif d’un bilan, peut être vendue. La donnée-flux est

beaucoup plus difficile à monétiser, parce qu'elle se périme. Les modèles économiques des grandes plateformes numériques reposent sur l'illusion que les données accumulées hier valent encore demain.

Une raison politique. L'ontologie objectale permet la propriété ; la propriété permet le contrôle ; le contrôle permet le pouvoir. Si l'on admettait que la donnée n'existe vraiment qu'au moment de son prélèvement, et qu'elle se périme inexorablement, la question politique se déplacerait : ce ne serait plus *qui possède les données*, mais *qui a accès au flux*, *qui décide de ce qui est prélevé*. Cette question dérange ; il est confortable de ne pas la poser.

Partie II — Le flux comme primat ontologique

Le flux : une description, pas une métaphore

Le flux n'est pas une métaphore. C'est une description littérale de la condition fondamentale du réel sous laquelle toute donnée est produite.

Le flux possède quelques propriétés structurantes. Il est *continu* : entre deux mesures, la réalité mesurée n'a pas cessé d'exister. Entre deux clics de l'utilisateur, le comportement n'a pas été suspendu. Il est *dense* : entre deux instants, aussi proches soient-ils, le nombre d'événements est potentiellement infini. Aucune annotation n'épuise sa source. Il est *irréversible* : l'instant *t* passé est passé pour toujours. Tout prélèvement non effectué à *t* est définitivement perdu — on peut le reconstituer par inférence, on ne peut pas le ressusciter. Et il est *indépendant* : le flux est indifférent à l'annotation. Qu'on regarde ou non, qu'on mesure ou non, il continue. La donnée doit beaucoup à l'annotateur ; le flux ne lui doit rien.

Ce dernier point implique que le volume de données produit par l'humanité — aussi considérable soit-il, aussi vite qu'il croît — ne représente pas la production *du flux* mais la production de *prélèvements sur le flux*. Le flux lui-même n'est pas mesurable ; il est, par définition, ce dont on tire les mesures. La civilisation contemporaine ne produit pas de plus en plus de réel : elle documente de plus en plus finement certaines portions de ce réel, en laissant le reste s'écouler dans l'obscurité.

Une précaution s'impose ici. Quand cet essai parle du « flux du réel », il n'engage pas une thèse métaphysique sur la nature ultime du réel. Que cette caractéristique appartienne au réel lui-même ou à notre condition de le percevoir, l'essai tient dans les deux cas. La précaution est posée dans la préface méthodologique ; elle vaut pour tout ce qui suit.

Whitehead et le *datum*

Le philosophe britannique Alfred North Whitehead, dans son maître-livre *Process and Reality* (1929), construit une métaphysique dans laquelle,

contre toute la tradition aristotélicienne, le devenir est plus fondamental que l'être. Whitehead nomme l'unité élémentaire du réel non pas *substance* mais *occasion actuelle* — un événement ponctuel de devenir, qui s'actualise puis devient ressource pour les occasions suivantes.

Et voici la trouvaille remarquable pour notre propos : pour désigner ce qu'une occasion actuelle *transmet* aux occasions suivantes, Whitehead précise philosophiquement le mot *datum*. Dans sa terminologie, le *datum* est "ce qui est donné" à une occasion ultérieure par les occasions qui la précèdent. Ce terme ne désigne pas une chose ; il désigne un *résidu de processus* — ce qui demeure d'une occasion advenue et qui peut être repris dans l'occasion suivante. Le *datum* est donc, par essence, *rétrospectif* : c'est ce qui *a été donné*, pas ce qui *est*. Et il est *relationnel* : il n'a de signification que pour l'occasion qui le reprend.

Cette intuition — formulée près d'un siècle avant l'ère du *Big Data* — est probablement plus juste que toutes les définitions managériales contemporaines de la donnée. Elle dit explicitement ce que la littérature standard occulte : la donnée est ce que le passé lègue au présent, sous forme d'un fragment isolé qui devra être recontextualisé pour reprendre sens.

Bergson et la durée : ce que la mesure perd

Henri Bergson, dans l'*Essai sur les données immédiates de la conscience* (1889), distingue rigoureusement deux modes du temps. Le *temps spatialisé* est celui de l'horloge, de la mesure, de l'inscription : découpé en instants discrets, juxtaposés comme des points sur une ligne. C'est le temps des données. La *durée pure* est celle du vécu, de la conscience, du processus : continue, qualitative, indivisible. C'est le temps du flux.

Bergson soutient que la mesure scientifique procède par réduction de la durée au temps spatialisé. Pour mesurer, il faut figer ; pour figer, il faut interrompre la continuité. La donnée est ce qui reste de la durée après qu'on l'a spatialisée. Elle conserve la grandeur, elle perd le mouvement.

Cette perte n'est pas une critique de la science. Bergson rappelle simplement que *la mesure ne contient pas tout ce qu'elle mesure*. Cette intuition devrait être au fondement de toute littérature sérieuse de la donnée. Elle signifie qu'aucune base de données, même exhaustive, ne *contient* le réel qu'elle prétend décrire. Elle en contient des coupes, des instantanés. Le reste — la durée elle-même — est resté dans le flux, où il a continué sans nous.

Partie III — La métaphore de la sauvegarde

L'analogie la plus éclairante pour saisir la nature processuelle de la donnée nous vient d'un objet familier de l'informatique : la sauvegarde de fichier.

Soit un fichier F présent sur un disque, et une sauvegarde S de ce fichier effectuée à l'instant t_0 sur un disque externe. À l'instant t_0 , F et S sont identiques. À l'instant $t_1 > t_0$, le fichier F a été modifié, mais S n'a pas changé : ils ont commencé à diverger. À l'instant $t_2 \gg t_0$, les modifications se sont accumulées ; S , qui était la copie de F , n'en est plus qu'une archive d'un état antérieur. Pour disposer d'une sauvegarde à jour, il faut effectuer un *nouveau* prélèvement.

Cette analogie transpose directement à la donnée. Le fichier F est le flux du réel : il évolue en permanence, sans s'arrêter. La sauvegarde S est la donnée : figée au moment de son prélèvement. Dès l'instant où elle est inscrite, elle commence à diverger du flux. La divergence s'accroît avec le temps — lentement pour les flux lents (la géologie, la démographie), rapidement pour les flux rapides (les marchés financiers, le comportement en ligne).

L'analogie nomme avec précision ce qui se passe dans les systèmes algorithmiques contemporains. Un grand modèle de langage, entraîné sur un corpus figé à une date donnée, est littéralement une sauvegarde d'un état du flux textuel mondial. À chaque instant qui passe, ce corpus s'éloigne du flux qui, lui, continue. La technique qui consiste à brancher ces modèles sur des recherches en ligne en temps réel est une tentative de *re-prélever* dans le flux pour compenser ce décalage — reconnaissant implicitement, par son existence même, que le modèle sans elle serait une sauvegarde périmée.

On ne demande pas d'une sauvegarde qu'elle soit belle, complète ou intelligente. On lui demande d'être *fidèle au moment de sa création* et d'être *fraîche pour l'usage qu'on en fait*. Une sauvegarde datant d'hier est précieuse si le fichier original a été endommagé aujourd'hui. Une sauvegarde datant de cinq ans est inutile pour restaurer un fichier que l'on travaille depuis lors — elle représente un état révolu. La valeur d'une donnée n'est donc pas dans la donnée : elle est dans le rapport entre la donnée, le flux dont elle a été tirée, et l'usage qu'on en veut faire. Toute donnée est un point ; tout réel est une ligne.

Partie IV — L'acte d'annotation

Annoter, c'est choisir

Si la donnée naît du geste qui prélève, alors ce geste appelle examen. Annoter, c'est toujours choisir : choisir *quoi* observer dans le flux (l'objet), *comment* l'observer (la méthode), *quand* l'observer (la fréquence), *avec quels instruments* (la médiation technique), *selon quelles catégories* (la grille).

Chacun de ces choix opère une réduction. L'objet choisi exclut tous les autres objets possibles ; la méthode retenue exclut toutes les autres

méthodes ; la fréquence laisse hors-champ tout ce qui se passe entre deux observations ; la grille classificatoire impose ses distinctions à un flux qui n'en avait pas.

Chaque prélèvement est un *acte composé* : le choix de l'objet, le choix du moment, le choix de l'instrument, le choix de la catégorie. Supprimer n'importe lequel de ces choix, c'est supprimer la donnée elle-même. La donnée est un acte, pas un fait pur.

C'est pourquoi l'idée d'une "donnée brute" — *raw data* — est un oxymore, comme l'a montré la collection d'études historiques réunies par Lisa Gitelman. Toute donnée a été travaillée par quelqu'un, quelque part, avec des instruments et des catégories qui portent leurs propres choix. Cela était vrai pour les tablettes cunéiformes d'Uruk, qui classaient les moutons selon des catégories administratives précises. Cela est vrai pour les capteurs d'un smartphone, qui mesurent certaines fréquences d'accélération et pas d'autres. Il n'y a pas de données innocentes — il y a des données dont on a oublié l'auteur.

Le non-prélevé

Tout ce que l'on choisit d'annoter exclut une infinité d'autres choix possibles. Le non-prélevé est, par définition, invisible dans la base de données ; il ne fait pas partie de l'analyse ; il n'est pas comptabilisé dans les statistiques. Il reste dans le flux, indistinct.

Cette zone d'ombre — ce que les chercheurs en *data studies* appellent les *missing data* ou *uncollected data* — est constitutive de toute base de données. Une base est définie autant par ce qui s'y trouve que par ce qui n'y figure pas. Un recensement qui ne pose pas de question sur la langue parlée à la maison rend invisible la diversité linguistique d'une population. Un capteur urbain placé dans les quartiers aisés et absent des quartiers pauvres produit une "ville statistique" biaisée. Un corpus d'entraînement constitué principalement de textes en langue anglaise produit un modèle qui "parle" certaines langues mieux que d'autres.

Le non-prélevé n'est pas le contraire des données : c'est leur doublure invisible. Toute donnée se découpe sur un fond de non-données. La littératie de la donnée devrait inclure une *littératie du non-prélevé* — la capacité à se demander, devant n'importe quel jeu de données : qu'a-t-on choisi de ne pas mesurer ? qu'a-t-on choisi de ne pas voir ? qui en a décidé, et dans quel contexte ?

Qui annote ?

Le flux est indifférent ; l'annotation, jamais. Elle a un sujet — donc un point de vue, donc des intérêts.

Quatre grandes classes d'annotateurs structurent le paysage contemporain. Les *annotateurs humains experts* — scribes, comptables, médecins,

statisticiens, scientifiques, journalistes — sont lents et coûteux, mais réflexifs et capables d'auto-correction : ils savent, en principe, qu'ils annotent et pourquoi. Les *annotateurs humains involontaires* — chacun de nous, quand nous remplissons un formulaire, cliquons sur un site, géolocalisons une application — produisent des annotations sans nécessairement le savoir et le plus souvent sans en saisir l'usage ultérieur. Les *capteurs automatiques* — thermomètres, caméras, satellites, GPS — sont rapides et peu coûteux, mais aveugles à leur propre cadre : un capteur placé au mauvais endroit produit des données impeccablement précises et complètement fausses. Les *algorithmes d'annotation* — reconnaissance d'images, transcription audio, classification de textes — sont puissants mais reproduisent et amplifient les biais de leurs corpus d'entraînement.

Chaque classe d'annotateur produit des données aux caractéristiques différentes : granularité, fraîcheur, fiabilité, biais. Une littérature complète de la donnée devrait inclure la capacité à identifier, devant toute donnée, *quel type d'annotateur l'a produite*, et à anticiper les biais propres à ce type.

Partie V — Le corpus comme condition d'intelligibilité

Une donnée seule ne dit rien

L'ontologie processuelle radicalise une intuition déjà présente dans la pyramide DIKW. Elle va plus loin : la donnée n'est pas seulement insignifiante hors contexte, elle est *proprement vide* hors corpus.

Une mesure de température — “37,2 °C” — n'est pas une information à laquelle il manquerait du contexte. C'est un symbole sans référence, qui attend d'être branché sur d'autres mesures de température (la série temporelle du même sujet), des normes (la plage normale chez l'humain adulte), des classifications (les seuils de la fièvre), d'autres données concomitantes, un sujet identifié, un horodatage. Ces éléments forment le *corpus* de la donnée : l'ensemble des autres prélèvements qui lui donnent sens. La donnée seule est un mot sans phrase ; le corpus est la grammaire qui la rend lisible.

Il s'ensuit que la valeur d'une donnée individuelle dépend très largement de la densité du corpus dans lequel elle s'inscrit. La même mesure de température dans un hôpital disposant de millions de mesures historiques et de normes cliniques établies, et dans un contexte isolé sans comparaison possible, n'a pas la même valeur — même si elle est identique en tant que signal. Ce n'est pas la donnée qui change, c'est le corpus qui l'illumine ou l'éteint.

On voit alors apparaître un effet de rendement croissant : plus le corpus est dense, plus chaque donnée nouvelle est précieuse. C'est l'un des moteurs

économiques du *Big Data*. Et ce rendement est auto-entretenu : un corpus riche attire des données de meilleure qualité, qui enrichissent le corpus, qui augmente encore sa valeur. Inversement, les organisations à faible corpus subissent un cercle vicieux où leurs données isolées valent peu. Cette dynamique éclaire pourquoi la concentration mondiale des données entre les mains de quelques grandes plateformes et de quelques États n'est pas seulement un phénomène de volume — c'est un phénomène d'écart de valeur qui s'auto-entretient.

La mesure comme constellation de prélèvements

Il importe de distinguer la *donnée brute* (le prélèvement) de la *mesure interprétée* (la donnée mise en relation à un référentiel). Le thermomètre indique "37,2". C'est un signal. Pour qu'il devienne mesure, il faut une unité, une référence physique, une calibration de l'instrument, un sujet. Chacun de ces éléments est lui-même un *autre prélèvement*, opéré sur un *autre fragment du flux*, qui vient cadrer le prélèvement principal. Une mesure n'est pas une donnée pure : c'est une constellation de prélèvements articulés.

Cela signifie qu'une donnée n'est jamais qu'à moitié donnée. Elle exige toujours, pour fonctionner, d'autres données. La sauvegarde ne devrait donc pas préserver *un* fichier, mais le *système de fichiers* dans lequel il s'inscrit — c'est précisément ce que les bibliothécaires appellent les métadonnées et les archivistes le *contexte de production*.

Partie VI — La péremption temporelle

La demi-vie de la donnée

L'industrie elle-même a fini par reconnaître la péremption — sous le nom de *data freshness* ou *data decay*. Plusieurs auteurs ont proposé de doter chaque type de donnée d'un attribut intrinsèque de décroissance, sur le modèle de la demi-vie radioactive. La donnée, comme un atome instable, se désintègre. Elle ne disparaît pas physiquement — elle reste, gravée dans son support — mais elle perd sa capacité à représenter le flux dont elle a été extraite.

Les taux de décroissance varient considérablement selon le type de flux. Très lents pour les données géologiques ou démographiques structurelles. Lents pour les données patrimoniales stables. Moyens pour les données socioprofessionnelles. Rapides pour les données comportementales. Très rapides pour les données financières, météorologiques, médicales aiguës, ou de trafic réseau. La donnée n'est pas persistante. Elle est périssable — comme un fruit, comme un poisson hors de l'eau.

Trois âges de la donnée

À partir de cette propriété, on peut distinguer trois âges dans la vie d'une donnée, correspondant à trois usages différents.

Premier âge — la donnée fraîche. Tant que l'écart entre la donnée et le flux dont elle provient reste suffisamment petit, la donnée peut être utilisée comme représentation du réel actuel. C'est l'usage opérationnel : prendre une décision, formuler une recommandation, déclencher une action. Un capteur de température dont la dernière mesure date de deux minutes peut fonder une décision médicale immédiate.

Deuxième âge — la donnée mûrissante. À mesure que la donnée s'éloigne de son flux, elle perd sa capacité représentationnelle pour le présent, mais conserve une utilité statistique : agrégée à d'autres données du même type, elle alimente des modèles, révèle des tendances, soutient des prédictions. Ce n'est plus la réalité actuelle qu'elle représente, mais un comportement passé dont la trace peut informer une projection.

Troisième âge — la donnée archive. Une fois trop éloignée du flux pour avoir une utilité même statistique sur le présent, la donnée devient pure trace du passé. Elle ne représente plus, elle *commémore*. Sa valeur n'est plus opérationnelle ni prédictive, mais historique.

Cette distinction a une conséquence pratique : beaucoup de problèmes contemporains de gouvernance des données viennent de ce qu'on confond ces trois âges. On prend pour de l'opérationnel ce qui n'est plus que de l'archive ; on relègue à l'archive ce qui pourrait encore servir. On évalue la dangerosité d'un individu sur la base d'un dossier vieux de dix ans. On solvabilise un client sur la base d'un historique qui ne reflète plus sa situation. On répond à une question sur le présent avec un modèle entraîné sur un corpus figé deux ans plus tôt. Et l'on redécouvre, sous couvert d'innovation, d'anciennes pratiques — réponses éprouvées à des contraintes climatiques, alimentaires, d'aménagement ou de bien-être — qu'on avait reléguées à l'archive et qui n'avaient rien perdu de leur valeur opérationnelle.

Le paradoxe de la conservation

Reste un paradoxe : pourquoi conserver une donnée périmée ? Trois bonnes raisons coexistent. L'intérêt statistique : même périmée pour l'opérationnel, la donnée enrichit le corpus historique qui éclaire les données présentes. L'intérêt mémoriel : conserver est une fonction culturelle en soi ; les sociétés archivent pour pouvoir se penser dans le temps. L'intérêt comparatif : confronter l'état initial et l'état actuel permet de mesurer un changement.

Ces raisons légitimes ne doivent pas masquer un usage problématique : conserver une donnée périmée *en faisant comme si elle était encore opérationnelle*. C'est la confusion des âges, et elle est l'un des risques majeurs de la gouvernance contemporaine des données.

Partie VII — Du flux d'observation : le maillon manquant

La limite de la donnée isolée

L'ontologie processuelle reconstitue bien ce qu'est une donnée : un prélèvement sur le flux, daté, partiel, portant les marques de son auteur. Mais quelque chose manque. Une donnée seule, même bien comprise, ne rend pas compte de ce qui se passe lorsque les prélèvements deviennent systématiques. Entre l'entaille sporadique sur un os préhistorique et les capteurs IoT qui produisent en continu des millions de points par seconde, il y a un changement qualitatif, pas seulement quantitatif. À un certain seuil de densité et de cadence, on ne change pas seulement de quantité de données : on change de nature.

Ce quelque chose d'autre a un nom : le **flux d'observation**.

Le triptyque complet

Là où l'ontologie processuelle initiale posait deux plans — le flux du réel et la donnée comme prélèvement sur ce flux — l'ontologie complète en requiert trois.

Le **flux du réel** est ce qui se passe, indépendamment de toute annotation humaine. Continu, immense, indifférent. Le réel en tant qu'il s'écoule, tel que Héraclite le pensait, tel que Bergson le nommait durée.

Le **flux d'observation** est un flux construit par une entité humaine — ou par une organisation, un dispositif technique sous contrôle humain — par accumulation de prélèvements cadencés effectués sur le flux du réel selon des critères intentionnels. Le flux d'observation n'est pas le flux du réel : c'est une dérivation partielle, fidèle aux critères qui l'ont défini, et manipulable. On peut le stocker, le rejouer, le consolider, le transmettre, l'interrompre. Il dure tant qu'on l'entretient.

La **donnée consolidée** est le grain informationnel synthétique extrait du flux d'observation par traitement — moyenne, indicateur, statut, équation — à haute valeur d'usage et réutilisable dans d'autres contextes.

Le triptyque *flux du réel* → *flux d'observation* → *donnée consolidée* constitue la chaîne complète de production de la donnée selon l'ontologie processuelle.

Ce que le flux d'observation éclaire

Introduire ce maillon manquant a plusieurs conséquences importantes.

Il rend l'ontologie *opérationnellement praticable*. Si toute donnée se périme, on pourrait craindre que prélever soit vain. La réponse est

: on prélève en continu, pour reconstituer un flux. La péremption individuelle est compensée par la cadence collective. L'ontologie passe d'une philosophie de la perte (le flux nous échappe) à une philosophie de la reconstruction partielle (nous bâtissons des canaux qui en captent ce qui nous est utile).

Il *réconcilie l'ontologie avec la pratique technique*. L'industrie du traitement en continu construit déjà, sans le théoriser ainsi, des flux d'observation — comme le faisaient avant elle les registres paroissiaux du XVI^e siècle, et avant eux les journaux astronomiques babyloniens depuis le VIII^e siècle avant notre ère.

Il *déplace la question de propriété*. L'humain n'est pas central parce qu'il possède la donnée, mais parce qu'il *construit les canaux* par son intention. Il n'a pas le réel ; il a des intentions, et ses intentions structurent ce qu'il en capte. Cette formulation — l'humain comme constructeur de canaux, pas comme propriétaire des eaux — ouvre une économie politique radicalement différente de celle de la donnée-objet.

La définition complète de la donnée selon l'ontologie processuelle peut maintenant être énoncée :

La donnée est le grain élémentaire d'un flux d'observation que l'humain construit par son intention, à partir du flux du réel. Ce flux d'observation est fidèle au flux du réel relativement aux critères qui l'ont défini, sans jamais en être une réplique totale. De ce flux d'observation peuvent être extraites des données consolidées à haute valeur d'usage, qui alimentent décisions et connaissances, et qui peuvent à leur tour circuler entre entités pour engendrer de nouveaux flux d'observation dérivés.

Les propriétés du flux d'observation — ses critères de construction, ses analogies, sa dynamique inter-entités, les compétences qu'il implique — sont l'objet du document *Creuser le canal*.

Partie VIII — Conséquences plus larges

Pour les systèmes algorithmiques

L'ontologie processuelle éclaire les grandes tensions contemporaines autour des systèmes d'intelligence artificielle. Un grand modèle de langage est, architecturalement, une sauvegarde : son corpus d'entraînement est daté, sa connaissance s'arrête à une certaine date, et toute interrogation portant sur des événements ultérieurs révèle le décalage structurel entre la donnée figée et le flux qui a continué. Les techniques qui consistent à le brancher en temps réel sur des sources extérieures sont des tentatives de rebranchement sur le flux. Elles

témoignent, par leur existence, de ce que l'ontologie processuelle permet de nommer : le modèle sans elles est une sauvegarde qui diverge.

La question philosophique ouverte — peut-on imaginer un système d'intelligence qui apprendrait du flux du réel en continu, sans corpus figé ? — appelle une réponse que le concept de flux d'observation permettra de préciser : un tel système ne pourrait pas apprendre directement du flux du réel, car ce flux n'est pas directement observable. Il apprendrait des flux d'observation que des humains maintiennent actifs. L'intelligence du système dépendrait alors de la qualité des canaux — et donc, en dernière instance, de l'intention de ceux qui les creusent.

Pour le droit des données

Le droit des données personnelles — illustré par le RGPD européen et ses équivalents mondiaux — repose tout entier sur l'idée que la donnée est un objet appartenant à la personne concernée. Cette personne peut consentir à son traitement, accéder à ses données, les rectifier, les effacer. Toutes ces opérations supposent un objet stable.

L'ontologie processuelle perturbe ce cadre. Si la donnée est un prélèvement sur le flux de mon comportement, le rectifier ne change rien au flux — qui a déjà continué. Le “droit à l'oubli” est, à proprement parler, un droit à la suppression des prélèvements, non un droit à l'arrêt du flux. Cette distinction ne disqualifie pas le droit existant, mais elle suggère des évolutions qui en dépassent les catégories actuelles : un droit au non-prélèvement (en amont de l'acte, et non plus seulement de ses suites), un droit à la péremption automatique des prélèvements au-delà d'une certaine fraîcheur (attaché à la nature de la donnée, non à la demande de la personne), un droit à la transparence sur le geste d'annotation lui-même (qui, comment, selon quels critères).

