

# Mode d'emploi pour un *Atlas Mnémosyne* \* des services de renseignement

Alexandre Papaemmanuel

Enseignant à Sciences Po, président de la commission digital défense de l'AA-IHEDN.

La mémoire a exploré les rives de l'artificialité avant l'intelligence, offrant la capacité pour l'homme de travailler à mémoriser l'ordre du monde en suivant les préceptes du poète grec Simonide de Céos <sup>(1)</sup>. La mémoire, désormais externalisée à la machine, augmente l'homme naviguant dans un atlas numérisé, synonyme de capitalisation, d'archivage et de corrélation automatisée. Ce renversement n'est pas anodin car « si les technologies numériques peuvent libérer un potentiel inexploré de la mémoire par le fait de son extériorisation, elles peuvent également provoquer le contraire, c'est-à-dire bloquer les possibilités scellées dans la mémoire » <sup>(2)</sup>. Puisque l'art de la mémoire n'est plus ce combat « pour exercer les capacités du cerveau », il devient une course au nombre de serveur et d'algorithme pour soutenir l'analyste de renseignement dans son appréhension des rémanences, dans son rapport immersif avec la donnée au service d'une forme de « connaissance traversière » fondée sur la « puissance intrinsèque de montage qui consiste à découvrir [...] des liens que l'observation directe est incapable de discerner » <sup>(3)</sup>.

Or, dans un monde où l'image – l'information comme la donnée – s'avère foisonnante, où le temps court s'impose et où l'immédiateté prime, l'acteur étatique en mesure de connaître pour anticiper pourra forcer le destin en sa faveur. Le renseignement est l'un des outils idoines pour saisir l'urgence du monde, ses soubresauts et ses lignes de rupture, et anticiper ses lignes de fractures futures. Pareille action suppose de pouvoir maîtriser l'agrégation de données de plus en plus massives pour éclairer, prévoir et comprendre afin de garder un temps d'avance.

En effet, l'homme n'est plus capable de traiter cette masse générée par des capteurs toujours plus nombreux et performants. Les crises sont désormais gérées par ceux qui sont en mesure de mieux capter, analyser, exploiter, maîtriser et protéger une somme toujours plus importante de données.

\* Œuvre du critique d'art allemand Aby Warburg.

<sup>(1)</sup> Pacôme Thiellement : *Le Théâtre de Camillo, L'art de la mémoire et sa revanche* ; Vertige, 2008.

<sup>(2)</sup> Bernard Stiegler : « Le numérique empêche-t-il de penser ? », *Esprit*, janvier 2014.

<sup>(3)</sup> Georges Didi-Huberman : *Atlas, ou le gai savoir inquiet* ; Paris, Les Éditions de Minuit, 2011.

Le volume et la variété des ressources conduisent donc les services à faire évoluer leurs méthodes de renseignement impliquant d'importants coûts culturels et humains, techniques et financiers, et juridiques. L'ancien directeur de la DGSI, Patrick Calvar, l'a rappelé le 10 mai 2016 à l'Assemblée nationale : « Nous ne manquons pas de données ni de métadonnées, mais nous manquons de systèmes pour les analyser. [...] La moindre perquisition nous permet de récupérer des milliers de données. Nous avons donc besoin d'outils de *big data* pour répondre immédiatement à nos besoins <sup>(4)</sup>. » Les services de renseignement (SR) doivent faire face aux défis de la « dataïfication » et repenser leur méthode, outillage et obligation face à la donnée pour que l'État puisse saisir son temps.

### **Les loci de mémoire pour méthode**

Pour les services de renseignement, l'enjeu réside dans la maîtrise du flux de données produite dans des volumes inédits afin de produire une information pertinente, supposant une nouvelle organisation.