Pour l'éthique du stockage et la politique des communs

La doctrine “stocker tout, on saura quoi en faire plus tard” suppose que les données conservées gardent leur valeur. L'ontologie processuelle suggère un renversement : stocker ce dont on a besoin, libérer le reste. Puisque la donnée se périmé, son stockage massif est largement illusoire — on accumule des sauvegardes qui divergent du fichier à mesure qu'on les garde. Le coût environnemental de cette accumulation — en surface, en énergie, en eau — est réel et croissant. Il ne se justifie que si la valeur informationnelle des stocks perdure dans le temps.

Plus fondamentalement, l'ontologie processuelle reformule la question de l'appropriation. Si la donnée est un prélèvement sur le flux du comportement humain collectif, et si ce flux n'appartient à personne, alors la propriété revendiquée sur “leurs” données par les grandes plateformes apparaît sous un jour nouveau. Ces données ne sont pas leur création : elles sont leurs *prélèvements* sur un commun. Le flux du

comportement humain, comme l'air ou l'eau, est un bien commun au sens d'Elinor Ostrom. Les prélèvements opérés sur lui peuvent être appropriés en tant que prélèvements particuliers, mais ne sauraient priver les autres acteurs du droit d'opérer leurs propres prélèvements.

Tant qu'on pense la donnée comme un objet, on est condamné à une économie politique de la propriété — où les positions oligopolistiques paraissent légitimes. Si on pense la donnée comme un prélèvement sur un commun, l'économie politique devient celle des droits d'usage — où le débat démocratique reprend ses droits sur la question : qui construit les canaux, selon quelles règles, au service de quels intérêts ?

Conclusion — Le filet et le fleuve

Le réel est un fleuve. La donnée est ce qu'un filet en tire. Notre civilisation a passé deux siècles à raffiner ses filets, à les multiplier, à les automatiser. Elle a oublié de regarder le fleuve.

Cet oubli est philosophiquement compréhensible : la métaphysique occidentale a privilégié la substance contre le processus, l'être contre le devenir, l'objet contre le flux. Mais réintroduire le fleuve dans notre pensée de la donnée n'est pas un geste de nostalgie philosophique. C'est une condition pratique : comprendre ce que l'on a vraiment entre les mains quand on travaille avec des données, ce qu'on peut en inférer et ce qu'on ne peut pas, ce qu'on possède et ce qu'on ne possède pas.

Reformulons en résumé l'essentiel.

Qu'est-ce qu'une donnée ? Non pas un objet préexistant qu'on collecte, mais un prélèvement opéré sur un flux continu du réel par un acte d'annotation qui le détache de ce flux pour le fixer dans une représentation symbolique — grain élémentaire d'un flux d'observation que l'humain construit par son intention.

Quelle est sa valeur ? Non pas une valeur intrinsèque et durable, mais une valeur déterminée conjointement par sa fraîcheur, la densité du corpus qui la rend interprétable, et la qualité du geste d'annotation qui l'a produite.

Quel est son destin ? Non pas la conservation indéfinie, mais une trajectoire en trois âges — opérationnel quand elle est fraîche, statistique quand elle mûrit, mémoriel quand elle est devenue trop éloignée du flux pour le représenter encore.

De là, deux questions s'ouvrent, que l'essai traite chacune en son document. Celle du temps : comment une chose figée peut-elle vieillir ? (*Les deux temps de la donnée.*) Et celle du canal : si les prélèvements systématiques constituent des flux d'observation, comment le construit-on ? Selon quels critères ? Avec quelles propriétés ? Et avec quels effets sur

ce qu'il permet de voir — et sur ce qu'il cache ? Creuser ce canal avec discernement est l'objet de l'exploration qui suit.

Tout s'écoule, rien ne demeure. — Héraclite

Et nos données, à mesure que le temps passe, s'éloignent inexorablement du flux qu'elles prétendent encore représenter.

Repères

Pour les références philosophiques (Whitehead, Bergson, Héraclite, Deleuze, Simondon), épistémologiques (Gitelman, Bowker, Leonelli, Berry) et politiques (Zuboff, D'Ignazio, Ostrom) mobilisées dans ce document, voir la *Bibliographie raisonnée*.

Sur la dualité flux/lot dans l'industrie des données : la coexistence entre traitement par lots (*batch processing*) et traitement en continu (*stream processing*) est documentée par les plateformes techniques (Confluent, Pathway, Apache Flink). Sur la métaphore de la demi-vie appliquée à la donnée : voir l'entrée correspondante dans la bibliographie raisonnée.

Les deux temps de la donnée

L'acte, le support, et le courant qui fait vieillir ce qui ne bouge pas

« *Le temps est invention ou il n'est rien du tout.* »

— Henri Bergson, *L'Évolution créatrice* (1907)

Préambule — Un paradoxe à résoudre

Les documents précédents ont, à plusieurs reprises, énoncé deux affirmations qui paraissent se contredire.

D'un côté, la donnée est **figée, inerte, arrêtée**. Une fois prélevée et inscrite, elle ne change plus. « La température relevée est de 38,2 °C » : cela restera vrai pour toujours, gravé, définitif. C'est même cette fixité qui fait l'utilité de la donnée — on peut la consulter, la transmettre, la rejouer, parce qu'elle ne bouge pas.

De l'autre côté, ces mêmes documents ont dit que la donnée **se périme**, qu'elle **dépérit**, que sa **fraîcheur varie**. Avec le temps, elle représente de moins en moins fidèlement le réel dont elle a été extraite. Elle vieillit.

Comment une chose figée peut-elle vieillir ? Tant que ce paradoxe apparent n'est pas résolu, il fragilise tout l'édifice de l'ontologie processuelle. Si la donnée est figée, elle ne peut pas se périmer. Si elle se périme, c'est qu'elle n'est pas vraiment figée. Il faut choisir — ou bien dissoudre la contradiction.

Ce document a pour objet de la dissoudre. Sa thèse tient en une phrase : **la donnée n'a pas un temps, elle en a deux**. Dans l'un, elle est éternelle. Dans l'autre, elle vieillit. Et ces deux temps ne se contredisent pas, parce qu'ils ne mesurent pas la même chose.

Établir cette thèse demande de remonter à la fabrique même de la donnée — aux deux opérations dont elle résulte. Car si la donnée a deux temps,

c'est qu'elle naît de deux gestes : un **acte** qui la fait naître, un **support** qui la fait durer. Le premier l'ancre dans le temps du réel ; le second l'installe dans le temps de l'inscription. La donnée est la tension entre ces deux ancrages — et sa péremption, nous le verrons, n'est rien d'autre que l'éloignement inéluctable du premier, à mesure que le flux qui le portait continue de couler.

Ce document, comme l'ensemble de l'essai, applique la distinction fond/forme posée dans la Préface méthodologique. Le vocabulaire de « référentiel » employé épisodiquement dans les documents 3 et 4 — au sens du cadre de la donnée — est ici abandonné au profit de la distinction bergsonienne plus juste des deux temps.

Partie I — Les deux opérations qui font une donnée

Une donnée ne naît pas d'un seul geste

L'intuition courante se représente la production d'une donnée comme un geste unique : on « relève » une mesure, on « collecte » une information, on « enregistre » un fait. Un seul verbe, un seul mouvement.

Cette représentation est trop sommaire. Produire une donnée, c'est accomplir **deux opérations de natures différentes**, que la rapidité des dispositifs modernes a fondues au point de les rendre indistinctes, mais qui restent logiquement séparées.

La première consiste à **arrêter une valeur**. Dans le flux du réel, la température d'un corps varie sans cesse, sans jamais s'immobiliser sur aucune valeur. Arrêter une valeur, c'est décider que, à cet instant précis, cette grandeur continue *vaut* 38,2 — c'est pratiquer une coupe dans un écoulement qui ne comporte, de lui-même, aucune coupure. Nous nommerons cette première opération l'**acte d'inscription**.

La seconde consiste à **fixer durablement** la valeur arrêtée. Arrêter une valeur ne suffit pas : il faut encore que cet arrêt persiste, qu'il ne soit pas aussitôt repris par l'écoulement. Il faut confier la valeur à quelque chose qui la maintiendra — une entaille dans un os, un signe sur une tablette, une trace sur un disque. Nous nommerons cette seconde opération la **fixation sur un support**.

Ces deux opérations sont distinctes. La première est un geste de **découpe** : elle sépare un instant du continu. La seconde est un geste de **conservation** : elle arrache l'instant découpé à l'usure du temps. On peut concevoir l'une sans l'autre — un calcul mental arrête une valeur sans la fixer ; une bande magnétique vierge est un support qui n'a encore rien arrêté. Mais une donnée, pour exister pleinement, a besoin des deux.

L'acte d'inscription : la coupe dans le continu

L'acte d'inscription est l'instant où une valeur cesse d'être un moment du flux du réel pour devenir une donnée. C'est un acte de **découpe**, et cette découpe mérite qu'on en mesure la portée.

Le flux du réel est **continu**. La température d'un corps ne procède pas par paliers ; elle varie de manière indivisible, sans jamais se poser. Entre 38,1 et 38,3, elle est passée par une infinité de valeurs, dont aucune n'a duré le moindre instant assignable. Le réel, en lui-même, **ne contient pas de 38,2 °C** : il contient un devenir thermique qui ne s'arrête sur rien.

L'acte d'inscription impose un arrêt là où le réel n'en connaît aucun. Il dit : *ici, maintenant, c'est 38,2*. Il extrait du devenir une valeur stable, une grandeur nette, un nombre. Ce faisant, il accomplit ce que Bergson nommait la **spatialisation du temps** : il transforme la durée — continue, qualitative, indivisible — en une suite de points fixes, comparables, manipulables. La donnée naît exactement là, au point où la durée pure est convertie en temps spatialisé.

Cette conversion a un prix. Elle gagne la maniabilité — un nombre, on peut le comparer, le calculer, le transmettre — et elle perd le mouvement. La donnée conserve la grandeur ; elle abandonne le devenir dont cette grandeur fut un instant. La donnée est une trace, et toute trace est plus pauvre que le pas qui l'a produite.

Deux propriétés de l'acte d'inscription doivent être retenues, car elles commanderont toute la suite.

L'acte est daté. Il a lieu à un instant précis du flux du réel. Il prélève *ce* moment-là, et pas un autre. La donnée porte donc en elle, indélébile, la marque de l'instant où l'acte a eu lieu. C'est par rapport à lui que se mesurera ensuite tout le vieillissement.

L'acte est unique et non répétable. Le geste qui a fixé « 38,2 » à l'instant *t* est lui-même un événement du flux du réel — il appartient au flux, il s'y est produit une fois, il ne s'y reproduira jamais à l'identique, puisque le flux a continué. On ne peut pas refaire le même acte d'inscription : on ne peut qu'en accomplir un nouveau, à un nouvel instant, sur un réel qui a changé. Cette propriété, en apparence abstraite, a une conséquence pratique majeure que nous retrouverons : la fraîcheur perdue ne se rattrape pas, car l'acte qui aurait pu la capter est passé.

La fixation : l'arrachement à l'écoulement

La fixation sur un support est l'opération par laquelle la valeur arrêtée est confiée à une matière qui la maintiendra.

Fixer une donnée, c'est l'**arracher à l'écoulement**. Tant que la valeur n'est qu'arrêtée — tenue dans une mémoire vive, dans un esprit, dans un signal fugace — elle reste menacée : le moindre incident la dissout. La fixation la met à l'abri. Elle inscrit la valeur dans quelque chose qui résiste

au temps : la dureté d'un os, la cuisson d'une tablette d'argile, la stabilité magnétique d'un disque. À partir de cet instant, la donnée est **soustraite au flux**. Elle ne s'écoule plus avec lui. Elle demeure.

C'est la fixation qui donne à la donnée ses propriétés les plus précieuses : la donnée fixée peut être **consultée** à volonté, **rejouée**, **transmise**, **dupliquée**. Toutes ces opérations supposent qu'un support la maintienne. Sans fixation, pas de consultation possible, pas de mémoire, pas de cumul, pas de flux d'observation. La fixation est la condition matérielle de tout usage différé de la donnée.

La fixation a ses propriétés propres, à énoncer avec netteté.

La fixation n'est jamais absolue : elle est relative à la résistance du support. Aucune donnée n'est figée « pour l'éternité » dans l'absolu. Elle est figée *autant que son support tient*. Une inscription tracée sur le sable d'une plage tient quelques instants ; une entaille gravée dans un os tient depuis près de quarante millénaires ; une trace sur un support magnétique tient quelques années avant de se dégrader. Le « figé » a donc lui-même une durée de vie. Il existe en réalité **deux morts possibles** de la donnée, et la défaillance du support est l'une d'elles — nous y reviendrons.

La fixation est indifférente au sens. Le support ne sait pas ce qu'il porte. La roche gravée ignore qu'elle porte une température ; le disque ignore qu'il porte un relevé médical. Toute l'intention, tout le sens, tout l'arbitraire de la donnée viennent de l'acte, jamais du support. Le support se contente de tenir.

La fixation est duplicable. On peut copier la valeur fixée d'un support vers un autre. La copie est rigoureusement identique à l'original. Ce point, anodin en apparence, ouvre un paradoxe que nous examinerons plus loin.

Partie II — Les deux temps

Pourquoi deux opérations engendrent deux temps

Chacune des deux opérations rattache la donnée à un temps différent.

L'**acte d'inscription** rattache la donnée au **temps du réel**. Car l'acte a lieu dans le flux du réel, à un instant daté de ce flux. Il prélève l'état du réel à *cet* instant. La donnée, par son acte d'origine, est donc indissolublement liée à un moment du flux du réel — le moment où la coupe a été pratiquée. Et ce flux, lui, continue de couler après l'acte. La donnée garde donc un pied dans un courant qui ne s'arrête pas.

La **fixation sur un support** rattache la donnée au **temps de l'inscription**. Car la fixation soustrait la donnée à l'écoulement et l'installe dans un support qui la maintient identique à elle-même. Dans ce temps-là — le temps propre du support, le temps de la chose inscrite — la donnée ne

bouge plus. Elle est ce qu'elle est, indéfiniment, aussi longtemps que le support tient.

Voilà la résolution du paradoxe du préambule. La donnée est figée *et* périssable parce qu'elle a un pied dans chacun de ces deux temps. Figée dans le temps de l'inscription, où le support la maintient immobile. Périssable dans le temps du réel, où le flux continue de s'éloigner du point que l'acte avait marqué. Les deux énoncés sont vrais en même temps, parce qu'ils parlent de deux temps différents.

Le temps de l'inscription : l'éternité de ce qui est gravé

Dans le temps de l'inscription, la donnée est **éternelle**. Cette éternité n'est pas une figure de style ; elle est exacte, à une condition près que nous lèverons dans la section suivante.

« La température relevée à l'instant t est de 38,2 °C. » Cet énoncé, une fois l'acte correctement accompli et la valeur correctement fixée, est vrai. Il l'était hier, il l'est aujourd'hui, il le sera dans mille ans. Aucun écoulement du temps ne le rendra faux. Le réel a beau continuer, le corps a beau se réchauffer ou se refroidir : *que la température relevée à l'instant t ait été de 38,2 °C* demeure un fait acquis, définitif.

C'est cette éternité qui fait que l'on peut **se fier** à une donnée, la **citer**, la **rejouer**, construire sur elle. Si la donnée changeait dans le temps de l'inscription, aucune mémoire, aucune comptabilité, aucune science ne seraient possibles. La civilisation de la donnée tout entière repose sur cette éternité-là.

Reste à bien comprendre ce qui est éternel. Ce n'est pas la valeur *en tant qu'elle décrit le réel actuel*. C'est la valeur *en tant qu'elle a été relevée à un instant donné*. L'éternité de la donnée est l'éternité d'un **constat daté**, non l'éternité d'une description valable pour toujours. La donnée ne dit pas « le corps est à 38,2 °C » ; elle dit « le corps fut relevé à 38,2 °C à l'instant t ». Le premier énoncé se périmé ; le second, jamais.

Le temps du réel : le dépérissement de ce qui ne bouge pas

Dans le temps du réel, la donnée **dépérit**. Et il faut être ici d'une grande précision, car c'est le point le plus subtil de tout le raisonnement.

La donnée, dans le temps du réel, ne bouge pas non plus. *Elle ne bouge pas*. La valeur 38,2 inscrite sur le support reste 38,2. Ce qui bouge, ce n'est pas la donnée — c'est le réel. Le corps continue de vivre, sa température continue de varier, le flux poursuit son cours. Et à mesure que le flux s'éloigne de l'instant t où l'acte avait planté son ancre, l'écart se creuse entre ce que la donnée affirme et ce que le réel est devenu.

Le dépérissement de la donnée n'est donc **pas un mouvement de la donnée**. C'est un mouvement *du réel* qui s'éloigne d'elle. La donnée est un point fixe ; le réel est un courant ; et ce que nous nommons «

péremption » ou « perte de fraîcheur » n'est rien d'autre que **la mesure de l'écartement croissant entre le point fixe et le courant**. Le dépérissement n'est pas une propriété de la donnée ; c'est une **relation** entre une donnée immobile et un réel mobile.

L'inversion : ce que la donnée fait au cadre scientifique classique

Cette situation met en jeu quelque chose que le cadre conceptuel habituel ne décrit pas.

Dans la méthode scientifique classique, on dispose d'un **cadre fixe** et l'on y mesure le **mouvement d'un objet**. La route est le cadre fixe, la voiture est l'objet mobile : on mesure la position changeante de la voiture par rapport à la route. Le ciel des étoiles lointaines est le cadre fixe, la planète est l'objet mobile. Toujours le même schéma : cadre stable, objet mobile, et c'est le mouvement de l'objet que l'on veut saisir.

La donnée renverse exactement ce schéma. Ici, **l'objet est fixe et le cadre est mobile**. La donnée — le prélèvement inscrit — est le point immobile. Le flux du réel — ce qui devrait jouer le rôle de cadre — est, lui, en mouvement perpétuel. On n'a donc plus un objet qui se déplace dans un cadre stable ; on a un objet figé dont le cadre s'éloigne.

Là où la science classique mesure le mouvement d'un objet dans un cadre fixe, la donnée donne à constater l'éloignement d'un cadre mobile autour d'un objet fixe.

Le vieillissement d'une donnée n'est pas son mouvement propre ; c'est l'effet du déplacement de son cadre. Cette inversion explique une difficulté que ressent souvent l'esprit formé à la méthode scientifique : il est habitué à ce que le référentiel serve à *capter* le mouvement, et il se trouve ici devant un cas où le mouvement du cadre lui-même est ce qu'il faut prendre en compte. L'inconfort est réel ; il n'est pas le signe d'une erreur, mais celui d'une situation authentiquement inhabituelle.

La péremption comme écartement

Réunissons les sections précédentes en une image unique.

La donnée a deux ancrages. L'un, posé par l'acte d'inscription, la rattache à un instant précis du flux du réel — l'instant t où la coupe a été pratiquée. L'autre, posé par la fixation sur un support, la rattache au temps de l'inscription, où elle demeure éternellement identique à elle-même.

Ces deux ancrages sont *immobiles*. La donnée ne bouge dans aucun des deux temps. Mais le flux du réel, lui, ne s'arrête pas : il s'écoule autour du point que l'acte a marqué, comme une rivière s'écoule autour d'un piquet planté dans son lit. Le piquet ne descend pas le courant ; le courant passe devant lui. À mesure que l'eau s'éloigne, l'écart se creuse entre l'instant où le piquet fut planté et l'état présent de la rivière.

Tant que cet écart reste petit — c'est-à-dire tant que l'instant de l'acte est récent —, la donnée représente fidèlement le réel. La donnée est *fraîche*. Mais le flux ne s'arrête jamais : l'écart grandit, non parce que la donnée bougerait, mais parce que le temps du réel emporte tout, sauf elle.

La péremption d'une donnée est exactement cet écartement. Ce n'est ni une dégradation de la donnée, ni une dégradation du support : la donnée reste lisible, le support tient bon. C'est l'écart grandissant entre le point que l'acte a marqué dans le flux et l'état présent de ce flux. La donnée ne se périme pas par l'intérieur ; elle se périme parce que le monde s'éloigne d'elle.

Cette image dit aussi pourquoi la péremption est **inéluçtable** mais à **vitesse variable**. Inéluçtable, parce que le flux du réel ne s'arrête jamais. À vitesse variable, parce que tout dépend de la rapidité du courant : un flux du réel qui change lentement — la composition d'un sol, la pyramide des âges d'une population — s'éloigne lentement de l'instant où l'acte fut posé ; un flux qui change vite — un cours boursier, un rythme cardiaque à l'effort — s'en éloigne en quelques instants. La vitesse de péremption n'est pas une propriété de la donnée : c'est la vitesse du fragment de réel dont elle fut extraite.

Partie III — Ce que la distinction éclaire

L'illusion de la répétabilité, ou les sources

Une objection naturelle peut surgir ici. Si l'acte d'inscription est unique et non répétable, comment se fait-il que, dans la pratique, on puisse prélever « deux fois la même donnée » ? Je compte les cernes d'un tronc d'arbre : j'obtiens 87. Je les recompte le lendemain : j'obtiens 87. Si je relève la température d'un corps deux fois à quelques secondes d'intervalle, l'instrument affiche deux fois 38,2 °C. La répétabilité semble être un fait d'expérience.

Cette objection mérite d'être affrontée, car sa réponse révèle quelque chose d'important.

D'abord, la rigueur exige de dire que ce ne sont pas « deux fois la même donnée » : ce sont **deux données distinctes qui portent la même valeur**. Chacune procède d'un acte d'inscription séparé, daté différemment, accompli par un geste différent. Ce qui se répète, c'est la valeur arrêtée, jamais l'acte. La distinction entre les deux temps autorise sans peine cette précision : deux actes peuvent coïncider dans le temps du réel sur la même valeur — parce que le flux du réel n'a pas eu le temps de bouger significativement entre eux.

Mais c'est précisément ce *parce que* qui doit être examiné, car il révèle la véritable nature de l'illusion. Quand je recompte les cernes du tronc,

j'obtiens la même valeur non parce que le flux du réel se serait arrêté, mais parce que sa vitesse de transformation, sur l'objet observé, est si lente à l'échelle humaine que l'écart entre les deux actes reste imperceptible. À l'échelle du tronc, un jour n'est rien. À l'échelle d'un siècle, le tronc se sera désagrégé et la « répétabilité » se sera évanouie. La péremption était bien à l'œuvre — simplement à un rythme que la perception humaine ne capte pas.

C'est la même structure d'illusion que celle qu'Aristote subissait, sans le savoir, sur la chute des corps. Aristote voyait toujours l'objet lourd tomber plus vite que l'objet léger, et il en concluait à une loi du réel. Mais ce qu'il voyait n'était pas la chute pure : c'était la chute *dans l'air*, et l'air biaisait toute son observation sans qu'il pût accéder à un environnement de comparaison. Il fallut attendre Galilée et le vide pour que l'illusion se dissipe. De même, l'humain qui observe la répétabilité d'un prélèvement ne perçoit pas une absence de péremption : il perçoit une péremption ralentie *au-delà de ce que sa fenêtre temporelle peut saisir*.

De cette analyse découle une notion utile : celle de **source stable**. Une source stable est un fragment du réel dont la péremption est si lente qu'elle peut être traitée, à l'échelle de l'usage humain, comme négligeable. Non parce qu'elle est absente — elle est inéluctable, le flux ne s'arrête jamais — mais parce qu'elle est imperceptible. La grammaire d'une langue, les théorèmes mathématiques, les grandes régularités physiques, les faits historiques bien établis, certaines propriétés géologiques : autant de sources dont les prélèvements peuvent être tenus pour valides longtemps, pourvu qu'on garde présent à l'esprit que cette validité est une *durée* et non une *éternité*.

Cette notion a une portée que l'essai doit nommer. Les sources stables sont, pour une société, ce que les terrains de fondation sont pour un bâtiment : suffisamment fermes, à l'échelle d'usage, pour qu'on puisse y bâtir des édifices durables — quand bien même on sait, en dernière analyse, que le sol lui-même finira par bouger. Une civilisation se construit en partie sur sa capacité à identifier ces sources, à les protéger, à les transmettre. D'autres sources, à péremption rapide, peuvent aussi être exploitées — une poche de pétrole, un effet de mode, un avantage concurrentiel — mais elles appellent une autre posture : savoir qu'elles s'épuiseront, prévoir leur relais, ne pas leur demander la stabilité qu'elles ne peuvent pas offrir. Ce qui est coûteux, ce n'est pas de prélever sur une source instable, c'est de la confondre avec une source stable et d'y bâtir comme si elle l'était.

La métaphore aquatique de l'essai s'enrichit ici naturellement. L'humain n'est pas propriétaire des eaux qui s'écoulent dans le fleuve. Il n'est pas non plus propriétaire des canaux qu'il a creusés dedans — au sens fort, ces canaux sont des artefacts qu'il maintient mais qui dépendent du fleuve. Mais les **sources** qui alimentent ses canaux — les fragments du réel dont la péremption ralentit jusqu'à devenir négligeable —, celles-là, il peut les reconnaître, les nommer, les transmettre. La source est ce qui apaise dans le réseau métaphorique : ni le torrent qui emporte, ni le canal qu'il faut

entretenir, mais le point d'origine identifié auquel on peut toujours revenir.

Cette idée appelle, au plan de la littérature, une compétence de la famille que le document *Creuser le canal* rassemble. Elle s'ajoute à la liste : savoir identifier les sources stables, et ne pas confondre la lenteur de leur péremption avec son absence.

Les deux morts de la donnée

La donnée peut **mourir de deux façons distinctes**, qu'il importe de ne pas confondre.

La première mort est la **péremption** : la donnée devient fausse pour décrire le présent. C'est la mort dans le temps du réel. La donnée demeure parfaitement lisible, parfaitement intacte sur son support — mais elle ne dit plus le réel actuel. Le relevé « 38,2 °C » d'il y a trois jours est toujours là, exact en tant que constat daté, mais il ne renseigne plus sur la fièvre présente. Cette mort-là n'atteint pas l'existence de la donnée ; elle atteint sa **validité représentative**.

La seconde mort est la **disparition** : le support se dégrade, et la donnée cesse purement et simplement d'exister. C'est la mort dans le temps de l'inscription. La roche s'érode, l'argile se brise, le support magnétique se démagnétise, le format devient illisible faute de machine pour le lire. La donnée n'est plus fausse : elle n'est plus *du tout*. Cette mort-là atteint l'**existence**.

Ces deux morts sont indépendantes. Une donnée peut être parfaitement vivante quant à son support et parfaitement morte quant au réel : c'est le cas de toute archive ancienne. Inversement, une donnée pourrait être encore fraîche quant au réel mais déjà perdue par défaillance du support.

Conserver une donnée et la maintenir valide sont deux tâches différentes, qui appellent des moyens différents. Lutter contre la disparition relève de la qualité et de la redondance des supports — c'est une affaire d'archivage. Lutter contre la péremption relève du re-prélèvement — c'est une affaire de fraîcheur, et aucun support, si durable soit-il, n'y peut quoi que ce soit. On peut graver une donnée dans le matériau le plus résistant : cela la protège de la disparition, cela ne la protège pas d'une seconde de péremption.

Une variante : la donnée inaccessible

Une troisième situation mérite d'être nommée, qui n'est ni la péremption ni la disparition au sens strict, mais une autre forme d'extinction. Imaginons une donnée fixée sur une onde électromagnétique qui s'est propagée dans l'espace au-delà du champ de prélèvement humain. L'inscription est toujours là : l'onde continue de voyager, transportant le signal qu'on lui a confié. Son support n'a pas disparu — il est même, à l'échelle humaine, d'une durabilité quasi infinie. Et la donnée n'est pas

périmée au sens du temps du réel : ce qu'elle dit reste exact relativement à l'instant où l'acte d'inscription a été pratiqué.

Mais elle est devenue **inaccessible**. Le lecteur n'a aucun moyen, depuis sa position, de la rejoindre — l'onde s'éloigne plus vite qu'il ne saurait la rattraper, ou s'est dispersée hors de portée des instruments. Pour lui, tout se passe comme si la donnée n'existait plus, alors même que ni le support ni la fraîcheur ne sont en cause.

Cette troisième forme d'extinction — l'inaccessibilité — ne contredit pas les deux morts précédemment posées ; elle en éclaire la portée. Une donnée peut être parfaitement vivante en elle-même, et néanmoins perdue *pour un lecteur situé*. C'est, en miroir, une conséquence du fait établi au document suivant : tout lecteur est partiel, borné par ce qu'il peut atteindre. La donnée, comme la trace, n'existe pleinement que dans la rencontre avec un lecteur capable de la lire — et cette rencontre, parfois, est rendue impossible par la seule géométrie de l'espace.

Le paradoxe de la duplication

La donnée peut être **dupliquée**, et la copie est identique à l'original — fait qui n'a aucun équivalent dans le monde des objets matériels. Si l'on duplique une pomme, on obtient deux pommes — deux objets distincts. Mais si l'on duplique la donnée « 38,2 °C relevé à l'instant t » d'un support vers un autre, on n'obtient pas « deux données » au sens où l'on aurait deux pommes. On obtient **la même donnée, portée par deux supports**.

La distinction des deux temps explique pourquoi. La duplication est une opération qui porte sur la **fixation** — elle copie l'inscription d'un support vers un autre. Elle ne touche pas à l'**acte**. Or l'acte est unique, daté, non répétable. Dupliquer le support ne crée pas un second acte ; cela ne fait que multiplier les porteurs d'un acte unique.

On en tire une conclusion : **la donnée n'est pas dans le support, elle est portée par lui**. Le support est multiple, matériel, périssable, duplicable. Ce qu'il porte — le constat daté issu de l'acte — est unique, invariant, indifférent au nombre de supports qui le portent. La donnée est donc une entité d'un genre singulier : elle a besoin d'un support matériel pour exister et durer, mais elle ne s'identifie à aucun support particulier. Elle est, en un sens précis, *l'acte rendu persistant* — et un acte ne se duplique pas, il se conserve.

Cette analyse a une conséquence pour les questions de propriété et de circulation déjà abordées. Quand une donnée est transmise d'une entité à une autre, l'entité d'origine ne s'en dépossède pas : elle en garde une copie, et les deux copies sont la même donnée. Ce qui circule, ce n'est pas un objet qui passerait de main en main ; c'est un constat qui se rend présent en plusieurs lieux à la fois. La donnée ne se cède pas comme une chose ; elle se *partage* comme un fait.

La règle peut dès lors se fixer en quatre verbes, que toute la suite de l'essai observera. L'**acte** se conserve : unique et daté, il ne se duplique ni ne se multiplie. L'**inscription** se duplique : la duplication est l'opération élémentaire, qui copie la fixation d'un support vers un autre. Les **porteurs** se multiplient : la multiplication est l'effet agrégé des duplications, comme la propagation sera l'agrégat des transports. La **donnée**, elle, se partage : une et une seule, présente en plusieurs lieux, sans que nul ne s'en dépossède.

Pourquoi la fraîcheur ne se rattrape jamais

La distinction des deux temps établit enfin avec rigueur un principe énoncé de façon intuitive : **la fraîcheur ne peut pas être obtenue après coup.**

La fraîcheur est la proximité entre l'instant de l'acte et le présent du réel. Or l'acte est unique et non répétable : il a eu lieu à l'instant *t*, il n'aura jamais lieu à un autre instant. Pour disposer d'une donnée fraîche, il n'y a qu'un seul moyen : accomplir un **nouvel acte d'inscription**, maintenant, sur le réel présent. Ce nouvel acte produit une nouvelle donnée — datée de maintenant, fraîche par construction. Mais il ne *rafraîchit* pas l'ancienne : il en crée une autre, à côté.

La fraîcheur, autrement dit, n'est pas un attribut que l'on pourrait restaurer ; c'est la signature d'un acte récent. On ne peut pas rendre récent un acte ancien. On ne peut que poser un acte nouveau.

Ce principe a une conséquence importante. Toutes les techniques qui semblent « produire de la fraîcheur » sans nouveau prélèvement — la régression, la prévision des séries temporelles, l'analyse des tendances et de la saisonnalité — ne produisent jamais, en réalité, de la fraîcheur. Elles produisent autre chose, qui demande un nom précis — et c'est l'objet de la partie suivante.

Partie IV — Fraîcheur, vraisemblance et prédiction

Fraîcheur et vraisemblance : deux choses à ne pas confondre

Il existe un vaste ensemble de techniques — régression linéaire, prévision des séries temporelles, contrôle statistique des procédés, analyse de la saisonnalité, étude des tendances — qui permettent, à partir d'un ensemble de données anciennes, d'estimer ce que vaudrait une donnée si on la prélevait maintenant.

Ces techniques sont précieuses, légitimes, souvent remarquablement efficaces. Mais soyons d'une grande rigueur sur ce qu'elles produisent. Elles ne produisent **pas de la fraîcheur**. Elles produisent ce qui doit se nommer d'un autre mot : de la **vraisemblance**.

La fraîcheur est le fruit d'un **contact réel** avec le présent du flux du réel. Elle s'obtient par un acte d'inscription récent — un nouveau prélèvement. Une donnée fraîche est garantie par le réel lui-même : elle dit ce que le réel est, parce qu'elle vient d'en être extraite.

La vraisemblance est le fruit d'un **calcul sur un modèle**. Elle est une donnée consolidée. Elle s'obtient sans contact avec le présent du réel, à partir d'un ensemble de données anciennes et d'hypothèses sur la manière dont le réel se comporte. Une donnée vraisemblable n'est pas garantie par le réel : elle est garantie par un modèle, c'est-à-dire par l'hypothèse que le réel a continué de se comporter comme le modèle le suppose.

Une donnée vraisemblable est donc une **donnée d'une autre nature** : non plus un prélèvement sur le flux du réel, mais un prélèvement sur un modèle du flux du réel. Elle n'a pas touché le réel présent ; elle a touché une représentation calculée de ce que le réel présent *devrait* être. Un calcul effectué à partir de flux d'observation construits à partir du flux du réel.

Cela ne disqualifie nullement la vraisemblance. Dans une immense quantité de situations, un nouveau prélèvement est impossible, trop coûteux, trop lent, et la vraisemblance est alors le meilleur substitut disponible. Mais l'honnêteté conceptuelle — et la littératie de la donnée — exigent de ne jamais confondre les deux. Présenter une donnée vraisemblable comme une donnée fraîche, c'est attribuer à un calcul la garantie qui n'appartient qu'au contact avec le réel. C'est faire passer une hypothèse pour un constat.

Le risque, ou le prix de la vraisemblance

Puisque la donnée vraisemblable repose sur l'hypothèse que le réel a continué de se comporter comme le modèle le suppose, elle porte en elle un **risque** : celui que cette hypothèse soit fausse — que le réel, justement, ait cessé de ressembler au modèle.

Ce risque ne peut pas être supprimé. Il peut seulement être **maîtrisé** : estimé, borné, et mis en regard de ce que l'usage tolère.

La marge d'incertitude acceptable n'est pas la même selon les domaines, parce que le **coût d'une erreur** n'est pas le même. Deux illustrations le montrent.

Première illustration — la gestion d'un rayon de commerce. Prévoir le réassortiment d'un rayon de produits secs et celui d'un rayon de produits frais sont deux exercices de vraisemblance de même nature, mais qui ne tolèrent pas la même marge d'incertitude. Le produit sec se conserve longtemps : si la prévision est trop large, l'excédent restera vendable, et la commande suivante pourra être ajustée. L'erreur est *rattrapable*, donc une marge d'incertitude ample est acceptable. Le produit frais se conserve peu : un excédent devra être jeté, une erreur ne se rattrape pas.

Même technique, même nature de vraisemblance — mais une exigence de précision dictée par la flexibilité de l'usage.

Seconde illustration — le calcul d'une trajectoire. Tracer la courbe d'un virage routier et calculer la trajectoire d'une sonde spatiale sont deux exercices de prévision de même nature, aux exigences incomparables. Pour le virage, une valeur approchée du nombre π à deux décimales suffit : le véhicule dispose d'une marge — il peut ralentir, corriger, la trajectoire est flexible. Pour la sonde qui doit exploiter une assistance gravitationnelle, il faut une valeur de π à de nombreuses décimales : la trajectoire n'a aucune flexibilité, un écart minime au départ devient un écart immense à l'arrivée.

Ces illustrations disent une chose de fond : **la qualité requise d'une donnée vraisemblable ne se définit pas dans l'absolu, mais relativement à l'usage et au risque que cet usage peut tolérer.** C'est une déclinaison directe de l'optimum observationnel qu'établit le document *Creuser le canal* : assez précis pour que le risque soit acceptable, pas plus précis que ce que l'usage exige.

Ce que l'on prédit n'est pas la donnée

Reste à éclairer un dernier point, le plus subtil, et peut-être le plus important pour la justesse conceptuelle de toute la série.

Quand on emploie une technique de prévision, le langage courant dit qu'on « prédit une donnée ». Cette expression est trompeuse, et la distinction des deux temps permet de dire pourquoi.

Une donnée est un prélèvement figé — un constat daté, arrêté par un acte, fixé sur un support. En tant que telle, elle n'a pas d'avenir : elle ne va nulle part, elle ne deviendra rien d'autre que ce qu'elle est. Une donnée passée restera éternellement cette donnée passée. **On ne peut donc pas prédire une donnée** : il n'y a rien, dans une donnée, qui soit susceptible d'être prédit, puisqu'elle ne se transformera jamais.

Ce que l'on prédit, en réalité, ce n'est pas la donnée : c'est **l'état futur du fragment du flux du réel** que les critères de prélèvement avaient découpé. On ne prédit pas un nombre ; on anticipe un état du monde. Et le résultat de cette anticipation — la valeur calculée — n'est pas une donnée fraîche : c'est une donnée vraisemblable, un prélèvement opéré non sur le réel mais sur un modèle.

Posons ici un garde-fou. Ce que l'on anticipe n'est jamais « le flux du réel » en général. Le flux du réel dans son ensemble est immense, polymorphe, et reste hors de toute prise. Ce que l'on anticipe, c'est uniquement **l'état futur du fragment de réel que les critères du flux d'observation avaient cadré**. La prévision hérite intégralement du cadrage du prélèvement. Si les critères ne captaient pas un certain paramètre, la prévision est aveugle à ce paramètre — et donc aveugle à tout événement futur qui en dépendrait.

On ne prédit pas une donnée — un prélèvement figé n'a pas d'avenir. On anticipe l'état futur probable du fragment de réel que les critères de prélèvement avaient découpé. Le résultat de cette anticipation n'est pas une donnée fraîche, mais une donnée vraisemblable : un prélèvement opéré sur un modèle, et non sur le réel. La fraîcheur ne s'obtient que par un nouveau contact avec le réel ; tout le reste est inférence, et l'inférence se paie en risque.

Conclusion — Le point fixe et le courant

Au terme de ce parcours, le paradoxe du préambule est dissous. La donnée est figée *et* périssable, sans contradiction, parce qu'elle vit dans deux temps à la fois.

Elle est figée dans le **temps de l'inscription**, parce qu'un support la maintient identique à elle-même : « 38,2 °C relevé à l'instant *t* » restera vrai à jamais, constat daté soustrait à toute révision. C'est le socle de toute mémoire.

Elle est périssable dans le **temps du réel**, parce que le flux dont l'acte l'avait extraite ne s'arrête jamais : le réel continue, s'éloigne du point que l'acte avait marqué, et l'écart grandissant entre ce point fixe et ce courant est ce que nous nommons péremption. La donnée ne bouge pas ; c'est le monde qui la laisse derrière lui.

Cette double vie tient à la fabrique même de la donnée. Elle naît de deux opérations : un **acte d'inscription**, qui pratique une coupe dans le continu du réel et la date irrévocablement ; une **fixation sur un support**, qui arrache la valeur arrêtée à l'écoulement et l'installe dans l'éternité de l'inscrit. L'acte ancre la donnée à un instant du flux ; le support l'ancre dans le temps de l'inscription. La donnée est la tension entre ces deux ancrages — et son vieillissement, l'éloignement inéluctable du premier, à mesure que le flux qui le portait continue de couler.

De cette distinction découlent plusieurs clarifications. La donnée peut mourir de deux façons — périmer dans le temps du réel, disparaître dans le temps de l'inscription — et lutter contre l'une ne protège jamais de l'autre. La donnée peut être dupliquée sans être multipliée, parce qu'elle n'est pas dans son support mais portée par lui. La fraîcheur ne se rattrape jamais, parce qu'elle est la signature d'un acte récent. Et ce que l'on appelle « prédire une donnée » est en réalité anticiper l'état futur d'un fragment de réel cadré par des critères : le résultat n'est pas une donnée fraîche mais une donnée vraisemblable, garantie par un modèle et non par le monde.

De cette double leçon — la péremption qui ralentit jusqu'à paraître stable, la transmission qui s'accélère jusqu'à paraître instantanée — se dégage

une vérité plus large encore. **La péremption existe à toutes les échelles du flux du réel. L'humain n'en perçoit qu'une bande — celle où le rythme du réel est commensurable au sien. En deçà, le réel paraît immuable ; au-delà, il paraît instantané. C'est entre ces deux seuils qu'il pose ses canaux, plante ses sources, construit sa civilisation.**

La leçon, au fond, est simple. Une donnée est un point fixe planté dans un courant. Sa force est d'être un point fixe : c'est ce qui nous permet de bâtir sur elle. Sa limite est d'être plantée dans un courant : c'est ce qui fait qu'elle nous échappe à mesure que le courant avance. Comprendre la donnée, c'est tenir ces deux vérités ensemble — et savoir, à chaque fois qu'on s'appuie sur une donnée, depuis combien de temps le courant a quitté le point où elle fut plantée.

« *Le temps est invention ou il n'est rien du tout.* » La donnée, elle, est inscrite dans deux temps à la fois : celui qui la garde intacte, et celui qui, sans jamais la toucher, l'éloigne du monde.

Ces deux temps de la donnée sont, dans le fond, aussi deux temps du lecteur. Quand quelqu'un rencontre une trace — une inscription, une donnée, un document — il rencontre un acte accompli par quelqu'un d'autre, dans un autre présent, pour des raisons qu'il peut ne plus connaître. L'écart entre l'instant de l'acte et l'instant de la lecture, c'est ce que le document suivant explore : ce que le lecteur peut faire d'une trace dont il n'a pas été le témoin.

Repères

Pour les références mobilisées dans ce document — Bergson (durée pure / temps spatialisé), l'ontologie processuelle, Gitelman, et les artefacts archéologiques évoqués (os entaillé sur plusieurs dizaines de millénaires) —, voir la *Bibliographie raisonnée* en fin d'essai. Le statut de méthode (fond/forme) est posé dans la *Préface méthodologique*.

La trace et le lecteur

Ce qui rend la donnée possible — Aux fondations de l'essai

« *La nature ne fait rien en vain, et rien de superflu.* »

— Aristote, *De Caelo*

« *Le monde est fait pour aboutir à un livre.* »

— Stéphane Mallarmé

Préambule — Une question jamais posée

Ce document occupe, dans l'essai, une position singulière. Il a été rédigé en suivant ; il pourrait pourtant être positionné en premier — Les documents qui le précèdent ont tous présupposé, sans jamais l'interroger, qu'un prélèvement était possible. Ils ont décrit comment l'humain prélève, annote, construit des flux d'observation, mais ils n'ont pas demandé : *qu'est-ce qui, dans le monde et dans l'humain, rend ce prélèvement possible en premier lieu ?*

Ce document pose cette question. Il est aux autres ce que les fondations sont à un bâtiment : invisibles une fois l'édifice dressé, mais ce sans quoi l'édifice ne tiendrait pas. Le lecteur qui aborde l'essai par son commencement chronologique peut donc le découvrir là où il se présente, sans dommage ; celui qui voudra ensuite comprendre pourquoi tout le reste tient gagnera à y revenir, et à le relire comme on relit une introduction quand on a fini un livre.

Tout l'essai repose sur une thèse : la donnée n'est pas un objet préexistant que l'on collecterait, mais le résultat d'un prélèvement opéré sur le flux du réel par un acte d'annotation. Cette thèse, si on l'examine, contient une supposition qu'elle n'a jamais justifiée. Elle suppose qu'il y a, dans le réel, quelque chose qui *peut* être prélevé ; et elle suppose qu'il existe un

être qui *peut* prélever. Ces deux suppositions ont été traitées comme des évidences. Elles n'en sont pas.

Si le flux du réel était un pur écoulement sans épaisseur — un présent perpétuel qui ne garderait aucune marque de ce qu'il a été —, que pourrait-on en prélever ? Rien, ou presque : un instantané sans profondeur, immédiatement dissous. Pour qu'un prélèvement ait un sens, il faut que le réel *porte quelque chose* — qu'il soit, d'une manière ou d'une autre, déjà marqué, déjà structuré par son propre passé. Il faut que le réel soit *lisible*.

Et si rien, dans l'univers, n'était capable de viser, de découper, d'annoter, d'inscrire — s'il n'y avait que de la matière et des forces —, qui prélèverait ? Personne. Le réel pourrait être aussi lisible qu'on voudra : sans lecteur, sa lisibilité resterait lettre morte. Il faut qu'il existe un être capable de lire.

La donnée, donc, ne tient pas par elle-même. Elle tient sur deux piliers, et ce document a pour seul objet de les mettre au jour :

- **Premier pilier — la lisibilité du réel.** Le flux du réel garde trace de lui-même. En s'écoulant, il laisse des marques. Cette propriété est ce qui rend le réel *prélevable*.
- **Second pilier — l'existence d'un lecteur.** Le flux du réel a produit, par son propre écoulement, un être capable de lire ces marques. Cet être est ce qui rend la lisibilité *effective*.

La donnée naît à la rencontre exacte de ces deux piliers : un réel qui porte des traces, et un être que le réel a engendré pour les lire. Tout l'essai repose là. Ce document descend jusqu'à ce socle.

Partie I — La lisibilité du réel

Le réel garde trace de lui-même

Commençons par le premier pilier, et par un fait que l'on ne remarque pas, justement parce qu'il est partout.

Le flux du réel, en s'écoulant, ne s'efface pas entièrement derrière lui. Il laisse des **marques**. Le passage de l'eau creuse la pierre, et la pierre garde la forme du passage. La croissance de l'arbre dépose des couches de bois, et le tronc garde l'empilement des couches. La coulée de lave refroidit, et la roche garde la mémoire de la coulée. Le dépôt des sédiments forme des strates, et la falaise garde l'ordre des dépôts. Le pas dans la boue laisse une empreinte, et la boue séchée garde l'empreinte.

Ce fait est si général qu'il faut un effort pour le voir comme un fait, et non comme une nécessité allant de soi. On peut imaginer — par l'esprit — un monde où il n'en serait pas ainsi : un monde où chaque instant

s'effacerait totalement, où la matière ne conserverait aucune empreinte de ce qu'elle a subi. Un tel monde serait possible à concevoir. Il ne serait pas connaissable.

Notre monde n'est pas ainsi. Notre monde est un monde où **le passé laisse des marques dans le présent matériel**. Nous nommerons cette propriété la *lisibilité du réel*. L'état présent de la matière n'est pas un pur instantané sans profondeur ; il est, partout, chargé de la mémoire de ce qui l'a produit. La pierre érodée n'est pas seulement une pierre : elle est une pierre *qui porte l'inscription de la pluie*. Le tronc coupé n'est pas seulement du bois : il est du bois *qui porte l'inscription des saisons*.

C'est cette propriété — et elle seule — qui rend le réel prélevable. On ne prélève pas un instant pur ; on ne pourrait rien en faire. On prélève un état du réel *en tant qu'il porte les marques de son passé*. Le prélèvement a un sens parce que le réel a une épaisseur.

Qu'est-ce qu'une trace ?

Le mot juste pour ces marques est **trace**. Trois caractères la définissent.

La trace est causée. Elle n'est pas faite *exprès*. Aucune intention ne préside à sa formation. La pluie n'érode pas la pierre pour laisser un message ; elle l'érode parce que telle est la conséquence physique du ruissellement. L'arbre ne forme pas ses cernes pour qu'on les compte ; il les forme parce que la croissance alterne avec les saisons. La trace est le produit aveugle d'un processus causal. C'est son premier caractère, et le plus important : la trace n'est voulue par personne.