#### **Une fondation**

La massification de la donnée numérique transcende les pratiques et habitudes de collecte, d'archivage et de catalogage de tous les SR. Ce *tsunami* ne peut s'agréger intelligiblement que traduit en ontologie, étape indispensable pour devenir une « procédure génératrice et constructrice de sens, capable d'assembler non seulement des matériels, mais aussi des temps hétérogènes » <sup>(5)</sup>. Traduire la donnée technique en modèle de données ou ontologie est le premier obstacle d'une intégration efficiente du renseignement, sinon condamné à reproduire des tâches ancillaires classificatoires. L'intégration du renseignement fut l'une des priorités de la réforme pensée par le *Director of National Intelligence (DNI)* James Dean Clapper <sup>(6)</sup> et mise en œuvre par Robert Cardillo, son adjoint en charge de l'intégration du renseignement en tant qu'*Intelligence Integrator*, fonction devant unir « les rôles d'analyse et de collecte pour élever le partage de l'information et la collaboration entre ces deux fonctions essentielles » <sup>(7)</sup>. L'échelon de coordination du renseignement américain a conduit cette réforme en s'appuyant sur l'intégration des données hétérogènes grâce à leurs modélisations sous la forme d'une représentation canonique intuitive correspondant aux besoins des SR. Le modèle de donnée ou l'ontologie est l'étape de traduction de données techniques brutes en objet métiers (conflit, base aérienne, munition, alerte...), préliminaire indispensable pour agréger, naviguer et visualiser l'information. Naviguer au milieu de ces métadonnées

<sup>(4)</sup> Audition de Patrick Calvar, directeur général de la sécurité intérieure, Commission de la défense nationale et des forces armées, mardi 10 mai 2016.

<sup>(5)</sup> Teresa Castro : « Atlas : pour une histoire des images "au travail" », *Perspective*, 2014.

<sup>(6)</sup> Floran Vadillo et Alexandre Papaemmanuel : *Les Espions de l'Élysée* ; Tallandier 2019.

<sup>(7)</sup> *Statement by the Director of National Intelligence James R. Clapper to the United States Intelligence Community workforce*, 20 août 2010.

permet une pensée compositionnelle, distribuée dans les espaces et dans le temps, ouvrant le chemin de la confrontation, de la comparaison et de la mise en relation. Cette étape permet de corrélérer toutes les données techniques sans discrimination avant de s'appuyer sur des produits de renseignement finis et qualifiés, « car des événements apparemment insignifiants dans une analyse basée sur une seule source de renseignement peuvent s'avérer importants lorsqu'ils sont intégrés dans une analyse *multi-INT* »<sup>(8)</sup>. Cette stratégie ayant pour première étape la gouvernance de la donnée doit s'appuyer sur une organisation adaptée, qui redonne une marge de manœuvre aux SR face aux stratégies de désintermédiation incarnées par des acteurs privés (*data brokers*, agrégateurs de données commerciales), porteurs de la promesse d'un renseignement privatisé, réactifs en contrepartie d'une exploitation massive des données personnelles.

### **Une organisation**

Si l'idée est séduisante et dans l'air du temps, la donnée doit être couplée à une finalité concrète et à des moyens de mise en œuvre. Les SR doivent donc « réinternaliser » certaines compétences, car le jugement et l'expertise humaine seront toujours indispensables. Quelques experts en métadonnées, de nombreux experts métiers, des délégués à la protection des données et des administrateurs des données, autant de fonctions stratégiques pour la mise en œuvre d'une gouvernance maîtrisée et compatible du cadre légal de l'exploitation et de la conservation des données. Dépassant la simple mise à disposition au plus grand nombre, la donnée doit alimenter la prise de décision procédant ainsi à une véritable recomposition interne des pouvoirs et des savoirs. Grâce à cette approche holistique par la donnée, chacun peut l'interroger et la croiser afin d'infléchir, à son échelle, la définition ou la mise en œuvre de cette politique publique si particulière. L'analyste thématique (contre-prolifération, contre-terrorisme...) et/ou géographique n'est plus isolée, mais au contact de l'expert capteur (drones, satellites, données de communication...) et du *datascientist*, l'échelon central est quant à lui dynamisé par l'échelon déployé sur le terrain, l'organisation partage ainsi en permanence autour d'un plateau virtuel. Cette transformation redonne son sens à l'action des agents tout en l'inscrivant concrètement et techniquement dans une stratégie globale, maîtrisée par l'État.