La trace est physiquement liée à ce qu'elle marque. Il y a, entre la trace et ce qui l'a produite, un lien réel, matériel, non arbitraire. L'empreinte a *réellement* été creusée par le pied ; le cerne a *réellement* été déposé par une année de croissance. C'est ce qui distingue la trace du signe arbitraire : le mot « arbre » n'a aucun lien physique avec un arbre, c'est une convention de langue ; le cerne, lui, est physiquement le produit de l'année qu'il signale.

La trace est lisible, mais elle n'est pas lue. La trace contient, en un sens, une information — l'âge de l'arbre est *dans* la disposition des cernes. Mais cette information n'est que *potentielle*. Elle est disponible ; elle n'est pas actualisée. Tant que personne ne vient lire les cernes, l'arbre porte son âge sans que cet âge soit *su* de quiconque. La trace offre une lisibilité ; elle n'accomplit pas une lecture.

Ce troisième caractère est décisif pour la suite : **une trace est du réel lisible, ce n'est pas encore une donnée.**

La trace, condition de toute connaissance

La lisibilité du réel n'est pas un fait parmi d'autres : c'est la condition de possibilité de **toute connaissance**, sans exception.

Tout savoir qui porte sur le passé repose sur des traces. La géologie lit les strates pour reconstituer l'histoire de la Terre. La dendrochronologie lit les cernes pour dater le bois. La paléontologie lit les fossiles. L'astronomie lit la lumière des étoiles lointaines, qui est une trace d'un état du cosmos vieux de millions ou de milliards d'années. L'archéologie lit les vestiges. La médecine lit les symptômes, qui sont les traces de la maladie dans le corps. La police scientifique lit les indices.

Aucune de ces connaissances ne serait possible dans un monde sans traces. Et le constat porte plus loin : même la connaissance du présent repose sur des traces, car toute perception est déjà la lecture d'une trace. Quand l'œil voit un objet, il ne touche pas l'objet : il reçoit la lumière que l'objet a réfléchi — une trace lumineuse, déjà séparée de sa source par le temps, infime, du trajet. Percevoir, c'est toujours lire une marque que le réel a laissée.

La lisibilité du réel est donc le sol de tout. Sans elle, pas de science, pas de mémoire, pas de perception, pas de donnée. C'est le premier pilier, et il est si fondamental qu'on ne le voit jamais — comme on ne voit pas le sol sur lequel on marche.

Les temps de la trace : au-delà de l'instantané et de l'éternel

Une dernière remarque sur ce premier pilier, qui prépare la Partie III.

Les traces que le réel laisse de lui-même se déposent à toutes les échelles de temps. Certaines se forment et s'effacent en une fraction de seconde — l'onde à la surface de l'eau, l'écho d'un son. D'autres mettent des saisons — les cernes d'un arbre. D'autres des siècles — l'érosion d'une falaise. D'autres des millions d'années — les strates géologiques. D'autres encore des milliards d'années — la lumière fossile du cosmos.

L'humain ne perçoit directement qu'une **fenêtre** de ces échelles. Ce qui se forme et s'efface trop vite lui échappe ; ce qui se forme trop lentement aussi. L'humain a un *seuil de l'instantané* et un *seuil de l'éternel*. Ces deux seuils ne sont pas des propriétés du réel. Ce sont des propriétés de l'humain. De même que l'humain ne perçoit qu'une bande du spectre lumineux et qu'une bande du spectre sonore, il ne perçoit qu'une bande du *spectre des temps*.

Ce point sera essentiel quand nous parlerons du second pilier. Il signifie que **le lecteur du réel est un lecteur situé** : il ne lit pas toutes les traces, il lit celles qui tombent dans sa fenêtre. La lisibilité du réel est totale — et il faut entendre l'énoncé en son statut exact : non l'inventaire, impossible à clore, de tout ce que le réel donnerait à lire, mais le constat qu'aucune connaissance n'a jamais eu d'autre voie : rien de ce que nous savons du monde ne l'a été autrement que par une trace. La lecture qu'un être donné peut en faire est, elle, toujours partielle, bornée par ce que cet être est.

Partie II — La trace n'est pas la donnée

Le seuil que rien ne franchit tout seul

Nous arrivons au point le plus délicat de ce document, et le plus important pour la solidité de tout l'essai.

Il est tentant — et l'intuition est belle — de dire que les traces *sont* des données. Les cernes de l'arbre ne sont-ils pas l'inscription des années ? Les strates ne sont-elles pas l'annotation des dépôts ? La pierre volcanique ne porte-t-elle pas, gravée en elle, l'information de l'éruption ? Le tronc coupé ne réunit-il pas, semble-t-il, les deux opérations que *Les deux temps de la donnée* a identifiées comme constitutives de la donnée : un acte d'inscription (une paire d'anneaux par année) et une fixation sur un support (le bois lui-même) ?

Cette tentation doit être nommée et écartée. Si les traces étaient des données, alors la donnée préexisterait à l'humain, elle traînerait partout dans le réel, et l'on retomberait exactement dans l'ontologie objectale que les cinq documents se sont employés à défaire. La phrase « nous vivons entourés de données » redeviendrait vraie — et avec elle, l'idée que la donnée est un objet que l'on ramasse, et non un prélèvement que l'on opère.

La distinction est donc à tenir fermement. **La trace n'est pas la donnée.** Entre elles, il y a un seuil, et ce seuil, rien ne le franchit tout seul. Il faut un acte.

Causé contre voulu

La différence entre la trace et la donnée tient en deux mots : la trace est **causée**, la donnée est **voulue**.

L'acte d'inscription, tel que *Les deux temps de la donnée* l'a établi, n'est pas n'importe quelle marque laissée dans la matière. C'est un acte qui **décide** : il arrête une valeur, il pratique une coupe intentionnelle dans le continu, il vise.

Or l'arbre ne décide rien. L'arbre ne vise rien. L'arbre ne pratique aucune coupe intentionnelle dans le réel : il *subit* le réel, et en le subissant, il en garde la marque. La formation des cernes est un processus causal aveugle. Il n'y a, dans l'arbre, aucune intention de consigner les années, aucun projet de mémoire, aucun destinataire visé. L'arbre ne s'inscrit pas lui-même : il est inscrit, passivement, par l'alternance des saisons.

C'est pourquoi les cernes, malgré les apparences, ne réunissent pas les deux opérations constitutives de la donnée. Ils ont bien quelque chose qui *ressemble* à une fixation sur un support — le bois retient les couches. Mais ils n'ont pas d'acte d'inscription, parce qu'il n'y a pas eu d'acte du tout : il y a eu un processus. Et un processus n'est pas un acte. Un processus

se déroule ; un acte se décide. La trace est le résidu d'un processus ; la donnée est le produit d'un acte.

La trace, donc, possède la *matérialité* de la donnée — elle est dans la matière, elle dure — mais elle n'en possède pas l'*origine*. Il lui manque ce qui fait qu'une donnée est une donnée : le geste intentionnel qui vise, découpe et inscrit.

Les trois étages : flux, trace, donnée

Posons maintenant clairement l'architecture complète. Entre le flux du réel et la donnée, il n'y a pas deux étages, il y en a trois.

Premier étage — le flux du réel. L'arbre qui pousse, la pierre que la pluie frappe, les sédiments qui se déposent. Pur écoulement. Aucune trace encore distinguée, aucune donnée. Le réel en train de se faire.

Deuxième étage — la trace. Les cernes formés, la pierre érodée, les strates empilées. Le flux du réel a laissé sur la matière la marque de son passage. Le réel est désormais *lisible* — il porte une information potentielle. Mais cette information n'est lue par personne. Il n'y a toujours pas de donnée : il y a du réel devenu lisible, et rien de plus. La trace est un entre-deux : plus que le pur flux, moins que la donnée.

Troisième étage — la donnée. Un observateur survient. Il regarde le tronc coupé. Il applique un critère — « une paire d'anneaux, un clair et un foncé, vaut une année ». Il compte. Il inscrit un résultat : « cet arbre a quatre-vingt-sept ans ». Là, et seulement là, une donnée existe. Et elle n'était nulle part dans l'arbre avant que l'observateur ne la produise.

Ce dernier point doit être pesé. L'arbre ne « contenait » pas son âge. L'arbre contenait des cernes — une trace. L'âge, lui, est le produit d'un acte : l'acte de quelqu'un qui a su que les cernes étaient lisibles, qui a choisi de les lire, qui a appliqué une règle de lecture, et qui a arrêté un résultat. La donnée « quatre-vingt-sept ans » appartient à l'observateur et à son acte ; elle n'appartient pas à l'arbre. L'arbre a fourni la trace. L'observateur a produit la donnée.

Le saut entre le deuxième et le troisième étage — entre la trace et la donnée — est le saut que rien ne franchit tout seul. C'est le saut de la lecture. La trace est *lisible* ; la donnée est *lue*. Et entre lisible et lu, il y a tout l'écart qui sépare une possibilité de son actualisation — l'écart qu'un acte, et seulement un acte, peut combler.

Le « livre de la nature » : une métaphore convoquée puis dépassée

L'intuition que le réel est, en quelque sorte, déjà écrit — qu'il est un texte porteur de sens — est ancienne et vénérable. La pensée médiévale concevait le monde comme un second livre, à lire à côté de l'Écriture.

Galilée, au seuil de la science moderne, écrivait que le grand livre de l'univers est écrit en langage mathématique. L'idée traverse les siècles : le réel serait lisible parce qu'il serait, déjà, une forme d'écriture.

Cette métaphore a une part de vérité — c'est tout le premier pilier : oui, le réel est lisible. Mais prise à la lettre, elle induit une erreur. Un livre est *écrit par quelqu'un, pour être lu, avec une intention*. Dire que la nature est un livre, c'est suggérer qu'elle a été inscrite intentionnellement, qu'elle est un message, qu'elle contient des données qui attendent leur lecteur. Or ce n'est pas le cas. Le réel n'est pas écrit : il s'écoule, et son écoulement laisse des traces. Les traces ne sont pas des phrases ; ce sont des résidus causaux.

La formulation juste n'est donc pas « le réel est un livre que l'on lit », mais : **le réel n'est pas un livre — il est ce à partir de quoi un lecteur peut faire des livres**. La nature ne contient pas de données ; elle contient des traces, et les données sont ce que le lecteur produit en lisant les traces. La métaphore du livre confond la lisibilité — qui est dans le réel — avec la lecture — qui est dans le lecteur. C'est cette confusion que tout l'essai a pour fonction de dissiper.

Partie III — Le lecteur que le flux a produit

Sans lecteur, la lisibilité reste lettre morte

Nous avons établi le premier pilier — la lisibilité du réel — et sa limite — la trace n'est pas la donnée, parce qu'il y faut un acte de lecture. Cette limite désigne, en creux, le second pilier. Il faut un **lecteur**.

Imaginons un monde parfaitement lisible — un monde riche en traces, où chaque chose porterait gravée la mémoire de tout son passé — mais où n'existerait aucun être capable de lire. Un monde de pierres marquées, de troncs annelés, de strates ordonnées, et personne. Que vaudrait la lisibilité de ce monde ? Rien. Elle resterait une possibilité pure, jamais actualisée. Les traces seraient là, parfaites, exhaustives — et muettes.

La donnée exige donc, en plus d'un réel qui porte des traces, l'existence d'un être capable de les lire. Cet être doit posséder des facultés précises : **percevoir** les traces, les **distinguer**, les **interpréter**, **viser** une intention, et **inscrire** un résultat. Ces facultés réunies, c'est ce que nous appelons un lecteur.

La question devient alors : d'où vient ce lecteur ?

Le lecteur, produit du flux

Le lecteur n'est pas venu d'ailleurs. Il n'a pas été déposé dans le monde depuis l'extérieur du monde. Le lecteur — l'humain — est lui-même un

produit du flux du réel.

L'humain est un être vivant. Le vivant est apparu au sein du flux du réel, par un processus qui appartient entièrement à ce flux : l'évolution. L'évolution n'est rien d'autre qu'un long écoulement du flux du réel — des générations, des variations, des sélections, étalées sur des milliards d'années. Et ce processus a façonné des êtres dotés de sens : des yeux qui captent la lumière, des oreilles qui captent les vibrations, des systèmes nerveux qui distinguent, comparent, mémorisent. Puis, chez l'une de ces lignées, une cognition capable d'abstraction, d'intention, de langage — capable de viser, de découper, d'inscrire.

Les facultés mêmes qui font de l'humain un lecteur ont été produites par le flux du réel. L'œil qui lit la trace est le résultat d'un long travail du flux. La cognition qui interprète la trace est le résultat d'un long travail du flux. L'humain n'observe pas le flux du réel depuis une rive extérieure ; il est une portion du flux, que le flux a façonnée de telle sorte qu'elle est devenue capable de lire les autres portions.

Pourquoi le flux a-t-il produit un tel être ? Précisément à cause du premier pilier. Parce que le réel est lisible — parce qu'il porte des traces —, il était *avantageux*, pour un être vivant, de savoir les lire. L'animal qui sait lire la trace du prédateur survit ; celui qui sait lire la trace de la proie se nourrit ; celui qui sait lire la trace de la saison anticipe. La lisibilité du réel a, en quelque sorte, *appelé* l'apparition de lecteurs. Le premier pilier a rendu le second non seulement possible, mais favorisé.

Le réel se retourne sur lui-même

De ce qui précède découle une conséquence d'une grande portée, qu'il faut énoncer avec netteté tout en restant dans les limites du fond.

Si le lecteur est issu du flux, et s'il lit les traces que le flux laisse de lui-même, alors quelque chose de remarquable se produit : **le flux du réel, à travers le lecteur qu'il a produit, se lit lui-même.**

Avant l'apparition de lecteurs, le flux du réel s'écoulait et laissait des traces, mais ces traces n'étaient lues par personne — le réel se déroulait sans se connaître. Avec l'apparition d'un être capable de lire, une portion du flux est devenue capable de prendre connaissance des autres portions, et, à terme, d'elle-même. Le réel a produit, en l'humain, un point où il se retourne sur son propre écoulement et le déchiffre.

La donnée est l'unité élémentaire de ce retournement. Chaque fois qu'un humain prélève une donnée, le flux du réel, par l'intermédiaire de cet être qu'il a façonné, prend acte d'un fragment de lui-même. La connaissance, la science, la mémoire, l'essai que vous lisez en ce moment — tout cela est le flux du réel en train de se lire à travers l'un de ses produits.

Ce document s'en tiendra là pour le fond. La portée de ce retournement — ce qu'il signifie pour la condition de l'humain — déborde le cadre d'un

document de fondation. Elle sera reprise dans le document qui clôt l'essai. Ici, il suffit d'avoir établi le fait, et lui seul : le second pilier n'est pas extérieur au premier. Le lecteur n'est pas un visiteur venu du dehors lire un monde étranger. Le lecteur est le réel devenu, en un point, capable de se lire.

Un lecteur situé, donc une lecture partielle

Puisque le lecteur est façonné par l'évolution dans des conditions particulières, il est un **lecteur situé**. Ses facultés ne sont pas illimitées : il perçoit une certaine bande du spectre lumineux, du spectre sonore, du spectre des temps — et pas le reste.

La lecture du réel par l'humain est donc, par construction, **partielle**. Non par défaut de soin, non par insuffisance de méthode, mais par nature : un lecteur situé ne lit que les traces qui tombent dans sa fenêtre. Les traces trop rapides, trop lentes, trop ténues, d'une nature que ses sens ne captent pas, restent illisibles *pour lui* — quoique parfaitement réelles.

Ce fait fonde, à la racine, tout ce que l'essai a dit de la partialité de la donnée. Si tout flux d'observation est partiel, si tout prélèvement laisse un non-prélevé, si aucune donnée n'épuise le réel — c'est, en dernière analyse, parce que **le lecteur lui-même est partiel**. La partialité de la donnée n'est pas un accident corrigible ; elle est l'héritage de la condition du lecteur. L'humain ne lit pas tout le réel parce que l'humain n'est qu'une portion du réel, et qu'une portion ne peut pas tout embrasser.

Les instruments — la lunette, le microscope, le capteur, le thermomètre — étendent la fenêtre du lecteur. Mais ils ne suppriment pas la condition de lecteur situé : ils la déplacent. Tout instrument a, à son tour, sa propre fenêtre, ses propres limites. L'extension de la lecture est sans fin ; son achèvement est impossible.

Partie IV — Ce que cette fondation légitime

Les deux piliers sous chaque document de l'essai

Reprenons l'essai entier, et montrons que chacun de ses documents repose, sans l'avoir dit, sur les deux piliers que celui-ci vient d'établir.

De l'entaille à l'algorithme racontait plusieurs dizaines de millénaires d'histoire de la donnée. Cette histoire n'a pu commencer que parce que le réel était lisible et parce qu'un lecteur était apparu, capable de graver une entaille. Un os entaillé est un acte de lecture : il suppose les deux piliers.

Lire, écrire, comprendre le monde par la donnée examinait la littératie. Or la littératie n'est rien d'autre que la culture du second pilier : l'art d'être

un meilleur lecteur. Former à la littératie de la donnée, c'est affiner les facultés du lecteur que le flux a produit.

Le flux et l'annotation posait l'ontologie processuelle : la donnée est un prélèvement sur le flux. Ce document-ci en fournit le soubassement : le prélèvement est possible parce que le flux est lisible (premier pilier) et parce qu'il existe un préleveur (second pilier).

Creuser le canal développe le flux d'observation. Cette construction suppose un constructeur capable de viser, de choisir des critères, d'inscrire — c'est-à-dire un lecteur. Et la partialité reconnue du flux d'observation trouve ici sa cause profonde : le lecteur est situé.

Les deux temps de la donnée distinguait l'acte d'inscription et la fixation sur un support. Ce document-ci montre d'où vient l'acte : il vient d'un lecteur, lui-même produit du flux. Et il montre ce qui précède l'acte : la trace, sans laquelle l'acte n'aurait rien à inscrire.

Chaque document de l'essai, donc, supposait silencieusement la lisibilité du réel et l'existence du lecteur. Aucun ne l'avait dit. Ce document l'a dit. Il était les fondations ; il rejoint maintenant l'édifice.

Pourquoi l'ontologie processuelle exige cette fondation

Une question pourrait demeurer : pourquoi avoir tant attendu pour poser le socle ? Pourquoi ne pas l'avoir posé d'emblée ?

Parce que les fondations, dans l'ordre de la construction, se coulent en premier, mais que, dans l'ordre de la compréhension, on ne saisit leur nécessité qu'une fois l'édifice dressé. Tant qu'on n'a pas vu le bâtiment, on ne mesure pas ce qu'il faut pour le porter.

L'ontologie processuelle affirme que la donnée n'existe pas avant son prélèvement — qu'elle naît d'un acte. Cette affirmation est forte, et elle prête le flanc à une objection redoutable : *si la donnée n'existe pas avant l'acte, alors la connaissance serait une pure création de l'esprit, sans ancrage dans le réel ; l'humain ne lirait pas le monde, il l'inventerait*. C'est l'objection du constructivisme radical, et si on ne peut pas y répondre, l'ontologie processuelle s'effondre dans l'arbitraire.

Les deux piliers sont la réponse. Oui, la donnée naît d'un acte — mais cet acte est une **lecture**, non une invention. Il lit une trace *réelle*, physiquement déposée par le flux du réel lui-même. La donnée n'est pas arbitraire : elle est tenue, par la trace, à un ancrage matériel dans le réel. Et le lecteur qui accomplit l'acte n'est pas un esprit flottant au-dessus du monde : il est un produit du monde, façonné par lui pour le lire. La donnée est donc, tout à la fois, *construite* — par l'acte du lecteur — et *fondée dans le réel* — par la trace que le lecteur lit et par sa nature de lecteur-issu-du-réel.

Un exemple éclaire la distinction. Imaginons quelqu'un qui affirme, sans avoir rien examiné, par pure décision : « cet arbre a 87 ans ». C'est

l'invention pure. Aucune trace ne tient cette donnée ; n'importe qui peut affirmer le contraire avec la même autorité — c'est-à-dire aucune. Voilà ce à quoi ressemblerait la donnée si le constructivisme radical était vrai : un nombre suspendu dans le vide, sans rien dans le monde pour l'arbitrer.

Le dendrochronologue obtient parfois le même chiffre — 87 ans — mais par un tout autre chemin. Il a vu les cernes, appliqué la règle *un cerne par année*, compté. Sa donnée est construite : par son acte, sa règle, sa décision de compter ceci et pas cela. Mais elle est aussi *tenue* — par les cernes, qui sont là, dans le bois, déposés par le processus de croissance, vérifiables par quiconque reprend la même opération. Si l'on conteste son résultat, on peut recompter ; la trace est là pour arbitrer.

Construction sans invention : voilà la position exacte que les deux piliers rendent possible. La donnée est l'acte d'un lecteur — le constructivisme avait raison sur ce point — mais cet acte lit une trace réelle, et le lecteur lui-même est réel, façonné par le flux pour le lire — le constructivisme avait tort de le suggérer.

Et c'est pour cela que les deux piliers étaient indispensables. Ils tiennent l'ontologie processuelle à égale distance de deux abîmes. D'un côté, l'ontologie objectale, qui croit que la donnée préexiste dans le réel : les piliers la réfutent en montrant que le réel ne contient que des traces, jamais des données. De l'autre, le constructivisme arbitraire, qui croit que la donnée est une libre invention de l'esprit : les piliers le réfutent en montrant que tout acte de lecture est tenu par une trace réelle et accompli par un lecteur lui-même réel. Entre ces deux abîmes, il y a un sol étroit et ferme : la donnée est ce qu'un lecteur, issu du flux, produit en lisant les traces que le flux laisse de lui-même. Tout l'essai tient sur ce sol.

Conclusion — La vague et l'océan

Ce document était les fondations. Reste, pour finir, à regarder ce qu'elles portent.

Nous avons établi deux piliers. Le réel garde trace de lui-même : c'est la lisibilité du réel, et c'est ce qui le rend prélevable. Le flux du réel a produit un être capable de lire ces traces : c'est le lecteur, et c'est ce qui rend la lisibilité effective. La donnée naît à la rencontre exacte des deux — un acte de lecture, accompli par un être issu du flux, sur une trace déposée par le flux. Ni le réel seul ni l'humain seul ne produisent la donnée : il faut leur rencontre.

Nous avons aussi tenu une distinction sans laquelle tout l'essai s'effondrerait : la trace n'est pas la donnée. Le réel est plein de traces ; il est vide de données. Les données n'apparaissent qu'avec le lecteur, qu'avec son acte, qu'avec son intention. Il est donc faux de dire que nous

vivons entourés de données. Nous vivons entourés de réel qui s'écoule et qui, en s'écoulant, se marque lui-même. Les données, c'est nous qui les faisons, en lisant ces marques. Le monde ne nous tend pas des données ; il nous tend une matière lisible, et il a fait de nous, par surcroît, les êtres capables de la lire.

Reste une dernière chose à dire, et elle ouvre vers le document qui clôt l'essai. Le lecteur que le flux a produit occupe une position étrange. Il est une portion du flux du réel — fait de la même matière, soumis au même écoulement, périssable comme tout ce que le flux emporte. Et pourtant, parce qu'il sait lire, il s'éprouve comme distinct du flux : il l'observe, le mesure, le découpe en données, construit des canaux, et finit par ambitionner de le maîtriser. Un être issu du flux qui prétend gouverner le flux. Une portion du réel qui se retourne sur le tout dont elle provient.

C'est l'image d'une vague qui voudrait posséder l'océan. La vague est faite d'eau de l'océan ; elle n'existe que portée par lui ; elle y retournera. Et pourtant, le temps qu'elle se dresse, elle se distingue de lui, elle prend forme, elle paraît, un instant, le surplomber. La donnée est l'écume de ce mouvement : le point précis où la vague, fugacement dressée au-dessus de l'océan dont elle est faite, prend connaissance de l'océan.

Comprendre ce que cela engage — pour l'humain, pour la place qu'il accorde à la donnée dans sa vie collective — est l'affaire du document qui clôt l'essai. Celui-ci aura fait son travail s'il a montré le sol sur lequel tout le reste repose : il n'y a de donnée que parce que le réel se laisse lire, et parce que le réel a produit, en l'humain, son lecteur.

« *Le monde est fait pour aboutir à un livre.* » Non : le monde n'aboutit à aucun livre. Mais il a produit un être qui, lisant les traces que le monde laisse de lui-même, écrit des livres — et celui-ci en est un.

Repères

Pour les références mobilisées dans ce document — la sémiotique de Peirce (la trace comme indice), la métaphore galiléenne du livre de la nature, la notion d'*Umwelt* de Jakob von Uexküll, l'ontologie processuelle de Whitehead, les artefacts archéologiques (l'os entaillé évoqué Partie I) —, voir la *Bibliographie raisonnée* en fin d'essai. Le statut de méthode (fond/forme) est posé dans la *Préface méthodologique*.

Creuser le canal

Le flux d'observation : comment l'humain reconstruit le réel par la donnée

« On ne se baigne jamais deux fois dans le même fleuve. »

— Héraclite, fragment 91

« Donnez-moi un point d'appui et je soulèverai le monde. »

— Attribué à Archimède

Préambule — Le maillon manquant

Les documents qui précèdent ont construit, par étapes, une compréhension processuelle de la donnée. Le troisième a posé l'ontologie processuelle et a introduit la notion qui est l'objet du présent document, en l'énonçant sous la forme suivante :

La donnée est le grain élémentaire d'un flux d'observation que l'humain construit par son intention, à partir du flux du réel. Ce flux d'observation est fidèle au flux du réel relativement aux critères qui l'ont défini, sans jamais en être une réplique totale. De ce flux d'observation peuvent être extraites des données consolidées à haute valeur d'usage, qui alimentent décisions et connaissances, et qui peuvent à leur tour circuler entre entités pour engendrer de nouveaux flux d'observation dérivés.

Le présent document prend cette définition comme point de départ et l'approfondit. Il s'agit de comprendre en détail ce qu'est un flux d'observation : comment il naît, ce qui le distingue du flux du réel, comment il reproduit le mouvement, comment on le manipule, à quelles conditions il est de qualité, comment il peut transformer le réel qu'il observe, et ce qu'il implique pour la littérature de la donnée.

Le flux d'observation est le **maillon manquant** de l'ontologie processuelle. Sans lui, la théorie restait contemplative : elle disait que la donnée se périme, que le flux nous échappe, mais ne disait pas comment l'humain s'y prend, concrètement, pour faire de la donnée un usage fécond. Avec lui, la théorie devient praticable. Elle décrit non seulement la condition de la donnée, mais l'**art** de la donnée — l'art de creuser des canaux dans le fleuve du réel.

Note de méthode : ce document, comme l'ensemble de l'essai, se tient au fond — il dit ce qui est vrai du flux d'observation indépendamment des technologies qui le mettent en œuvre — et s'interdit la forme. La règle est posée dans la *Préface méthodologique* ; elle vaut pour tout ce qui suit. Les illustrations mobilisées (météorologie, cinéma, photographie, sol aride) y sont, conformément à la préface, des illustrations contingentes.

Partie I — Qu'est-ce qu'un flux d'observation ?

Définition et position ontologique

Un flux d'observation est un flux *construit*. Cette construction le distingue radicalement du flux du réel, qui, lui, n'est construit par personne.

Reprenons les trois plans de l'ontologie processuelle :

Le **flux du réel** est ce qui se passe, indépendamment de toute observation. Il s'écoule en permanence, à toutes les échelles, sans interruption. Il est immense, dense, irréversible, indifférent à l'humain. Personne ne l'a creusé ; personne ne peut l'arrêter.

Le **flux d'observation** est un flux que l'humain construit à partir du flux du réel. Il est constitué par l'accumulation cadencée de prélèvements effectués selon des critères intentionnels. Il est partiel, fidèle relativement à ses critères, et manipulable.

La **donnée consolidée** est le grain informationnel synthétique que l'on extrait d'un flux d'observation par traitement : une moyenne, une tendance, une équation, un statut, un indicateur. Elle est à haute valeur d'usage et peut elle-même alimenter un nouveau flux d'observation.

Le flux d'observation occupe une position intermédiaire et charnière. Il n'est ni le réel brut ni l'information finale : il est l'**espace de travail** que l'humain se constitue entre les deux. C'est dans cet espace que se joue tout l'art de la donnée.

La naissance d'un flux d'observation : l'intention et les critères

Un flux d'observation ne naît pas spontanément. Il naît d'une **intention** : une entité — un individu, une entreprise, une administration, un

laboratoire — identifie un besoin, une question, un sujet qu'elle veut suivre. Sans intention, il n'y a pas de flux d'observation, seulement des prélèvements épars.

De l'intention découlent les **critères**. Définir les critères d'un flux d'observation, c'est répondre à une série de questions : *Quoi ?* Quels faits bruts observables du flux du réel veut-on prélever ? *À quelle cadence ?* À quelle fréquence effectue-t-on les prélèvements ? *Avec quelle fidélité ?* Quelle précision exige-t-on de chaque prélèvement, quels instruments, quelle calibration, quelle marge d'erreur acceptable ? *Avec quel corpus ?* Quelles données de contexte accompagnent chaque prélèvement pour le rendre interprétable ? *Sur quel périmètre ?* Quelle portion du flux du réel observe-t-on, et laquelle laisse-t-on de côté ?

Ces critères ne sont pas neutres. Ils sont l'expression directe de l'intention. Un flux d'observation construit pour surveiller la santé d'un patient n'aura pas les mêmes critères qu'un flux construit pour optimiser une chaîne logistique. **Les critères sont la signature de l'intention dans le flux d'observation.**

C'est un point essentiel pour la littératie : devant tout flux d'observation, la question première n'est pas "que dit-il ?" mais "quelle intention l'a fait naître, et quels critères en découlent ?". Un flux d'observation ne ment jamais sur le réel ; mais il ne montre que ce que ses critères lui permettent de montrer.

La fidélité relative aux critères

Une précision conceptuelle décisive s'impose ici. On serait tenté de dire qu'un flux d'observation est "une copie partielle conforme" du flux du réel. C'est inexact. Aucun flux d'observation n'est conforme au flux du réel *en général* — il ne peut pas l'être, puisqu'il en est infiniment plus pauvre. La formulation juste est la suivante :

Un flux d'observation est fidèle au flux du réel relativement aux critères qui ont permis de le construire.

Une analogie photographique éclaire ce point. Un photographe qui capture un objet en mouvement avec une vitesse d'obturation lente obtiendra une image où l'objet apparaît sous forme de traînée. Cette photo n'est pas "infidèle" : elle est une réplique parfaitement fidèle du réel *relativement au critère choisi*. La traînée n'est pas une erreur ; c'est la conséquence exacte du critère. Un autre photographe, avec une obturation rapide, figerait l'objet net — autre critère, autre fidélité, autre image, du même réel.

Il en va de même pour tout flux d'observation. Ses limites ne sont pas des défauts : ce sont les *empreintes de ses critères*. Un flux d'observation qui prélève la température une fois par jour ne "manque" pas les variations horaires par négligence ; il les laisse hors champ par construction.

La conséquence pour la littérature est nette : critiquer un flux d'observation, ce n'est pas lui reprocher de ne pas tout montrer — aucun ne le peut. C'est se demander si **les critères qui l'ont défini sont adaptés à l'usage qu'on veut en faire**. La critique pertinente ne porte pas sur la fidélité — qui est toujours relative et toujours réelle — mais sur la pertinence des critères.

Les propriétés du flux d'observation

Le flux d'observation possède un ensemble de propriétés qui le distinguent point par point du flux du réel.

Le flux du réel est continu ; le flux d'observation est discret. Le réel ne saute pas ; il s'écoule sans trous. Le flux d'observation est fait de prélèvements ponctuels séparés par des intervalles. Entre deux prélèvements, le flux d'observation ne contient rien — c'est un pointillé là où le réel est une ligne pleine.

Le flux du réel est immense ; le flux d'observation est restreint. Le réel comporte une infinité de faits bruts observables. Le flux d'observation n'en retient qu'une sélection délibérée. C'est précisément ce qui le rend *utile* : il est lisible parce qu'il est pauvre. Une carte qui représenterait tout le territoire à l'échelle 1:1 serait inutilisable.

Le flux du réel est indifférent ; le flux d'observation est intentionnel. Le réel ne poursuit aucun but. Le flux d'observation est tout entier l'expression d'un but. Il porte la marque de l'intention qui l'a creusé.

Le flux du réel est irréversible ; le flux d'observation est rejouable. Le réel ne revient jamais en arrière. Le flux d'observation, une fois enregistré, peut être rembobiné, relu, rejoué autant de fois qu'on le souhaite.

Le flux du réel est incontrôlable ; le flux d'observation est manipulable. On ne peut ni arrêter le réel, ni l'accélérer, ni le mettre en pause. Le flux d'observation peut être activé, suspendu, accéléré, ralenti, filtré, enrichi, fusionné, supprimé. Il obéit à son constructeur.

Le flux du réel n'appartient à personne ; le flux d'observation appartient à son constructeur. Le réel est un commun absolu. Le flux d'observation, en tant qu'artefact construit, peut être approprié, conservé, transmis, échangé. Ce qui s'approprie n'est jamais le réel — seulement le canal qu'on en a creusé.

Cette série d'oppositions peut se résumer d'une formule : **le flux du réel est donné, le flux d'observation est fait**. L'humain est le constructeur de canaux, pas le propriétaire des eaux.

Partie II — La mise en mouvement

Le paradoxe de l'inerte et du mouvement

Une donnée est un prélèvement *figé, inerte* : une fois inscrite, elle ne change plus. Comment, dès lors, un flux d'observation — qui n'est qu'une collection de données figées — peut-il restituer le *mouvement* du flux du réel ?

La réponse tient en deux mots : la **lecture enchaînée**. Une donnée isolée est inerte. Deux données consécutives, lues l'une après l'autre, font apparaître une différence — donc un changement, donc un embryon de mouvement. Une longue série de données, lue dans l'ordre et à un rythme suffisant, recrée le mouvement continu du flux du réel au moment où les prélèvements ont été effectués.

Le flux d'observation n'est donc pas en mouvement *en lui-même* — il est, au repos, une collection de points figés. Il *se met en mouvement* quand on le lit. Le mouvement n'est pas dans les données ; il est dans l'acte de les parcourir. Une pellicule de film, posée sur une table, n'est qu'une bande de photos immobiles ; elle ne devient mouvement que lorsqu'elle défile dans le projecteur.

C'est une propriété remarquable : le flux d'observation contient le mouvement **en puissance**, sous forme repliée, et le déploie à chaque lecture. On peut le déplier autant de fois qu'on veut, dans l'ordre ou à rebours, vite ou lentement. Le mouvement du réel, qui était irréversible et unique, devient — dans le flux d'observation — reproductible et maniable.

L'analogie du cinéma et le seuil

L'analogie la plus éclairante pour comprendre cette mise en mouvement est le cinéma.

Le cinéma naît d'un seuil. Une succession d'images fixes projetées lentement n'est qu'un diaporama : l'œil perçoit des images séparées. Mais lorsque la cadence franchit un certain seuil — pour le cinéma argentique classique, historiquement situé autour de vingt-quatre images par seconde —, quelque chose bascule : l'œil ne perçoit plus des images, il perçoit du mouvement. On a changé de nature perçue.

Il en va exactement de même pour le flux d'observation. Des prélèvements trop espacés ne donnent qu'une succession d'instantanés sans lien perçu. Mais lorsque la cadence franchit un seuil approprié au sujet observé, le flux d'observation, à la lecture, restitue le mouvement du flux du réel. On bascule du relevé ponctuel au flux véritable.

L'analogie va plus loin. Le cinéma ne s'est pas arrêté à ce premier seuil. Son histoire est celle d'une accumulation d'améliorations : synchronisation du son, couleur, augmentation de la définition, du format. Chacune de ces améliorations a étendu le corpus et amélioré

la probité — et au terme de cette accumulation, on ne parle plus d'un diaporama devenu fluide, on parle d'un art à part entière. L'amélioration des critères d'un flux d'observation n'est pas seulement quantitative : au-delà d'un certain cumul, elle produit un **saut qualitatif**. Le flux d'observation change de nature, devient capable de choses qu'il ne pouvait pas faire avant. Nous reviendrons sur ce phénomène en Partie VII.

L'analogie a aussi ses limites. Le cinéma crée le mouvement par une propriété du cerveau humain : l'œil interpole entre les images. Le mouvement perçu au cinéma est une reconstruction perceptive ; il n'est pas dans la pellicule. Or, dans le cas d'un flux d'observation exploité par un système automatique de décision, il n'y a pas de cerveau qui interpole. Le système prend la succession de prélèvements pour une continuité, sans la prudence implicite du spectateur qui sait qu'il regarde un film. Entre deux prélèvements consécutifs, n'importe quoi peut s'être produit dans le flux du réel — un événement bref qui n'a laissé aucune trace, parce qu'il s'est glissé entre deux points.

L'analogie doit donc être tenue avec sa limite : **le flux d'observation produit une illusion fonctionnelle de continuité, jamais une continuité réelle**. Apprendre à ne jamais confondre le film avec le monde, même quand le film est parfait, est l'un des objets de la littérature du flux d'observation.

L'analogie de la compression d'image

Une seconde analogie éclaire un autre aspect : la **compression d'image**.

Quand on photographie une scène et qu'on enregistre l'image dans un format compressé, on obtient un fichier qui reproduit fidèlement la scène pour l'usage courant. Mais si l'on zoome, on finit par atteindre une limite : des aplats, des blocs, des artefacts apparaissent. La compression a conservé ce qui était perceptible à l'échelle prévue, et sacrifié le reste.

C'est le fonctionnement même du flux d'observation. Il est une **compression du flux du réel**. Il en reproduit fidèlement ce que ses critères ont retenu, à l'échelle d'usage prévue. Mais si l'on cherche à l'exploiter à une finesse supérieure à celle pour laquelle il a été construit, on n'extrait que des artefacts.

Là où l'analogie du cinéma éclaire la dimension *temporelle* (cadence, mouvement), la compression éclaire la dimension de *résolution*. Un flux d'observation possède une **résolution temporelle** (donnée par sa cadence) et une **résolution descriptive** (donnée par la richesse de son corpus et la précision de ses prélèvements). Au-delà de l'une ou de l'autre, on quitte le domaine de validité du flux.

Le seuil minimal : ne pas perdre le signal

De ces analogies découle une règle de fond. Il existe un **seuil minimal** de cadence en dessous duquel le flux d’observation cesse de représenter fidèlement le mouvement du flux du réel. La théorie du signal connaît ce principe sous le nom de critère de Nyquist-Shannon : pour représenter fidèlement un phénomène qui varie, il faut l’échantillonner à une fréquence suffisamment supérieure à la fréquence de ses variations. En deçà, on produit des artefacts trompeurs — le phénomène où les roues d’une voiture semblent tourner à l’envers dans un film, parce que la cadence est mal accordée à la vitesse de rotation.

Transposé au flux d’observation : **la cadence de prélèvement doit être accordée à la vitesse de changement du flux du réel observé**. Un flux du réel qui change vite exige une cadence rapide ; un flux qui change lentement tolère une cadence lente. Si la cadence est inférieure au seuil, le flux d’observation ment — non pas en disant le faux sur chaque point, mais en laissant croire à une continuité qu’il n’a pas captée.

Ce seuil minimal est un fait de fond. Mais respecter le seuil minimal ne suffit pas. Être au-dessus du plancher n’est pas être au bon endroit.

Partie III — L’optimum observationnel

Plus n’est pas mieux

Dans le discours qui entoure le *Big Data*, une conviction s’est largement répandue : prélever davantage serait toujours préférable. C’est cette conviction que le concept d’optimum observationnel vient contredire.

Considérons un flux d’observation météorologique. Le passage de relevés manuels peu fréquents à des relevés par capteurs plus précis a vraisemblablement amélioré la qualité des prévisions. Mais faut-il densifier indéfiniment ? Passer d’un relevé toutes les dix minutes à un relevé chaque seconde n’apporterait, selon toute vraisemblance, presque rien — le temps qu’il fait ne change pas significativement d’une seconde à l’autre. On aurait multiplié le volume de données, la puissance de calcul, le coût de stockage, pour un gain quasi nul.

Ce sur-prélèvement a un nom : la **sur-ingénierie**. C’est l’erreur symétrique du sous-prélèvement. Le sous-prélèvement fait passer sous le seuil de Nyquist : on perd le signal. La sur-ingénierie fait monter très au-dessus du nécessaire : on ne gagne rien en signal, on perd en coût.

Définition de l’optimum observationnel

Entre le plancher du sous-prélèvement et le gaspillage de la sur-ingénierie, il existe une zone où le flux d’observation est construit *au mieux*. Le point

central de cette zone est l'**optimum observationnel**.

L'optimum observationnel est le point où l'arbitrage entre le bénéfice (la qualité du résultat obtenu) et le coût (la quantité de moyens mobilisés : prélèvement, transport, stockage, calcul, temps) est le plus satisfaisant pour l'usage visé.

L'optimum n'est pas l'état où l'on prélève le plus. C'est l'état où l'on prélève *ce qu'il faut* : assez pour que le flux d'observation soit fidèle et utile à son usage, pas plus que ce que cet usage requiert. Être à l'optimum, c'est être en **situation optimale d'observation** : ni aveugle, ni noyé.

Trois variables au moins entrent dans cet arbitrage : la **cadence** (fréquence des prélèvements), le **corpus** (richesse des paramètres captés et des données de contexte), et la **probité** (précision et fiabilité de chaque prélèvement). Chacune a un coût, et chacune, au-delà d'un certain point, cesse d'apporter un bénéfice proportionné. L'optimum est le réglage conjoint de ces trois variables qui maximise l'écart favorable entre bénéfice et coût.

L'optimum dépend du contexte du sujet observé

Il n'existe pas d'optimum observationnel universel. L'optimum dépend entièrement du contexte : de la vitesse à laquelle change le fragment du flux du réel observé, et de la finesse que l'usage exige.

La question qui détermine l'optimum est toujours la même : *le flux du réel observé peut-il changer significativement entre deux prélèvements ?* Si oui, la cadence est trop lente. Si non, la cadence est inutilement rapide.

Un flux d'observation sur l'activité sismique d'une faille, un flux sur la fréquentation d'un musée, un flux sur l'évolution d'un glacier, un flux sur le trafic d'un réseau informatique : chacun a son optimum propre. L'optimum observationnel n'est pas une valeur ; c'est une *question* qu'on doit reposer à chaque fois.

Le concept affirme — et c'est de fond — qu'un optimum observationnel existe pour tout flux d'observation, et que le situer est nécessaire pour faire de la donnée un usage pleinement satisfaisant. Il ne détermine pas — et c'est de la forme — où se situe l'optimum dans tel cas concret. Ce point dépend du sujet, de l'usage, des coûts du moment, des contraintes de l'entité. Il dépend, en dernière instance, du jugement de l'humain qui construit le flux d'observation. C'est exactement la posture de 1950 sur l'automobile : on affirme qu'il faudra des routes (fond), on ne fixe pas leur largeur (forme).

Partie IV — La manipulation du flux d'observation

Un flux que l'on maîtrise

Le flux d'observation, parce qu'il est un artefact construit et enregistré, se prête à un ensemble d'opérations. Recensons-les brièvement, car chacune est une capacité que l'humain gagne sur le réel.

Stocker. Conserver le flux d'observation, alors que le flux du réel est perdu à chaque instant.

Rejouer. Relire un flux enregistré autant de fois qu'on le souhaite. Le mouvement du réel, unique et irréversible, devient reproductible : ralenti, accéléré, arrêt sur instant.

Activer et désactiver. Suspendre puis reprendre. Observer par intermittence, en fonction des besoins et des moyens.

Consolider. Extraire d'un flux des données consolidées — moyennes, tendances, statuts — qui résument à haute valeur d'usage ce que le flux contenait sous forme dispersée.

Épurer. Retirer des prélèvements redondants, écarter des paramètres devenus inutiles, réduire la cadence. L'opération qui permet de se rapprocher de l'optimum quand on s'aperçoit avoir sur-prélevé.

Enrichir et fusionner. Adjoindre des données de contexte, ou combiner deux flux pour en former un troisième, plus riche.

Transmettre. Diriger un flux vers une autre entité, qui l'exploitera selon ses propres critères. Ce point ouvre la dynamique inter-entités (partie suivante).

Supprimer. Détruire un flux. Capacité qui est aussi une responsabilité : ce qui est supprimé ne pourra jamais être re-prélevé à l'identique, puisque le flux du réel a continué.

Rejouer ou prévoir : les deux usages du flux d'observation maîtrisé

Parmi ces opérations, deux usages méritent un développement particulier.

Rejouer — la prise sur le passé. Un flux d'observation enregistré permet de revenir sur un moment révolu et de l'examiner à loisir. C'est par cette capacité que l'humain transforme l'irréversible en analysable.

Prévoir — la prise sur l'avenir. Un flux d'observation suffisamment dense et fidèle, observé sur une durée suffisante, fait apparaître des régularités, des tendances, des cycles. À partir d'elles, on peut formuler des prévisions : anticiper l'état probable du flux du réel à un instant futur. Ce n'est pas une prise directe sur l'avenir, mais une réduction de l'incertitude.

Ces deux usages reposent sur la même propriété de fond : un flux d'observation, contrairement au flux du réel, est un objet *maniable dans le temps*.

Un point important sur la fraîcheur. La fraîcheur d'un flux d'observation conditionne son rapport au réel *actuel*. Mais elle ne conditionne pas sa capacité à être *rejoué*. Un flux ancien, périmé pour représenter le présent, reste parfaitement jouable comme représentation fidèle de son époque. Un film de 1920 restitue le mouvement aussi bien qu'un film d'aujourd'hui ; il ne restitue simplement plus le présent.

La donnée consolidée : extraire la substance

Quand une entité traite et analyse son flux d'observation, elle n'en conserve pas l'intégralité des prélèvements bruts comme résultat. Elle en extrait une **synthèse** : moyenne, courbe de tendance, équation, statut, indicateur, diagnostic. C'est la donnée consolidée.

La donnée consolidée a trois caractéristiques propres. Elle est de **haute valeur d'usage** : elle répond directement à la question que l'intention avait posée. Elle est **contextualisée** : elle transporte avec elle les données annexes qui permettent de l'interpréter — la période sur laquelle porte la moyenne, les conditions de la mesure, la marge d'incertitude. Elle est **réutilisable** : parce qu'elle est synthétique et contextualisée, elle peut servir de point de départ à d'autres raisonnements, et notamment alimenter un nouveau flux d'observation.

Partie V — La dynamique : flux qui engendrent des flux

Le flux d'observation n'est pas une finalité

Un flux d'observation est un **point de passage** dans une dynamique qui peut se prolonger, sous deux formes.

La dérivation interne. D'un flux d'observation, on extrait des données consolidées. Mais ces données consolidées, accumulées à leur tour de manière cadencée, forment elles-mêmes un nouveau flux — un flux d'observation de second rang, qui n'observe plus directement le flux du réel mais le premier flux d'observation. Chaque niveau est une compression du précédent, plus pauvre en détail, plus riche en sens.

La transmission inter-entités. Un flux d'observation construit par une entité peut être dirigé vers une autre entité. Celle-ci le reçoit et le traite selon ses propres critères, qui ne sont pas ceux de l'entité d'origine. Elle en construit donc un nouveau flux d'observation dérivé, porteur de sa propre intention.

La substance n'est pas épuisable

Cette dynamique de transmission repose sur une propriété singulière de la donnée. Quand on extrait la substance d'une matière physique, la matière est consommée. La donnée ne fonctionne pas ainsi. Quand une entité a extrait d'un flux d'observation la donnée consolidée qui l'intéressait, elle en a tiré *sa* substance — celle que son intention cherchait. Mais le flux n'est pas vidé. Une autre entité, porteuse d'une autre intention, peut en extraire une substance toute différente.

Un flux d'observation sur les déplacements d'une flotte de véhicules a livré au transporteur sa substance logistique. Le même flux peut livrer à un urbaniste une substance sur l'usage des voiries, à un chercheur une substance sur les comportements de mobilité, à un assureur une substance sur l'exposition au risque. La dynamique inter-entités n'est pas le recyclage d'un déchet ; c'est la mise en circulation d'une ressource non épuisée.