### **Le théâtre de Camillo pour outil**

Camillo<sup>(9)</sup> a tenté de contenir dans une infrastructure à la fois la totalité des savoirs et les secrets de la beauté. Pour valoriser l'archive, mémoire précieuse des SR, en « devenir-voir » et « devenir-savoir »<sup>(10)</sup>, l'infrastructure *Big Data* alimentée par un écosystème de la donnée renouvelé, est au service d'un parcours à travers un

<sup>(8)</sup> Jean-Philippe Morisseau : « Le concept de l'*Activity Based Intelligence (ABI)* », *Geointblog*, 14 août 2020.

<sup>(9)</sup> Giulio Camillo : *Le Théâtre de la mémoire*, Eva Cantavenera et Bertrand Schefer (traduction) ; Éditions Allia, 2001.

<sup>(10)</sup> Georges Didi-Hubermann : *Atlas, ou le gai savoir inquiet (L'Œil de l'histoire, 3)* ; Paris, Les Éditions de Minuit, 2011.

nombre de données (images, interception, *tweet*, rapport...) utile et limité car, là où « l'archive invite le savant à se perdre dans l'infinitude des matériels accumulés »<sup>(11)</sup>, l'Atlas numérisé prométhéen garantit l'efficacité.

### **Une infostructure d'intégration**

L'infostructure numérique sert de trame sur laquelle l'organisation vient tisser ses données, ses capacités analytiques et l'exécution opérationnelle. Mais la technologie est un outil et non une fin en soi, c'est pourquoi l'action des SR doit s'appuyer sur une plateforme éprouvée, permettant la convergence de systèmes hétérogènes et de données isolées. Afin de se concentrer exclusivement sur leurs réorganisations internes et leurs performances, de grands acteurs industriels et régaliens français s'appuient déjà sur un système d'exploitation de la donnée robuste, répondant à trois qualités essentielles : une couche d'intégration, de stockage et de calcul, véritable armature au service d'une connexion immédiate à toutes les sources de données, quels que soient leurs volumes ou leurs structures ; une couche de gestion donnant au SR les moyens d'organiser et d'orchestrer en toute sécurité la totalité de son patrimoine de données ; une couche d'analyse de la donnée apportant aux questions critiques, des réponses graphiques et aisément interprétables (résilience économique d'un pays, ingérence internationale dans un conflit, etc.).

En dotant les analystes de tableaux de bord sur-mesure et de rapports dynamiques et interactifs, la capitalisation sera dynamique et le graphe relationnel d'un agent pourra être complété par un autre, la carte sera vivante et enrichie régulièrement avec de nouvelles données. Dépassant le rapport de renseignement figé et dactylographié, l'action de la communauté nationale du renseignement (CNR) gagne alors en performance tout en étant totalement synchronisée grâce à une source de vérité unique et partagée en interne, et avec ses partenaires (du RIM à la FIR<sup>(12)</sup>, de la FIR à la CNR, de la CNR à un tiers de confiance en opération coalisée, etc.) animant ainsi une communauté de confiance de la donnée. Gagnant en performance, les SR se positionnent alors au centre du jeu et peuvent ainsi animer un écosystème de la donnée foisonnant.

### **Un écosystème de la donnée**

La maîtrise de ces données suppose également un écosystème de la donnée plus abouti, affranchi des cloisons historiques entre les acteurs et les savoirs du renseignement au service d'une pensée buissonnière. D'autant que la menace protéiforme oblige tous les acteurs à l'interopérabilité. L'Académie du renseignement doit permettre, en s'appuyant sur des « SR plateformes », de brasser des ressources très diverses, militaires, industrielles et académiques pour créer une culture du renseignement actionnable et performative rompant avec la pratique du renseignement,

<sup>(11)</sup> Teresa Castro : *op. cit.*

<sup>(12)</sup> RIM : renseignement d'intérêt militaire ; FIR : fonction interarmées du renseignement.

cloisonnée, ayant, par pragmatisme, nécessité, mais parfois par manque de volonté, laissé les débats se tenir en milieux clos <sup>(13)</sup>. Une approche foisonnante s'appuyant sur une stratégie de plateforme permettrait de faire émerger un écosystème du renseignement autour de laboratoires de recherche fondamentale, d'universitaires et d'ingénieurs, autant d'acteurs de l'information qui au mieux s'ignorent, au pire nourrissent une méfiance mutuelle.