Une intelligence distribuée

Cette dynamique dessine, à grande échelle, quelque chose qui dépasse chaque entité prise isolément. Si de nombreuses entités construisent chacune leurs flux d'observation, et si ces flux peuvent circuler et se féconder mutuellement, alors se constitue une **représentation distribuée du flux du réel** — fragmentaire, faite de pièces hétérogènes, mais collectivement plus riche que ce qu'aucune entité ne pourrait construire seule.

Cette représentation distribuée n'est pas un stock centralisé. C'est un tissu vivant de flux d'observation entretenus en parallèle par une multitude d'entités, dont chacune reste responsable de ses propres canaux. Sa richesse tient à la diversité des intentions et à la possibilité de les mettre en relation.

Mais cette intelligence distribuée a une condition : elle ne se constitue que si les intentions qui creusent les canaux sont *plurielles et indépendantes*. Un tissu de flux concentré entre quelques mains — quelques plateformes, quelques États — n'est pas une intelligence distribuée : c'est une surveillance distribuée. La différence n'est pas technique ; elle est politique. Elle tient à la question de savoir qui décide des critères, au service de quelles intentions, avec quelle possibilité de contestation.

Une architecture distribuée des modèles

Cette dynamique inter-entités permet de revenir sur une question laissée ouverte dans *Le flux et l'annotation* : peut-on concevoir une intelligence artificielle qui apprendrait du flux du réel en continu, plutôt que d'un corpus figé ?

La question, telle que posée, contient une erreur. Une intelligence artificielle ne peut pas apprendre du flux du réel directement — ce flux n'est pas observable en tant que tel. Il ne l'est que par les flux d'observation que des entités humaines maintiennent. Une « intelligence en continu » n'est donc jamais branchée sur le réel : elle est branchée sur des canaux entretenus par d'autres.

Cette précision posée, le concept de flux d'observation dessine l'architecture que la question appelait. **Tout corpus d'entraînement est un assemblage de flux d'observation figés** : on ne présente pas « le monde » à un modèle, on lui présente ce que des canaux humains, à une date donnée, ont déposé. Or les flux diffèrent par leur taux de péremption — certains observent un réel à péremption lente (la grammaire d'une langue, les théorèmes, les grandes régularités), d'autres un réel à péremption rapide (l'actualité, les marchés, l'état d'un projet). De cette différence découle un arbitrage : **entraîner et donner accès ne répondent pas à la même question**. Un flux lent justifie l'entraînement, dont le coût est amorti par la durée de validité du modèle. Un flux rapide appelle plutôt un accès — on branche le modèle sur le canal tenu à jour, plutôt que de lui faire absorber ce que ce canal a déposé hier.

Cette logique dessine, à grande échelle, non pas un modèle universel mais une constellation de modèles spécialisés, différenciés sur deux axes : le **thème** — un modèle par domaine, entraîné sur les flux pertinents pour son terrain — et la **durée de fraîcheur** — certains modèles, ancrés dans des corpus lents, restent valides longtemps ; d'autres sont rafraîchis fréquemment, ou remplacés par de simples connecteurs au flux entretenu. C'est l'incarnation algorithmique de l'intelligence distribuée décrite plus haut. Le dessin a une portée politique qu'il faut nommer : une concentration de l'IA autour de quelques très grands modèles tout-terrain, entraînés sur d'énormes corpus indifférenciés, n'est pas une intelligence distribuée mais une **surveillance distribuée**. La forme architecturale ressemble ; la pluralité des intentions, elle, fait défaut. Et conformément à la règle de méthode, ce qui précède est de fond — péremption différenciée des corpus, arbitrage entraînement/accès, supériorité d'une architecture plurielle pour la pluralité des intentions — et ce qui reste de forme, l'essai ne le tranche pas : quels découpages thématiques, quels protocoles de connexion, sous quelles régulations. La carte précise reste à dresser.

L'IA n'est donc ni le sommet d'une pyramide cognitive, ni un instrument unique. Elle est une famille d'instruments différenciés, articulée à un tissu humain qui maintient les flux dont elle dérive et auxquels elle se connecte. Son avenir ne se joue pas dans l'autonomisation d'un modèle qui se passerait des humains. Il se joue dans la qualité, la pluralité et l'indépendance des canaux que ces humains entretiennent.

Partie VI — La boucle : quand l'observation transforme le réel

Observer pour décider, décider pour agir

Jusqu'ici, nous avons décrit le flux d'observation comme un canal qui prélève dans le flux du réel sans l'affecter. Cette description est incomplète. Le flux d'observation, le plus souvent, n'est pas construit pour la seule contemplation. Il est construit pour **décider**. Et la décision, mise en œuvre, devient **action**. Et l'action, s'exerçant sur le monde, **modifie le flux du réel** lui-même.

Le schéma complet n'est donc pas une flèche, mais une **boucle** :

Flux du réel → prélèvements → flux d'observation → données consolidées → décision → action → modification du flux du réel → nouveaux prélèvements...

Quand un flux d'observation est utilisé pour décider et agir sur le fragment du réel qu'il observe, il forme un **circuit fermé**.

Une illustration : le sol aride

Une entité veut savoir quelle espèce d'arbre planter sur une zone de sol aride. Elle construit un flux d'observation : nature du sol, paramètres météorologiques, humidité, ensoleillement. De ce flux, elle tire une donnée consolidée : telle espèce d'arbre est adaptée. C'est une décision. L'entité plante.

Les arbres grandissent. Et, en grandissant, ils modifient le fragment du réel qui avait été observé : ils retiennent l'humidité, ombragent le sol, atténuent l'aridité. Le flux du réel de cette zone n'est plus ce qu'il était. Si l'entité construit alors un nouveau flux d'observation avec les mêmes critères, les prélèvements révéleront un réel transformé. L'écart entre les premiers prélèvements et les nouveaux n'est pas un défaut de mesure : c'est l'**effet de l'action**.

Un point mérite d'être relevé. Les premiers prélèvements — ceux du sol aride initial — basculent immédiatement, du fait de l'action, au troisième âge de la donnée : l'âge mémoriel. Ils ne représentent plus le présent du réel, puisque ce présent a été délibérément modifié. La boucle de rétroaction accélère le vieillissement de la donnée : ce n'est plus seulement l'écoulement passif du temps qui périmé les prélèvements, c'est l'action elle-même, issue de leur exploitation, qui les relègue à l'archive.

L'observation n'est pas neutre

Quand un flux d'observation est inséré dans un circuit fermé, **l'observation cesse d'être neutre**. Observer pour agir, c'est s'engager à transformer

ce qu'on observe. Le flux d'observation n'est plus un miroir tendu au réel ; il est le premier maillon d'un processus qui va modifier ce réel.

Cela ne disqualifie nullement l'observation — la boucle, dans l'exemple du sol aride, est parfaitement vertueuse. Mais cela impose une lucidité. Celui qui construit un flux d'observation destiné à la décision n'est pas un spectateur extérieur. Les critères qu'il choisit ne font pas que déterminer ce qu'il verra : ils déterminent, par le relais de la décision et de l'action, ce que le réel deviendra. Observer, quand c'est pour décider, c'est déjà commencer à agir.

Partie VII — Le saut systémique

Nous avons vu, avec l'analogie du cinéma, que l'accumulation d'améliorations sur un flux d'observation peut produire, au-delà d'un certain seuil, un saut qualitatif.

Le fait de fond est le suivant : **l'amélioration des moyens par lesquels un flux d'observation est construit peut produire un changement systémique** — un changement qui ne se réduit pas à une amélioration graduelle, mais transforme la nature de ce que le flux d'observation permet de faire, et parfois la nature de l'activité tout entière qui s'appuie sur lui.

Le cinéma en est l'exemple paradigmatique. Mais le phénomène est général. Le passage de relevés manuels peu fréquents à des relevés par capteurs précis et denses n'a pas seulement "amélioré un peu" tel ou tel flux d'observation — il a rendu possibles des choses qui n'existaient pas avant. Au-delà d'un certain cumul, on ne dit plus "les relevés sont un peu meilleurs" ; on dit que telle activité a changé de régime.

Plus près de nous, deux exemples vécus par beaucoup illustrent des sauts du même ordre, sur des paramètres différents.

Le premier porte sur le **stockage et l'accès à la donnée**. Un grand modèle de langage — ce que l'on désigne aujourd'hui par *LLM*, pour *Large Language Model* — condense en une entité unique trois fonctions qui, dans toute l'histoire humaine, avaient été tenues séparément : la **bibliothèque** où l'information est rangée, le **bibliothécaire** qui sait la retrouver, le **conteur** qui sait la restituer en langue compréhensible. À première vue, ce qui change est l'efficacité : ce qui exigeait trois lieux, trois métiers, trois actes, tient désormais dans une seule interaction. Mais le saut systémique est ailleurs. Chacun des trois rôles, intégré aux deux autres, *change de nature*. La bibliothèque ne range plus des livres : elle compresse un corpus dans les poids d'un réseau, d'où l'on ne peut plus extraire un exemplaire intact. Le bibliothécaire ne consulte plus un catalogue : il génère par calcul une réponse probable. Le conteur ne lit plus un texte : il synthétise du texte qui *ressemble* à ce que le corpus aurait pu produire. Et avec la fusion des

trois rôles disparaît leur séparation — qui jusque-là rendait possibles la traçabilité (savoir d'où vient une information), la critique (contester un classement), le retour à la source (relire le texte original), et le contre-pouvoir (un rôle pouvait corriger l'autre).

Le second exemple porte sur la **vitesse de circulation des signaux**. Quand un utilisateur valide une action en ligne — confirmer un rendez-vous, par exemple —, la notification de confirmation peut arriver sur son téléphone *avant même* que la page web qui a déclenché l'action n'ait fini de s'actualiser. Pourtant, le signal a parcouru un trajet plus long : serveur applicatif, serveur de notification, opérateur, téléphone. La cause de cette inversion apparente est simple : à la vitesse à laquelle le signal voyage aujourd'hui, le temps de transit est devenu si court qu'il passe sous le seuil de la perception humaine. L'expérience subjective n'est plus celle d'un *envoi puis d'une réception* ; c'est celle d'une **simultanéité**. Deux portions du flux du réel — l'écran de l'ordinateur, l'écran du téléphone — sont modifiées par un seul geste, en apparence en même temps.

Chacun de ces deux franchissements de seuil — à sa manière propre — rend possible une **automatisation** à laquelle l'humain a, jusqu'ici, dû procéder lui-même.

Le saut sur la vitesse et la synchronie automatise la **circulation des données** entre dispositifs. Quand la transmission devient imperceptible, des classes entières de tâches répétitives — réserver une place, suivre une commande, prendre un rendez-vous, remplir un formulaire administratif, recouper deux registres, déclencher un rappel, transmettre une pièce justificative — peuvent être confiées à des machines articulées entre elles, qui se passent les données dans un temps dont l'humain n'a plus à se soucier. La déclaration d'impôt pré-remplie est, à cet égard, exemplaire : ce que l'utilisateur devait jadis recopier à la main — revenus, prélèvements, situations — circule désormais entre l'employeur, la banque, l'administration, et lui parvient déjà inscrit. Sa tâche, qui était de *produire* la déclaration, se réduit à la *valider*.

Le saut sur le stockage et l'accès automatise, lui, des **opérations cognitives** que la séparation des trois rôles avait toujours laissées à un agent humain. Lire un courriel et formuler sa réponse, qualifier un texte, mettre en forme un document mal rédigé, convertir le format d'une donnée, synthétiser un dossier volumineux, extraire d'un ensemble la partie pertinente, traduire d'une langue à une autre — autant d'opérations qui exigeaient jusque-là un processus dédié, c'est-à-dire quelqu'un qui lisait, jugeait, reformulait. Elles peuvent désormais être déléguées à un dispositif qui condense en lui-même les fonctions de la bibliothèque, du bibliothécaire et du conteur.

Dans les deux cas, c'est le même geste qui se généralise : « *je confirme, j'accepte* ».

Ces automatisations prolongent, dans le terrain numérique, un mouvement plus ancien. Le métier à tisser avait, en son temps, automatisé une

opération répétitive jusqu'alors humaine — et cette automatisation avait transformé non seulement le textile, mais l'organisation du travail et la manière de se vêtir d'une société. La chaîne de montage, l'imprimerie, le téléphone avaient chacun, en leur temps, déclenché des sauts comparables. Le numérique en déclenche un nouveau, qui porte cette fois sur les **tâches d'annotation, de coordination et de circulation administrative** — ces tâches que le document premier de l'essai désignait comme « *ce qui a, en commun, d'être répétitif* ». Le terrain change ; la structure du phénomène est la même.

Les répercussions sont à double tranchant ; nommons les deux versants. D'un côté, un bénéfice réel : la charge mentale s'allège, le temps libéré peut se reporter sur ce qui compte, et un travailleur indépendant peut, en intégrant ces dispositifs dans son organisation, déléguer ce qui ne le mobilise pas — relances, suivi, comptabilité — pour se concentrer sur son cœur de métier. L'automatisation administrative est, en quelque sorte, la réponse contemporaine à la sur-administration que la même société avait engendrée. De l'autre côté, un dessaisissement : l'usager qui valide ne sait plus ce qui circule, qui voit quoi, quelles données ont alimenté quel pré-remplissage, ni selon quels critères. Sa charge mentale se déplace, son agenda se constitue ailleurs, son attention se libère — ou se laisse capter par ceux qui maintiennent les canaux à sa place.

Cet exemple est, conformément à l'avertissement de méthode, une illustration de la forme — datée, révisable, possiblement caduque dans dix ans, quand l'usage et la technique auront changé. Ce qui reste de fond, lui, ne change pas : un saut de cadence ou de vitesse, au-delà d'un certain seuil, ne se réduit pas à une amélioration graduelle — il transforme la nature de ce qu'un flux d'observation rend possible, la manière dont l'humain s'y articule, et avec elle l'organisation de la société qui s'y appuie.

Ce que le concept affirme — et qui est de fond — tient en peu de mots : certains flux d'observation peuvent, par l'amélioration de leurs moyens de construction, conduire à un changement systémique. Cela était vrai pour le cinéma, ce sera vrai pour des flux d'observation que nous ne savons pas encore nommer. Ce que le concept ne dit pas — parce que c'est de la forme — c'est *quels* flux connaîtront un tel saut, *quand*, *dans quelle mesure*, et *avec quelles conséquences*.

Le concept pointe une possibilité — le saut systémique — pour que le lecteur la garde présente à l'esprit. Il ne prescrit pas et ne prédit pas.

Le concept permet de **pointer** des écueils : la sur-ingénierie, le sous-prélèvement, l'illusion de continuité, la boucle qui n'est pas neutre, le saut systémique aux effets imprévus. Mais il ne peut pas en **décrire précisément** les contours, car ils dépendent des formes concrètes que prendront les choses. Le concept est une lampe : il éclaire la zone où il faut regarder, il avertit qu'il y a là des écueils. Il ne dessine pas la carte détaillée des récifs. Cette carte, c'est à chaque humain, dans chaque

contexte, de la dresser — et c'est précisément ce que la littérature du flux d'observation doit lui rendre possible.

Partie VIII — Ce que le flux d'observation apporte à la littérature

Du stock au flux : un changement de regard

Introduire le flux d'observation dans la littérature de la donnée, c'est opérer un changement de regard fondamental.

La littérature standard apprend à manipuler la donnée comme un *stock* : on l'identifie, on la collecte, on la nettoie, on l'analyse, on la communique. Le verbe directeur est *avoir*.

La littérature enrichie par le concept de flux d'observation apprend à penser la donnée comme un *flux* : elle vient d'un écoulement (le flux du réel), elle est captée dans un canal construit (le flux d'observation), elle se met en mouvement quand on la lit, elle se périmé relativement à un usage, elle peut être rejouée, transmise, dérivée. Le verbe directeur n'est plus *avoir* mais *faire couler*.

Ce changement ne rend pas caduque la littérature standard : on a toujours besoin de savoir nettoyer une donnée ou lire un graphique. Mais il l'enveloppe dans une compréhension plus juste. L'apprenant formé au flux d'observation ne se demande plus seulement "cette donnée est-elle exacte ?" mais "de quel canal vient-elle ? avec quelle intention a-t-il été creusé ? est-il à l'optimum observationnel ? que laisse-t-il hors champ ?"

La chaîne étendue

Le deuxième document a présenté la pyramide DIKW (*Data* → *Information* → *Connaissance* → *Sagesse*) comme outil pédagogique central, et en a signalé les limites. Le concept de flux d'observation permet de proposer une chaîne plus complète :

Flux du réel → *Flux d'observation* → *Données élémentaires*
(prélèvements) → *Données consolidées* → *Information* →
Connaissance → *Sagesse*

Les niveaux supérieurs restent identiques. Mais la littérature doit remonter **deux étages plus bas** que ne le faisait la pyramide classique : non seulement comprendre la donnée et son contexte, mais comprendre le flux d'observation qui la produit — son intention, ses critères, sa cadence, ce qu'il laisse hors champ — et le flux du réel dont ce flux d'observation est une dérivation partielle.

Les compétences de la littératie processuelle

Nous tenons maintenant le matériel pour exposer en pleine force les compétences que cette littératie processuelle ajoute aux cadres existants. Elles forment un corps cohérent qu'aucun cadre standard de la *data literacy* ne couvre.

Lire le flux du réel derrière la donnée. Devant tout jeu de données, identifier le processus du réel qui a été annoté. Quel phénomène a été observé ? À quelle échelle ? Avec quelle fréquence ? Cette compétence préalable conditionne toute interprétation correcte.

Reconnaître un flux d'observation. Savoir distinguer un prélèvement isolé d'un flux d'observation entretenu, et, devant un flux d'observation, comprendre qu'il s'agit d'une construction et non d'une fenêtre transparente sur le réel.

Remonter à l'intention et aux critères. Devant tout flux d'observation, reconstituer l'intention qui l'a fait naître et les critères qui en découlent — quoi, à quelle cadence, avec quelle fidélité, quel corpus, quel périmètre.

Juger la fidélité relative. Comprendre qu'un flux d'observation est fidèle relativement à ses critères, jamais absolument ; ne pas lui reprocher ses limites, mais évaluer si ses critères conviennent à l'usage visé.

Critiquer la pertinence des critères. La fidélité d'un flux d'observation à ses critères n'est pas l'assurance de sa pertinence pour un usage donné. Savoir interroger : *les critères qui ont défini ce flux sont-ils adaptés à la question que je me pose ?* — question différente, et souvent plus difficile.

Identifier le non-prélevé. Devant toute base de données, poser la question symétrique : qu'a-t-on choisi de ne pas mesurer ? Quels flux n'ont pas été observés ? Qui a écarté quoi, selon quels critères ?

Reconstituer le corpus nécessaire. Identifier l'ensemble des autres données qui donnent sens à une donnée isolée — unités, normes, références, métadonnées. Comprendre que la donnée seule est incomplète, et que sa valeur dépend de la densité du corpus disponible.

Évaluer la fraîcheur, distinguer les trois âges. Estimer l'écart temporel entre la donnée et le flux dont elle provient, et juger si cet écart est compatible avec l'usage envisagé. Savoir si une donnée est utilisable de manière opérationnelle, statistique ou seulement mémorielle.

Identifier les sources stables. Reconnaître, parmi les flux d'observation, ceux dont la péremption est si lente qu'elle peut être tenue, à l'échelle d'usage, pour négligeable — grammaire d'une langue, théorèmes, grandes régularités physiques. Ces sources fournissent les terrains de fondation sur lesquels une connaissance ou un édifice peuvent s'appuyer dans la durée. Les confondre avec des sources à péremption rapide, ou l'inverse, est l'erreur à éviter : chaque type appelle une posture distincte.

Situer l'optimum observationnel. Savoir poser, pour un flux d'observation

donné, la question de l'équilibre entre bénéfice et coût ; reconnaître le sous-prélèvement et la sur-ingénierie ; chercher le point où l'on est en situation optimale d'observation.

Penser le geste d'annotation. Devant toute donnée, savoir qui l'a annotée, avec quel instrument, selon quelle grille, dans quelle intention, avec quels biais propres à cette classe d'annotateur.

Distinguer la mise en mouvement et le rapport au présent. Comprendre que la fraîcheur conditionne le rapport d'un flux d'observation au réel actuel, mais non sa capacité à être rejoué fidèlement comme représentation de son époque.

Reconnaître l'illusion de continuité. Savoir que le flux d'observation produit une représentation fonctionnellement continue du réel, mais non une continuité réelle. Ne pas confondre le film avec le monde, même quand le film est de haute qualité.

Penser la boucle. Savoir reconnaître quand un flux d'observation est inséré dans un circuit fermé, et mesurer que, dans ce cas, le choix des critères engage l'avenir du réel.

Anticiper le saut systémique. Garder présente à l'esprit la possibilité qu'une amélioration des moyens produise un changement de nature, et non une simple amélioration graduelle.

Comprendre la dynamique inter-flux. Savoir qu'un flux d'observation peut être dérivé, transmis, retraité par d'autres entités selon leurs propres critères ; que la qualité d'un flux dérivé dépend de la chaîne complète des prélèvements amont ; que de cette circulation naît une intelligence distribuée dont la richesse dépend de la pluralité des intentions qui alimentent les canaux.

Reconnaître la nature et l'âge des corpus qui alimentent un modèle d'IA. Devant tout résultat produit par un système d'IA, savoir poser les questions : sur quels flux d'observation ce modèle a-t-il été entraîné ? quelle est la fraîcheur du corpus ? quelles intentions ont creusé les canaux dont ce corpus provient ? la question posée porte-t-elle sur un domaine à péremption lente, où la mémoire du modèle reste valide, ou sur un domaine à péremption rapide, où il faudrait un accès direct au flux entretenu ? Cette compétence est la version contemporaine, appliquée aux systèmes d'IA, de la lecture critique des flux d'observation.

La finalité : un individu en situation d'observer

La finalité de tout ce parcours rejoint celle que le deuxième document assignait à la littératie de la donnée. Mais le concept de flux d'observation la précise : être *data-literate*, au sens plein, ce n'est pas seulement savoir manipuler des données. C'est savoir se mettre, soi-même, en **situation d'observer** : comprendre que toute donnée vient d'un canal qu'une intention a creusé, savoir creuser ses propres canaux, les régler

à l'optimum, les rejouer pour comprendre le passé et les extrapoler pour entrevoir l'avenir, lire les canaux que d'autres ont creusés sans les confondre avec le réel, savoir que ses observations, quand elles guident l'action, transforment le monde qu'elles observent.

C'est une compétence technique. C'est aussi une compétence civique : un citoyen qui sait cela ne se laisse pas imposer un flux d'observation comme "la réalité". Devant tout chiffre présenté comme un fait, il sait poser les questions de fond : *quel flux d'observation alimente ce chiffre ? qui en a défini les critères ? que ne montre-t-il pas ? à quelle fréquence est-il rafraîchi ?* Ces questions sont politiques autant que techniques. Elles sont le cœur d'une démocratie informée.

Et c'est, peut-être, une compétence existentielle. L'humain ne possède pas le réel et ne le possédera jamais — le réel s'écoule, immense et indifférent. Mais l'humain peut, par son intention, en creuser des canaux : des répliques partielles, fidèles à ce qu'il a choisi d'y voir, qu'il peut rejouer, partager, mettre au service de ses décisions. Il ne sera jamais le propriétaire des eaux. Il peut être le constructeur, lucide et responsable, de ses canaux.

Conclusion — L'art du canal

Le flux du réel est un fleuve immense, qui s'écoule sans nous, sans fin, sans retour. Les documents précédents avaient appris à le regarder, et à comprendre que nos données n'en sont que des prélèvements — des seaux d'eau tirés du courant, qui se périment dès qu'on les pose sur la berge.

Ce document a montré autre chose. L'humain ne se contente pas de puiser des seaux. Il **creuse des canaux**. Par son intention, il détourne du fleuve une part d'eau, il lui donne une largeur, un débit, un tracé. Le canal est infiniment plus pauvre que le fleuve ; il ne contient que ce que le tracé a admis. Mais le canal, lui, l'humain le maîtrise. Il peut le suivre, le remonter, le rejouer, le relier à d'autres canaux, le confier à un voisin, le combler quand il n'en a plus l'usage. Le canal est l'œuvre humaine par excellence : une portion de réel rendue maniable.