Pour ne pas rester qu'incantatoire, cet écosystème doit s'arrimer aux SR plateformes grâce à leur *APIsation* (*Application Programming Interface*), c'est-à-dire leur capacité à capter l'innovation extérieure sans pour autant exposer l'entièreté de leur patrimoine informationnel. L'*API* est un processus de structuration d'un programme informatique pour qu'il soit accessible par des logiciels tiers. Ce moyen technique permet la mise à disposition de services ou de données d'une entreprise, d'une administration ou d'un SR tout en garantissant aux développeurs d'économiser du temps et des ressources. En structurant leur biotope, le système d'*API* permet aux entreprises d'augmenter leur portefeuille de services et d'applications grâce à de nouveaux acteurs innovants. Les SR peuvent désormais donner accès à leurs plateformes par l'intermédiaire de ces interfaces de programmation à une *start-up* ou à un laboratoire (développant des algorithmes de détection de changement sur une photo ou des outils de reconnaissance du locuteur sur des flux audio...) <sup>(14)</sup>. Cette approche permet alors au SR d'éprouver la performance de ces acteurs de la science des données, innovant dans le domaine de l'intelligence artificielle (IA) et *Machine Learning* (*ML*). Si l'intégration d'une innovation extérieure peut ainsi être plus fluide, elle doit néanmoins être incorporée de manière responsable et efficace en s'appuyant sur des mécanismes d'évaluation et de validation continues, et de mise en concurrence des modèles algorithmiques. La plateforme agissant comme un socle d'intégration permet également la capitalisation des Retex des utilisateurs afin d'apporter des améliorations continues tout en contextualisant l'usage des traitements algorithmiques. L'IA encore beaucoup fantasmée peut devenir opérationnelle, car déployée dans le cadre d'un processus décisionnel représentatif (analyse d'image satellitaire, réorientation des capteurs...) et d'un contexte opérationnel réel (données techniques, architecture systèmes...) et non dans un environnement de laboratoire de R&D. Le champ des possibles est alors modélisable (impact des nouvelles technologies sur le domaine capacitaire, hybridation des modes d'actions, anticipation d'un plan de vol, détection précoce d'un déploiement ennemi...) et les scénarios d'anticipation sont fiabilisés par cette approche systémique sans jamais suspendre le cadre légal d'utilisation des techniques de recueil de renseignement.

### L'ascèse et le contrôle comme obligation

Les enjeux pour les SR, s'articulent autour de trois défis : la prise de décision basée sur des données est excessivement complexe à cause de la faible qualité

<sup>(13)</sup> A. Papaemmanuel : « Une culture du renseignement pour faire face au tsunami de la donnée », *La Tribune*, 2 juin 2016.

<sup>(14)</sup> *Chapter 5: AI and the Future of National Intelligence Blueprint for Action.*

des données à disposition ; les données risquent d'être partagées de façons inappropriées, soit trop peu pour mener à bien l'analyse et *in fine* la mission, soit au-delà de ce qui est nécessaire et proportionné ; les décisions prises grâce à ces données peuvent demeurer opaques aux analystes tout comme aux autorités de contrôle si l'accès aux données brutes ou aux mécanismes de fusion n'est pas transparent. Les machines accompagnent, mais la décision et la responsabilité restent humaines.

Le premier défi est un problème de confiance dans la qualité des données à disposition, le deuxième un problème de collaboration sécurisée, enfin le troisième un problème de transparence et de responsabilité.