Tout l'art de la donnée tient dans la qualité de ce creusement. Un canal mal tracé — trop étroit, et l'on manque le signal ; trop large, et l'on se noie dans une eau inutile — dessert son constructeur. Un canal bien tracé, réglé à l'optimum, est l'instrument le plus puissant dont l'humain dispose pour comprendre le passé, entrevoir l'avenir, et décider — en sachant que ses décisions, en retour, modifieront le fleuve.

Il restera toujours, au-delà de tous les canaux, le fleuve lui-même : démesuré, continu, indifférent, inappropriable. C'est une leçon de modestie. Mais c'est aussi une leçon de liberté : puisque le fleuve ne se possède pas, personne ne peut prétendre le posséder. Il n'y a que des

canaux, creusés par des intentions, que chacun peut apprendre à tracer, à lire et à partager.

Reste ce que ce savoir implique quand on pose la question à l'échelle du temps long — c'est l'objet du document *Les deux temps de la donnée* : qu'est-ce qu'observer sur la durée, que reste-t-il d'un flux d'observation quand les eaux ont depuis longtemps changé ?

« *On ne se baigne jamais deux fois dans le même fleuve.* » Mais on peut apprendre à en creuser des canaux.

Repères

Pour les références mobilisées dans ce document, voir la *Bibliographie raisonnée*, notamment les notes sur le critère de Nyquist-Shannon, sur le *dataïsme* (van Dijck) et sur l'argument du *N=All* (Mayer-Schönberger & Cukier). Le statut de méthode (fond/forme, statut des illustrations) est posé dans la *Préface méthodologique*.

La vague et l'océan

L'humain dans le flux — Conclusion de l'essai

« Nous sommes tissés de l'étoffe dont sont faits les rêves. »

— William Shakespeare, *La Tempête*

Préambule — Refermer l'arche

Cet essai a commencé par une entaille dans un os, voici plusieurs dizaines de millénaires, et il s'achève ici. Entre les deux, il aura tenté de comprendre ce qu'est une donnée.

Le document *La trace et le lecteur* s'est achevé sur une image : une vague qui voudrait posséder l'océan dont elle est faite. Et sur une promesse — que le dernier document dirait ce que cette image engage, pour l'humain et pour la place qu'il accorde à la donnée dans sa vie collective. Ce document tient cette promesse. Il occupe une position particulière : il est le seul de l'essai à assumer pleinement une part de méditation. Les six précédents construisaient un concept ; celui-ci en tire les conséquences pour la condition de l'humain.

Conformément à l'avertissement de méthode (préface), ce document s'en tient au fond. Il ne dira pas au lecteur ce qu'il faut faire. Il s'efforcera de rendre clairement *pensable* la situation dans laquelle il se trouve, en faisant que, mieux voyant, il devienne par là même mieux à même de décider. La seule proactivité que cet essai puisse légitimement viser n'est pas l'obéissance à des recommandations : c'est la lucidité qui rend capable de choisir.

Partie I — La vague

Un être issu du flux

Reprenons l'image, et regardons-la de près, car tout ce document en découle.

Le document *La trace et le lecteur* a établi un fait que rien ne dément : l'humain est un produit du flux du réel. Il n'a pas été déposé dans le monde depuis l'extérieur du monde. Ses sens, sa cognition, sa capacité même de lire les traces et d'inscrire des données — tout cela a été façonné, sur des milliards d'années, par l'évolution, qui est un long écoulement du flux du réel. L'humain est fait de la matière du monde, soumis à ses lois, emporté par son courant.

Une vague est exactement cela. Une vague n'est pas un objet posé sur l'océan. Une vague est de l'eau de l'océan, momentanément dressée. Elle n'a pas de matière propre : ce qui la compose, c'est l'océan lui-même, en un point, pris dans un mouvement. Elle se forme, elle se tient un instant, elle retombe. Et l'eau qui la formait n'a jamais cessé d'être l'océan.

L'humain est à l'océan du réel ce que la vague est à l'océan d'eau : une portion qui se dresse. Cette image n'est pas un ornement. Elle dit une vérité littérale, établie par *La trace et le lecteur* : il n'y a pas d'extériorité de l'humain au flux du réel. Quand l'humain prélève une donnée, ce n'est pas un sujet qui saisit un objet ; c'est le flux du réel qui, en l'un de ses points, prend acte de lui-même.

La vague qui se dresse : le détachement

Et pourtant, quelque chose, dans la vague humaine, la distingue de toutes les autres.

Une vague d'eau ne sait pas qu'elle est une vague. Elle se dresse et retombe sans rien savoir de l'océan, sans rien savoir d'elle-même. La vague humaine, elle, *sait*. Elle se dresse — et, en se dressant, elle se retourne sur l'océan dont elle est faite, et elle le lit.

Ce retournement a une conséquence vertigineuse. En se retournant sur l'océan, la vague se distingue de lui. Non pas qu'elle en sorte — elle ne peut pas en sortir, elle est faite de son eau — mais elle s'en *éprouve* distincte. Elle se vit comme un point de vue sur l'océan, donc comme autre que l'étendue qu'elle contemple. C'est le **détachement** : non pas une séparation réelle, mais une séparation vécue. L'humain reste une portion du flux ; mais parce qu'il lit le flux, il s'éprouve face à lui.

De ce détachement vécu naît tout ce qui fait la grandeur de l'humain — et tout ce qui fait son vertige. Parce qu'il s'éprouve distinct du flux, l'humain peut le prendre pour objet : l'observer, le mesurer, le découper en données, en construire des flux d'observation. La science, la technique, cet essai même, sont les fruits du détachement.

Mais le détachement a une pente. Celui qui s'éprouve distinct du flux en vient, presque insensiblement, à s'en croire le maître. De « je lis le flux », on glisse à « je domine le flux » ; de « j'en prélève des données », à « ces données sont à moi » ; de « j'en construis des canaux », à « je gouverne les eaux ». La vague, à force de se dresser, finit par oublier qu'elle est l'océan. C'est le détachement qui devient illusion — et c'est le cœur du vertige que ce document doit nommer.

La prétention : posséder, recréer, canaliser, détruire, reconstruire

La trace et le lecteur avait décrit cette prétention d'un trait : un être issu du flux qui ambitionne de le posséder, de le recréer, de le canaliser, de le détruire et de le reconstruire. Reprenons brièvement ces cinq verbes.

Posséder. L'humain veut faire du flux, ou de ce qu'il en prélève, une propriété. Mais on ne possède jamais le flux du réel. On possède, à la rigueur, des canaux qu'on en a creusés — des flux d'observation, des prélèvements inscrits. Le fleuve, lui, n'appartient à personne.

Recréer. L'humain veut reproduire le flux, en fabriquer des doubles maniables — exactement ce que sont les flux d'observation. Cette recreation est légitime et féconde, à une condition : se souvenir que le double n'est jamais l'original.

Canaliser. L'humain veut orienter le flux, le contraindre, le diriger. C'est un art réel — celui-là même que *Creuser le canal* a célébré. Mais canaliser une portion n'est pas maîtriser le tout.

Détruire. L'humain a acquis le pouvoir de modifier le flux du réel lui-même. Cette boucle peut réparer ; elle peut aussi ravager. L'humain est devenu une force du flux, capable d'infléchir le flux — une puissance sans précédent, et aveugle si elle n'est pas lucide.

Reconstruire. Ayant défait, l'humain veut refaire. Ambition noble, mais qui doit savoir sa limite : le flux du réel est irréversible. On ne reconstruit jamais l'état d'avant ; on construit un état nouveau, qui portera la trace de la destruction et de la réparation.

Ces cinq verbes dessinent la condition de l'humain contemporain face au flux : une puissance immense, et une méprise possible sur la nature de cette puissance. L'humain peut beaucoup. Mais il ne peut beaucoup que sur des *canaux*. Le flux du réel, lui, le déborde, le porte, et l'emportera. La vague peut façonner l'écume ; elle ne commande pas la marée.

Grandeur et vertige

Dire que l'humain est une vague qui se prend pour le maître de l'océan, ce n'est pas le rabaisser. C'est décrire, exactement, ce qui fait à la fois sa grandeur et son vertige — et les deux sont inséparables.

La **grandeur** : que le flux du réel ait produit un être capable de se retourner sur lui et de le lire est un fait prodigieux. Pendant des milliards d'années, le flux s'est écoulé. Puis, en un point, il a produit la vague humaine — des êtres capables d'en lire les traces et, à leur tour, d'inscrire des traces voulues — et cette vague a commencé à le déchiffrer.

Le **vertige** : cette même vague, parce qu'elle se dresse, peut oublier qu'elle est l'océan. Elle peut prendre son détachement vécu pour une séparation réelle. Elle peut croire qu'elle possède ce dont elle est faite, qu'elle commande ce qui la porte, qu'elle survivra à ce qui l'emportera.

La lucidité, dont ce document fera l'enjeu central, n'est rien d'autre que ceci : tenir ensemble la grandeur et le vertige. Savoir que l'on est une vague capable de lire l'océan — c'est la grandeur, et il ne faut pas la nier par fausse modestie. Et savoir que l'on n'est qu'une vague — c'est la mesure, et il ne faut pas l'oublier par orgueil.

Partie II — Le torrent

Un moment singulier

La vague humaine a toujours lu le flux. Mais elle vit aujourd'hui un moment singulier de cette lecture.

Jamais, dans les dizaines de millénaires parcourus depuis les premières entailles, le rythme de l'annotation n'a connu pareille accélération. Ce que l'essai a appelé le numérique — une combinaison sans précédent de cadence dans les prélèvements, d'étendue des flux couverts, de précision de chaque mesure, et de vitesse de circulation et de traitement, qui rend possible, pour une vaste classe de flux du réel, de tenir des canaux à jour à mesure même que le réel se transforme — a fait passer la vague d'un régime à un autre. Elle ne grave plus des entailles ; elle entretient en permanence des centaines de millions de flux d'observation.

Nous nommerons ce moment le **torrent**. Le mot n'est pas neutre. Un torrent, c'est de l'eau vive — une richesse, une force. Mais un torrent, c'est aussi ce qui peut emporter, noyer, échapper. Le torrent de données est les deux à la fois : une puissance de lecture du réel sans précédent, et un débit qui peut submerger la vague qui l'a déclenché.

Ce document ne porte pas de jugement de valeur sur le torrent. Il dit seulement, sur le fond, une chose : le torrent est un moment où la question de la lucidité, qui s'est posée à la vague humaine depuis toujours, se pose avec une acuité nouvelle. Plus la vague lit vite et lit beaucoup, plus il importe qu'elle sache *ce qu'elle fait en lisant*.

Les écueils, pointés et non décrits

L'essai, tout au long de son parcours, a fait surgir un certain nombre d'écueils. Les voici rassemblés, nommés, et laissés à la vigilance du lecteur. Conformément à l'avertissement de méthode, le concept peut *pointer* ; il ne dresse pas la carte détaillée.

L'écueil de la confusion entre la donnée et le réel. Tout l'essai a travaillé à distinguer le canal du fleuve, le prélèvement du flux, la trace de la donnée. L'écueil est de l'oublier — de prendre un flux d'observation pour le réel lui-même. Le torrent, par sa richesse, rend cette confusion plus tentante.

L'écueil de la donnée périmée traitée comme fraîche. Une donnée vieillit dans le temps du réel sans bouger dans le temps de l'inscription. Le torrent, qui conserve tout, met à portée de main des données dont rien, à leur surface, ne dit depuis combien de temps le réel les a quittées.

L'écueil de la sur-ingénierie. Prélever bien au-delà de l'optimum observationnel, accumuler des données dont le coût excède de loin le sens qu'on en tire. Le torrent y est exposé par nature : la facilité technique incite à prélever sans mesure.

L'écueil du non-prélevé oublié. Le torrent, par son abondance, donne l'illusion de l'exhaustivité — et c'est par là qu'il rend le non-prélevé plus invisible encore. Croire que ce qui n'est pas mesuré ne compte pas.

L'écueil de l'observation qui se croit neutre. Observer pour décider, c'est engager une transformation du réel. Le torrent, parce qu'il alimente d'innombrables décisions automatiques, multiplie les boucles — et donc multiplie les transformations du réel opérées sans que personne les ait pleinement voulues.

L'écueil du lecteur dessaisi. L'essai a placé le lecteur au centre. L'écueil est qu'il se dessaisisse : qu'il délègue la lecture au point de ne plus savoir lire. La délégation, qui est légitime, devient écueil quand elle devient renoncement.

Ces six écueils ne sont pas une liste de fautes à corriger. Ce sont des points de vigilance, valables quelles que soient les époques et les techniques. L'essai les pointe. Ce que le lecteur en fera, dans son temps et avec ses moyens, lui appartient.

Partie III — La lucidité

Ce que l'essai rend au lecteur

Que reste-t-il, au terme de ce parcours, entre les mains du lecteur ?

Pas une méthode : l'essai n'a pas prescrit de procédures. Pas des consignes : il s'est interdit le registre de l'exhortation. Pas des solutions : la forme, il l'a laissée à d'autres et à d'autres temps.

Ce qu'il rend au lecteur est autre chose, et c'est plus précieux qu'une méthode — parce que cela ne se périme pas. Il lui rend un **regard**. Une manière de voir la donnée, le flux, le torrent, qui ne dépend d'aucune technique et ne vieillira avec aucune.

Cette lucidité tient en quelques propositions, qu'il faut ici rassembler une dernière fois, non comme un résumé, mais comme ce que le lecteur emporte :

Il n'y a pas de données dans le monde — il y a du réel qui s'écoule, et des actes qui en prélèvent. Toute donnée a été produite par quelqu'un, dans une intention, selon des critères. Aucune n'est neutre, aucune n'est brute, aucune n'est le réel.

Toute donnée vient d'un flux et y reste suspendue. Elle naît datée, elle vieillit à mesure que le réel s'éloigne d'elle, sa valeur dépend de sa fraîcheur, du corpus qui l'entoure, de la probité de l'acte qui l'a produite.

L'humain ne possède pas le flux du réel — il en creuse des canaux. Ce qu'il maîtrise, ce sont des flux d'observation : des répliques partielles, fidèles à leurs seuls critères, maniables, périssables.

Observer n'est jamais innocent. Choisir des critères, c'est choisir ce qu'on verra et ce qu'on ne verra pas ; et quand l'observation guide la décision, c'est déjà choisir ce que le réel deviendra.

Le lecteur est au centre, et il est issu du flux. La donnée n'existe que par l'acte d'un être que le flux du réel a produit pour le lire. Cet être est partiel, situé, mortel — et c'est pourtant en lui, et nulle part ailleurs, que le réel se déchiffre.

Voilà la lucidité. Elle ne dit pas quoi faire. Elle dit comment voir.

La lucidité comme condition de la liberté

Pourquoi ce regard importe-t-il ? Pourquoi la lucidité, et non l'ignorance heureuse ?

Parce que la lucidité est la condition de la liberté. C'est la dernière thèse de fond de cet essai, et elle demande d'être énoncée clairement.

Une vague qui ignore qu'elle est une vague n'est pas libre : elle est entièrement portée par le mouvement qui la soulève, sans aucune prise sur lui. Une vague qui sait qu'elle est une vague n'échappe pas pour autant à l'océan — elle ne le peut pas, elle en est faite — mais quelque chose, en elle, a changé : elle peut prendre acte de sa condition, et, à l'intérieur de cette condition, choisir.

Il en va de même pour l'humain face au torrent. L'humain ne sortira pas du torrent : il en est fait, il en vit, il y est pris. La lucidité ne lui offre aucune échappée. Mais elle lui offre, à l'intérieur du torrent, la différence entre être emporté et naviguer. Celui qui sait que la donnée est un prélèvement daté ne lui obéit plus aveuglément. Celui qui sait que tout flux d'observation est partiel cherche ce qu'on ne lui montre pas. Celui qui sait que l'observation transforme le réel mesure la portée de ce qu'il déclenche. Celui qui sait qu'il est le lecteur ne se dessaisit pas de la lecture.

C'est en ce sens, et en ce sens seulement, que cet essai a voulu rendre le lecteur **proactif**. Non pas en lui dictant des actes, mais en lui rendant le regard qui fait qu'un acte peut être choisi plutôt que subi. La littératie de la donnée, dont le deuxième document cherchait la juste définition, trouve peut-être ici sa formule la plus simple : *être data-literate, c'est avoir reconquis assez de lucidité sur la donnée pour rester libre dans un monde qui en est fait.*

La responsabilité du lecteur

La lucidité ne va pas sans une chose qu'elle entraîne nécessairement avec elle : la responsabilité.

Tant que l'humain croyait la donnée neutre, objective, simplement trouvée dans le monde, il pouvait se sentir irresponsable de ce qu'elle disait : la donnée parlait, il se contentait de l'écouter. L'essai a défait cette croyance. Si la donnée est un prélèvement opéré par un acte intentionnel, alors quelqu'un a choisi : choisi de prélever ceci et non cela, à cette cadence, selon ces critères, pour cet usage. Et qui choisit, répond de son choix.

Cette responsabilité ne se délègue pas entièrement, même quand la lecture est déléguée. Le lecteur du torrent contemporain ne peut pas tout lire lui-même — il doit s'en remettre à des dispositifs, à des instruments, à d'autres lecteurs. Cette délégation est inévitable et légitime. Mais il lui reste la responsabilité de savoir *qu'il délègue*, de ne pas confondre ce qu'on lui présente avec le réel, de garder vivante la question — d'où vient cette donnée ? quel canal l'a produite ? que ne me montre-t-elle pas ? depuis quand le réel l'a-t-il quittée ?

Et chacun, dans le torrent, est aussi un producteur de données. Chacun, par sa seule existence numérique, alimente des flux d'observation qu'il n'a pas creusés. L'essai n'a pas dit ce qu'il fallait en faire — c'eût été de la forme. Mais il a rendu une chose claire, et qui est de fond : ces données ne sont pas tombées du ciel, elles ont été prélevées, et tout prélèvement procède d'une intention qui n'est pas nécessairement celle de la personne sur laquelle il porte. Savoir cela, c'est déjà ne plus être tout à fait un objet du torrent. C'est commencer à en être un sujet.

Pourquoi le fond, et pas la forme

Une dernière chose mérite d'être dite. La préface méthodologique a posé d'emblée une règle : cet essai énonce ce qui est vrai de la donnée indépendamment des techniques qui la produisent, et s'interdit de prescrire la forme concrète que prennent les choses dans le temps présent du lecteur. Cette règle a été annoncée comme un parti pris ; elle a été tenue tout au long du parcours. Mais la raison profonde n'en a pas été dite. Maintenant que l'essai s'achève, elle peut l'être.

Cette raison n'est pas une préférence d'auteur. C'est une nécessité interne du sujet.

Regardons la littératie de la donnée telle qu'elle est aujourd'hui enseignée, dans les cadres institutionnels et les manuels qui en font la littérature. Le deuxième document a montré qu'elle se présente, pour l'essentiel, comme une discipline de la **forme** : elle décrit des outils, des plateformes, des cadres de compétences, des certifications, des architectures techniques, des bonnes pratiques articulées à des technologies datées. Ce n'est pas un reproche — c'est une description. La littératie appliquée a sa nécessité ; elle équipe le lecteur de moyens concrets pour aujourd'hui.

Mais cette littérature subit, à l'échelle institutionnelle, exactement ce que l'essai a thématiqué à l'échelle de la donnée : elle se périme. Les outils évoluent plus vite que les manuels qui les enseignent. Les plateformes changent de version en quelques mois, là où il faut plusieurs années à un cadre institutionnel pour intégrer ces changements. Les compétences techniques apprises dans une formation sont, pour une part, déjà à actualiser au moment où l'apprenant en sort. Le savoir sur la donnée subit la péremption qu'il ne sait pas penser.

La boucle est saisissante. La littérature qui décrit la littératie de la donnée vieillit dans le temps du réel sans bouger dans le temps de l'inscription — exactement comme la donnée du chapitre 5. Elle reste exacte en tant que constat daté (« voici comment on enseignait la littératie de la donnée en telle année »), mais elle perd sa capacité représentative à mesure que les outils qu'elle décrit sont remplacés par d'autres. Et ce vieillissement est d'autant plus rapide que la matière qu'elle traite — les technologies du numérique — évolue à un rythme que les cycles d'écriture, de relecture, de publication, de diffusion ne peuvent suivre.

Un essai sur la donnée qui se câblerait, lui aussi, sur la forme reproduirait fatalement ce vieillissement. Il décrirait, en quelques chapitres, l'état d'un art qui sera dépassé avant la fin de sa première lecture publique. Il ferait, sur la péremption, un livre périssable. Ce serait une contradiction performative — dire que la donnée se périme avec un texte qui se périme.

C'est pourquoi la règle fond/forme posée en préface n'était pas un choix esthétique. C'était la **seule manière** dont un essai sur la péremption de la donnée pouvait éviter de se périmer lui-même. Pour traiter de ce sujet en particulier, il fallait s'inscrire dans le temps de ce qui ne se périme pas

avec les outils : le temps des concepts, le temps des structures, le temps du fond.

Cela ne dévalue pas la littérature appliquée. Les deux registres sont complémentaires, à la manière dont les trois âges de la donnée sont complémentaires. La littérature appliquée est dans le **premier âge** : elle équipe pour le présent, elle est opérationnelle, elle se périme à mesure que les outils changent — c'est sa fonction. Cet essai a tenté de se tenir dans un registre qui serait au fond ce que le **deuxième et le troisième âge** sont à la donnée : un savoir qui mûrit lentement, qui informe les décisions de longue portée, qui se transmet au-delà de la génération technique qui l'a vu naître. Les deux sont nécessaires ; ils n'ont simplement pas la même durée de validité.

La forme finit toujours par se périmer. Le fond, lui, vieillit beaucoup plus lentement — au rythme où change la condition de l'humain comme lecteur des traces que le réel laisse de lui-même. C'est-à-dire à un rythme qui se mesure non en années, mais en transformations anthropologiques. Tant que l'humain reste un être qui prélève, qui annote, qui creuse des canaux dans le fleuve du réel, ce que l'essai a énoncé sur la donnée demeurera vrai. Le jour où l'humain cessera d'être cela, l'essai aura, lui aussi, vieilli — mais alors d'autres questions, et d'autres essais, occuperont la place.

Cet essai n'est donc pas seulement *sur* la préemption de la donnée. Il est, jusque dans sa méthode, *cohérent avec* ce qu'il dit de la donnée. La forme qu'il prend est une conséquence directe de la thèse qu'il défend. C'est le seul mérite qu'il revendique — et le seul qu'il pouvait revendiquer sans se contredire.

Conclusion générale de l'essai — Ce que sait la vague

Cet essai aura tenu, d'un bout à l'autre, une seule thèse, déclinée en sept moments. La donnée n'est pas un objet que le monde nous tend. Elle est un acte : le prélèvement, par un lecteur que le flux du réel a produit, d'une trace que le flux du réel a laissée de lui-même. Tout le reste — la préemption, le corpus, le flux d'observation, l'optimum, les deux temps, les deux piliers — découle de cette thèse unique, comme les branches d'un seul tronc.

De cette thèse, l'essai a tiré une image ; regardons-la, pour finir, une dernière fois. L'humain est une vague de l'océan du réel. Il en est fait ; il y retournera ; il ne le quittera jamais. Mais c'est une vague qui, en se dressant, s'est retournée sur l'océan et a appris à le lire. Cette lecture est sa grandeur. L'oubli de sa condition de vague est son vertige. Et entre la grandeur et le vertige, il y a un fil étroit, qui est la lucidité : savoir, en même temps, qu'on lit l'océan et qu'on n'est qu'une vague.

L'humain n'est pas propriétaire des eaux. Il n'est pas non plus, en toute

rigueur, propriétaire des canaux qu'il creuse — ces canaux dépendent du fleuve qui les alimente. Mais il est, plus modestement et plus durablement, le **gardien des sources** : ces points où le réel, par la lenteur de sa transformation, lui offre un terrain de fondation. Identifier les sources, les protéger, les transmettre : c'est, sur le long terme, la part la plus humble et la plus précieuse de sa responsabilité de lecteur.

À l'heure du torrent, ce fil est plus étroit que jamais, et plus nécessaire. L'humain peut, dans le torrent, se laisser emporter — prendre ses canaux pour le fleuve, ses données périmées pour le présent, son abondance pour de l'exhaustivité, son observation pour de l'innocence. Ou bien il peut, dans ce même torrent, rester lucide — et la lucidité ne le sortira pas du torrent, mais elle fera de lui, à l'intérieur du torrent, un être qui navigue au lieu d'être charrié.

Cet essai n'aura rien prescrit. Il aura tenté de rendre au lecteur le regard. De faire que la donnée, le flux, le torrent, deviennent pour lui *pensables* — et qu'ainsi, mieux voyant, il devienne plus libre, et plus responsable de la part du monde qui passe entre ses mains.

Une vague qui sait qu'elle est une vague est déjà, par ce seul savoir, autre chose qu'une vague : elle est le point où l'océan, un instant, se regarde.

Ce que sait la vague ne la sauve pas de l'océan. Mais c'est ce savoir, et lui seul, qui fait d'elle autre chose qu'une eau qui monte et retombe. C'est ce savoir que cet essai aura voulu, modestement, raviver.

« *Nous sommes tissés de l'étoffe dont sont faits les rêves.* » Et de l'étoffe dont est fait le réel. La vague est l'océan qui, un instant, se dresse et se regarde. Cet essai aura été une goutte de ce regard.

Repères conceptuels

Pour les références mobilisées dans l'ensemble du parcours, voir la *Bibliographie raisonnée*. Les notions convoquées dans ce document — l'ontologie processuelle, le flux d'observation, l'optimum observationnel, les deux temps de la donnée, la trace et le lecteur — sont développées dans les six documents précédents.

Statut des illustrations et de l'image directrice : la vague et l'océan, le torrent, le canal et le fleuve sont des images destinées à rendre le concept intuitif et mémorable. Conformément à la préface méthodologique, elles relèvent de la forme du propos ; seul le fond qu'elles servent engage l'essai. De même, la description du « moment du torrent » vaut comme caractérisation générale d'une époque, non comme analyse datée : ce qui est de fond, c'est que l'accélération de l'annotation pose la question de la lucidité avec une acuité accrue ; les formes concrètes de cette accélération appartiennent au temps du lecteur.

Le réseau d'images en bref — Pour fixer en quelques traits le réseau métaphorique de l'essai :

- le **flux du réel** est le fleuve — ce qui s'écoule sans nous, immense, indifférent, inappropriable ;
- les **prélèvements** sont des seaux d'eau tirés du courant — figés dès qu'on les pose sur la berge ;
- les **flux d'observation** sont des canaux creusés dans le fleuve — portions d'eau détournée, maniables, entretenues par une intention ;
- les **sources stables** sont les points où le fleuve coule si lentement qu'on peut y bâtir des édifices durables ;
- la **vague** est la place de l'humain dans le tout — faite d'eau de l'océan, elle se dresse, le déchiffre, et y retourne ;
- le **torrent** est le moment contemporain — un débit jamais vu, qui peut emporter qui ne sait pas naviguer.

Ainsi s'achève l'essai **Le fleuve et le canal — La littérature de la donnée à l'épreuve du flux***, en sept documents et une préface méthodologique. Conçu et dirigé par David Bories, rédigé en utilisant l'intelligence artificielle générative.*

Bibliographie raisonnée

Appareil de références de l'essai

Cette bibliographie réunit les références appelées par les sept documents de l'essai. Elle est organisée pour rester utile dans la durée : d'abord les ouvrages de fond, dont la pertinence ne se périmera pas ; ensuite les pistes pour suivre l'évolution du champ par soi-même ; enfin des termes de recherche durables, accessibles quels que soient les portails et les outils de demain.

Ce qui ne changera pas

Ces ouvrages constituent un socle de référence stable, à travers les sept documents de l'essai.

Aux fondations philosophiques

Bergson, Henri — *Essai sur les données immédiates de la conscience*, Félix Alcan, 1889 ; *Matière et mémoire*, 1896 ; *L'Évolution créatrice*, 1907. De tous les philosophes du processus, Bergson est celui dont la pensée s'applique le plus directement au statut de la donnée — son titre même de 1889 contient le mot. La distinction de la durée pure et du temps spatialisé est le soubassement direct de la thèse des deux temps de la donnée (document 5).

Whitehead, Alfred North — *Process and Reality: An Essay in Cosmology*, The Macmillan Company, 1929 (Gifford Lectures 1927-1928 ; édition corrigée par D. Griffin et D. Sherburne, The Free Press, 1978). La métaphysique du processus la plus aboutie. La précision philosophique du mot *datum* y est centrale : ce qui est donné à une occasion ultérieure par les occasions qui la précèdent. Pour une introduction accessible : *Process Philosophy*, *Stanford Encyclopedia of Philosophy*.

Peirce, Charles Sanders — La sémiotique de Peirce, et notamment sa distinction du signe causé (l'indice) et du signe conventionnel, est le soubassement de la distinction trace/donnée développée au document 6. La fumée à l'égard du feu, l'empreinte à l'égard du pas : l'indice est un signe lié physiquement à son objet, non par convention.

Cassin, Barbara (dir.) — *Vocabulaire européen des philosophies — Dictionnaire des intraduisibles*, Seuil/Le Robert, 2004. L'entrée *datum* retrace la généalogie philosophique du concept de donnée. Un outil de fond.

Deleuze, Gilles — *Différence et répétition*, PUF, 1968. Pour la pensée du devenir contre la représentation.

Simondon, Gilbert — *L'Individuation à la lumière des notions de forme et d'information*, Jérôme Millon, 2005. Pour la pensée de l'individuation comme processus, et la conception de l'information comme différentiel plutôt qu'objet.

Jullien, François — *Du temps. Éléments d'une philosophie du vivre*, Grasset, 2001. La différence entre la pensée occidentale du temps-objet et la pensée chinoise du temps-propension — un contrepoint philosophique utile à la thèse processuelle.

Floridi, Luciano — *The Philosophy of Information*, Oxford University Press, 2011. La définition la plus aboutie de la position que cet essai renverse : la donnée y est fondée comme *diaphora*, une différence, un manque d'uniformité dans le réel — c'est-à-dire, en dernière analyse, comme quelque chose qui *est là*, fût-ce relationnellement, avant tout geste. L'essai soutient l'inverse : la différence est dans le flux, mais la donnée ne naît qu'au prélèvement qui l'inscrit. Lire Floridi, c'est mesurer ce que l'ontologie objectale a de plus solide — et où passe exactement la ligne de partage.

Histoire et anthropologie de la donnée

Schmandt-Besserat, Denise — *Avant l'écriture, la mémoire des choses*, Le Pommier, 2008. Pour l'origine des systèmes de comptage et leur lien avec la naissance de l'écriture. L'ouvrage de référence, accessible à un lectorat non spécialisé.

Goody, Jack — *La Raison graphique*, Minuit, 1979 ; *The Logic of Writing and the Organization of Society*, Cambridge University Press, 1986. Pour comprendre les effets cognitifs et sociaux de l'écriture, socle de toute réflexion sur la littérature.

Hacking, Ian — *The Taming of Chance*, Cambridge University Press, 1990 (trad. fr. *L'Émergence de la probabilité*, Seuil, 2002). Pour comprendre comment, au XIX^e siècle, la statistique est passée d'outil administratif à mode de pensée dominant.

Desrosières, Alain — *La Politique des grands nombres. Histoire de la raison statistique*, La Découverte, 1993. Pour les liens entre production de données, administration et pouvoir. Une sociologie historique rigoureuse et lisible.

Critique épistémologique de la donnée

Gitelman, Lisa (dir.) — « *Raw Data* » *Is an Oxymoron*, MIT Press, 2013. Neuf études de cas historiques montrant que toute donnée porte la trace de son contexte de production, de ses instruments et de ses catégories classificatoires. Le titre est devenu une formule fondatrice du champ.

Bowker, Geoffrey C. & Star, Susan Leigh — *Sorting Things Out: Classification and Its Consequences*, MIT Press, 1999. Étude fondatrice sur la manière dont les systèmes de classification façonnent les phénomènes qu'ils prétendent décrire. Éclaire le geste d'annotation comme acte structurant (document 4) et la trace comme dépôt situé (document 6). Bowker a par ailleurs contribué à l'anthologie « *Raw Data* » *Is an Oxymoron* dirigée par Gitelman.

Bowker, Geoffrey C. — *Memory Practices in the Sciences*, MIT Press, 2005. Sur les infrastructures de mémorisation scientifique — bases de données, archives, standards — et leur effet rétroactif sur ce qui peut être su. En écho aux deux temps de la donnée (document 5) et au statut des supports (document 5, partie III).

Leonelli, Sabina — *Data-Centric Biology: A Philosophical Study*, University of Chicago Press, 2016. Une approche relationnelle de la donnée, présentée comme alternative à la pyramide DIKW.

Berry, Gérard — *L'Hyperpuissance de l'informatique*, Odile Jacob, 2017. Pour comprendre les fondations techniques de l'informatique contemporaine et les présupposés ontologiques qu'elles portent.

Enjeux contemporains et politiques

Zuboff, Shoshana — *The Age of Surveillance Capitalism*, PublicAffairs, 2019 (trad. fr. *L'Âge du capitalisme de surveillance*, Zulma, 2020). La synthèse la plus documentée sur les enjeux contemporains de la captation massive de données comme modèle économique.

D'Ignazio, Catherine & Klein, Lauren F. — *Data Feminism*, MIT Press, 2020 (accès libre en ligne). Pour la critique du *raw data*, l'attention aux *missing data* et l'inscription des structures de pouvoir dans les données.

O'Neil, Cathy — *Weapons of Math Destruction*, Crown, 2016 ; trad. fr. *Algorithmes : la bombe à retardement*, Les Arènes, 2018. Pour les dommages concrets que des données mal comprises ou délibérément biaisées peuvent produire.

Cardon, Dominique — *Culture numérique*, Presses de Sciences Po, 2019 ; *À quoi rêvent les algorithmes*, Seuil, 2015. Un panorama dense mais accessible des transformations qu'Internet et les données numériques ont produites dans les sociétés.

Mayer-Schönberger, Viktor & Cukier, Kenneth — *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, 2013. Pour une introduction au tournant des données massives, et la formulation explicite de l'idée que la quantité compense la qualité (argument dit du *N=All*), à laquelle l'essai s'oppose.

Ostrom, Elinor — *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, 1990 ; trad. fr. *Gouvernance des biens communs*, De Boeck, 2010. Prix Nobel d'économie 2009. La théorie des communs, développée pour les ressources naturelles, est transposable aux flux comme communs (document 3).

Littératie de la donnée — textes fondateurs

Schiold, Milo — *Information Literacy, Statistical Literacy, Data Literacy*, IASSIST Quarterly, vol. 28, n° 2-3, été-automne 2004. L'acte de naissance le plus souvent cité du concept de *data literacy*.

Mandinach, Ellen B. & Gummer, Edith S. — *Data Literacy for Educators: Making It Count in Teacher Preparation and Practice*, Teachers College Press, 2016. L'ancrage du concept dans la pédagogie.

Ridsdale, Chantel et al. — *Strategies and Best Practices for Data Literacy Education*, Dalhousie University, 2015. La première revue systématique de la littérature sur le sujet, et un cadre conceptuel à quatre domaines (collecte, gestion, évaluation, application).

D'Ignazio, Catherine & Bhargava, Rahul — *Approaches to Building Big Data Literacy*, MIT Center for Civic Media, 2015. La grille des quatre verbes — lire, travailler avec, analyser, argumenter avec — pour une littératie capacitante.

Naur, Peter — *Datalogy, the Science of Data and Data Processes*, IFIP Congress 1968 ; *Computing, a Human Activity*, Addison-Wesley, 1992. Le précurseur non reconnu de la *data literacy*, qui définissait dès 1966 une discipline visant à former les individus à la nature et à l'usage des données.

Verdi, Ugo — *À la recherche d'une maîtrise des données : une étude des discours scientifiques et professionnels sur la data literacy*, thèse de doctorat en sciences de l'information et de la communication, Université Bordeaux Montaigne, 2024. L'état de l'art le plus complet disponible en langue française au moment de la rédaction de l'essai. Accessible via HAL.

Lehmans, Anne (dir.) — *La culture des données : un nouvel enjeu éducatif*, ISTE Editions, 2024. La synthèse francophone la plus récente.

Morrow, Jordan — *Be Data Literate*, Kogan Page, 2021. Une entrée par la dimension professionnelle.

Star, Susan Leigh & Griesemer, James R. — *Institutional Ecology, "Translations" and Boundary Objects, Social Studies of Science*, vol. 19, n° 3, 1989. Le concept d'*objet frontière*, appliqué à la *data literacy* par Verdi (2024).

Ackoff, Russell L. — *From Data to Wisdom, Journal of Applied Systems Analysis*, vol. 16, 1989. La formalisation de la pyramide DIKW.

Ce qui évolue et comment le suivre

Certains éléments cités dans l'essai sont par nature en mouvement permanent. Les portails ci-dessous offrent des points d'entrée stables pour suivre leurs évolutions.

Cadres institutionnels de littératie : Joint Research Centre de la Commission européenne pour le DigComp (versions successives 2.0, 2.1, 2.2, 3.0...) ; portail PIX et Éduscol pour le Cadre de référence des compétences numériques (CRCN) français ; UNESCO *Media and Information Literacy* pour les cadres en littératie médiatique et en IA ; portail CNIL pour le référentiel de formation à la protection des données.

Mesures internationales des compétences : OCDE pour le PIAAC (compétences adultes) et le PISA (élèves de 15 ans) ; Statistique Canada pour *Data Literacy: What It Is and How to Measure It in the Public Service*.

Recherche francophone en littératie de la donnée : portail HAL pour les archives ouvertes ; GTnum DEFI (Données pour l'Éducation, la Formation, l'Innovation), animé depuis 2022 par Anne Lehmanns et Camille Capelle à l'Université de Bordeaux ; revues *InterCDI*, *Hermès*, *Information and Learning Sciences* (Emerald), *Journal of Documentation*, *Journal of Community Informatics*.

Acteurs civiques durables du champ associatif : Open Knowledge Foundation (portail d'entrée stable, hébergeant notamment *School of Data*) ; *DataKind* ; *The Engine Room* ; *Our World in Data* (Oxford) ; *Open Data France*.

Études du déficit de compétences en données : rapports annuels d'éditeurs d'analytique (notamment le *Data Literacy Project* cofondé par Qlik) ; cabinets d'études comme Accenture ou Gartner.

Empreinte environnementale des infrastructures numériques : Agence internationale de l'énergie (AIE) pour les rapports annuels sur la consommation énergétique des datacenters ; *International Data Corporation* (IDC) pour les volumes mondiaux.

Archives primaires de l'histoire de la donnée : *National Archives* (Royaume-Uni, pour le *Domesday Book*) ; *National Archives and Records Administration* (États-Unis) ; *Archives nationales de France* ; *British Museum*, *Pergamonmuseum*, *Metropolitan Museum of Art* pour les tablettes cunéiformes ; *IBM History*, *Bibliothèque du Congrès* (fonds Hollerith), *ACM* (fonds Turing Award).

Pour approfondir par soi-même : termes de recherche durables

Les liens vieillissent, les noms d'institutions changent, les plateformes se renouvellent. Ces termes de recherche, eux, restent valides pour interroger les bases académiques et bibliothéconomiques disponibles à la date où l'on lit.

Histoire préhistorique de la cognition symbolique : *archéologie cognitive*, *Upper Paleolithic symbolic behavior*, *art pariétal datation*, *représentation symbolique préhistoire*.

Histoire de l'écriture et des systèmes d'annotation : *origine de l'écriture*, *proto-cunéiforme*, *calculi mésopotamiens*, *histoire des systèmes numériques*.

Histoire de la statistique et de la mesure : *histoire de la statistique*, *arithmétique politique*, *raison statistique*, *épistémologie de la mesure*.

Histoire et théorie de la littératie de la donnée : *data literacy définition*, *data literacy history*, *datalogie Peter Naur*, *boundary object literacy*, *DIKW model criticism*.

Cadres institutionnels et politiques de littératie : *DigComp competence framework*, *data literacy policy*, *CRCN compétences numériques*, *AI literacy framework*.

Approches critiques : *data feminism*, *datafication critique*, *data activism*, *algorithmic accountability*, *capitalisme de surveillance données*.

Enjeux contemporains de la donnée : *épistémologie de la donnée*, *ontologie processuelle*, *data studies*, *science and technology studies (STS)*, *data freshness*, *data decay*, *raw data myth*.

Pratiques pédagogiques : *data literacy education*, *data storytelling pedagogy*, *arts-based data literacy*, *serious games data*.

Infrastructures numériques et empreinte : *empreinte environnementale datacenters*, *sobriété numérique*, *green IT*, *data center energy consumption*.

Sources archéologiques et historiques citées dans l'essai

Premières représentations symboliques — Les peintures rupestres les plus anciennement datées à ce jour ont été découvertes en Asie du Sud-Est insulaire. Des travaux publiés en 2021 par l'archéologue Maxime Aubert et son équipe (Griffith University, Australie) ont daté une figure animale sur la paroi d'une grotte à Sulawesi (Indonésie) à plus de 45 500 ans, et suggèrent que d'autres représentations de la région pourraient remonter à 60 000 ou 65 000 ans. Référence : Aubert, M. et al., *Oldest cave art found in Sulawesi*, *Science Advances*, 2021.

Os entaillés préhistoriques — Deux artefacts parmi les mieux documentés : l'**os de Lebombo** (entre 42 000 et 43 000 ans, monts Lebombo, Afrique australe), péroné de babouin portant 29 encoches délibérées, interprété comme un bâton de comptage ou un suivi de cycles lunaires (travaux de Peter Beaumont) ; l'**os d'Ishango** (entre 20 000 et 25 000 ans, bassin du Congo), portant trois colonnes d'encoches organisées selon une structure interne complexe dont l'interprétation reste débattue (travaux de Jean de Heinzelin de Braucourt, repris et discutés par Alexander Marshack).

Jetons d'argile et naissance de l'écriture — Le système des *calculi* (jetons d'argile) a été retrouvé sur un territoire s'étendant de l'Anatolie à l'Iran actuel, avec des exemplaires remontant à 9 000 ans. Les enveloppes d'argile scellant ces jetons sont appelées *bullae*. Les travaux de Schmandt-Besserat ont démontré le lien direct entre ce système et l'invention de l'écriture cunéiforme.

Écriture administrative à Uruk — Le proto-cunéiforme apparaît vers 3 300 av. J.-C. à Uruk (Mésopotamie). Plus de 5 000 tablettes issues des couches archéologiques correspondantes ont été exhumées ; leur contenu est presque exclusivement administratif et comptable. L'historienne Eleanor Robson (University College London) a produit des analyses de référence sur ce corpus.

Premiers grands recensements — Le *census* romain (recensement des citoyens et de leurs biens, pratiqué depuis le VI^e siècle av. J.-C.) ; le *Domesday Book* anglais (1086, inventaire exhaustif des terres et ressources du royaume d'Angleterre, commandé par Guillaume le Conquérant).

Comptabilité en partie double — Formalisée par le frère franciscain Luca Pacioli dans son *Summa de arithmetica* (1494), elle est considérée comme l'un des outils intellectuels fondateurs du commerce moderne.

Grandes bibliothèques institutionnelles — Alexandrie (Égypte, III^e siècle av. J.-C.) ; Pergame (Anatolie) ; bibliothèques des temples égyptiens (Tebtunis, Oxyrhynque) ; bibliothèques monastiques bouddhistes (Nalanda, en Inde, fondée au V^e siècle, accueillit jusqu'à dix mille étudiants) ; bibliothèques du monde islamique médiéval (*Bayt al-Hikma*

de Bagdad, IX^e siècle) ; *amoxcalli* aztèques en Mésopotamie. La bibliothèque publique moderne — ouverte à tous sans condition — s'est développée à partir du XIX^e siècle, notamment en Grande-Bretagne (*Public Libraries Act*, 1850).

Premières analyses statistiques de populations — Les journaux astronomiques babyloniens (VIII^e siècle av. J.-C. à l'ère commune), série de relevés célestes et commerciaux couvrant plusieurs siècles, qui ont permis à Hipparque puis Ptolémée de calculer des régularités planétaires fines ; les *Natural and Political Observations upon the Bills of Mortality* de John Graunt (1662), première analyse statistique de données démographiques londonniennes, considérée comme l'acte de naissance de la démographie.

Histoire de la mécanisation du calcul — Machines à calculer mécaniques (Pascal, Leibniz, XVII^e siècle) ; travaux de Charles Babbage sur la machine analytique (1830-1840), premier concept de machine à programme général, commentés par Ada Lovelace ; mécanographie sur cartes perforées (Herman Hollerith, recensement américain de 1890) ; premiers ordinateurs électroniques (ENIAC, 1945) ; modèle relationnel d'Edgar Codd (1970), qui a structuré la gestion des bases de données pour des décennies. Pour la perspective théorique : David Landes, *The Unbound Prometheus*, 1969 ; Lewis Mumford, *Technics and Civilization*, 1934.

Note sur la « fraîcheur » et le « data decay »

La littérature professionnelle a fini par nommer la péremption des données. Quelques jalons : *Data Assets: Information Decay*, *Analytics Magazine* (INFORMS), 2014, pour la métaphore de la demi-vie ; Wang, R. Y. & Strong, D. M., *Beyond Accuracy: What Data Quality Means to Data Consumers*, *Journal of Management Information Systems*, vol. 12, n° 4, 1996, article fondateur qui distingue plusieurs dimensions de la qualité, dont la *timeliness*. Sur les biais algorithmiques produits par des historiques anciens : Angwin, J. et al., *Machine Bias*, ProPublica, 2016 (analyse du système COMPAS de prédiction de récidive aux États-Unis). Sur la formule « *Data is not persistent. It is perishable* », elle circule dans la littérature professionnelle sur la qualité des données depuis le milieu des années 2010, notamment dans les travaux sur la *data freshness* en contexte B2B et médical.

Note sur le critère de Nyquist-Shannon

La notion de seuil en deçà duquel un échantillonnage ne représente plus fidèlement un phénomène variable est un fait établi de la théorie du signal,

connu sous le nom de critère de Nyquist-Shannon. L'essai en retient le principe de fond — l'existence d'un seuil à respecter — sans en mobiliser le formalisme mathématique.

Note sur Van Dijck et la critique du *dataïsme*

La critique du discours qui présente la donnée comme valant intrinsèquement par sa seule masse a été formalisée sous le nom de *dataïsme* par José van Dijck : *Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology*, *Surveillance & Society*, vol. 12, n° 2, 2014.

Document rédigé en mai 2026.

Nous ne vivons pas entourés de données.

—

Un jour de mai 2026, je demandai à l'IA :

«

Produis un documentaire étayé, complet, argumenté, propre, sourcé, sur les origines de la donnée dans l'humanité depuis les origines les plus anciennes qui ont été repérées :

Comment sont apparues les premières données collectées par l'homme, sous quelle forme ? Pour quels usages ? Dans quels buts ?

Puis détaille les évolutions majeures qui ont conduit la civilisation contemporaine à être dépendante des usages de la donnée à travers le numérique et son flot incessant de données.

Comment l'homme a pris conscience de la donnée, de son potentiel ?

Comment l'a-t-il collectée, produite et exploitée au fil des millénaires ? »

L'IA s'exécuta.

Et l'essai commença.

—

Dire que l'auteur, au moment de produire cet essai, ne l'avait pas en tête, c'est ne pas tenir compte de près de 20 ans d'observation de ce qui se trame en sous-bas : au début des câbles, un écran, un clavier. Puis rapidement, du portable, un réseau, du tactile, du mobile. Et maintenant, des applications, des utilisateurs, des codes et une vie numérique omniprésente dont le flot de données charrie autant d'intérêts pour le présent que d'interrogations sur le devenir.

Dans le fond, cet essai dresse le portrait du monde qui bouge.

Sur la forme, il en fige l'image d'une société contemporaine.