### **Qualité et traçabilité des données**

Les analystes doivent tout d'abord pouvoir évaluer la véracité et la fiabilité des données. Toute décision prise sur la base d'informations inexactes risque d'impacter le succès opérationnel d'une mission pouvant aboutir à des erreurs lourdes de conséquences, comme une surveillance injustifiable. Ainsi, il en va des libertés individuelles et publiques de garantir que les analystes puissent s'appuyer sur des technologies fournissant les capacités suivantes :

- Les utilisateurs doivent pouvoir vérifier la provenance des données à leur disposition et ainsi déterminer eux-mêmes si ces données sont fiables. En respectant les permissions d'accès, les utilisateurs peuvent évaluer l'historique des données, suivre les ajouts, modifications et suppressions successives qu'elles ont suivies, permettant de ne pas répéter les mêmes erreurs.
- Les solutions technologiques doivent maintenir une version unique et canonique de la donnée, plutôt que de laisser se propager de multiples copies au sein d'une organisation. Afin de préserver cette unicité, ces solutions doivent offrir des outils de gestion des versions, permettant aux utilisateurs de proposer par eux-mêmes des modifications de données, une fois que celles-ci ont été vérifiées et approuvées.
- Les consommateurs de données doivent pouvoir s'assurer que ces dernières sont cohérentes et à jour. Pour ce faire, des contrôles automatiques vérifient que les données respectent certaines conditions préétablies. De plus, les consommateurs doivent pouvoir vérifier si les données sont fraîches et les rafraîchir le cas échéant. Ainsi, il sera garanti que les décisions prises ont été basées sur les meilleures informations à disposition.

### **Interopérabilité et partage**

Pour pouvoir partager des données, il faut paradoxalement être en mesure de les sécuriser. Les données doivent être sécurisées avec un très haut niveau de spécificité. Les permissions des utilisateurs peuvent être centrées sur des données individuelles ou des jeux de données entiers. Cette granularité crée différents degrés

d'accès pour différents groupes d'utilisateurs, par exemple par niveau d'habilitation ou grade. Pour chaque ressource individuelle, les organisations peuvent définir ce que les utilisateurs peuvent découvrir, lire, modifier et supprimer. Plus la granularité dans la sécurisation des données est permise, allant jusqu'aux métadonnées ou à l'attribut le plus primaire, plus le partage granulaire sera possible, car les droits d'accès seront fonction du niveau d'habilitation des utilisateurs et du niveau de classification des données (avec la possibilité de définir des groupes d'utilisateurs avec des accès différents sans aucun risque de partage non autorisé). Ainsi, les nouvelles technologies se doivent de mettre en place un contrôle des accès de haute précision :

- Les SR doivent pouvoir organiser leurs données selon les différents niveaux de classification. Le principe de proportionnalité requiert que les utilisateurs aient uniquement accès aux données strictement nécessaires aux vues de la finalité légitime de leur traitement. Cela suppose des contrôles d'accès hautement configurables, mais également un niveau d'orchestration permettant de configurer ces permissions dans la mesure du minimum nécessaire aux utilisateurs (basé sur leur rôle ou leur finalité de traitement et limité dans le temps).
- Le principe de minimisation des données implique que les utilisateurs aient accès aux données, nécessaires à leur finalité de traitement, les moins sensibles. Ce processus doit être dynamique : la même donnée doit apparaître à différents niveaux de rédaction en fonction de l'identité de l'utilisateur, mais aussi en fonction de la finalité spécifique au moment de l'accès à cette donnée. Cela révèle que le minimum de données confidentielles et donc de protéger la vie privée de traitement ou corrélation abusifs.
- Afin d'être conforme aux exigences de gouvernance des données et dans le but de se protéger contre le risque de réidentification des données, les administrateurs ont besoin d'avoir une compréhension claire des activités de traitement sur leurs plateformes. Savoir exactement où sont les données personnelles est un prérequis permettant de s'assurer que des données qui ne doivent pas être croisées ne le soient, ni le seront.

### **Transparence et responsabilité**

Afin de garantir la confiance entre les administrés et les administrations qui les protègent, aucun compromis entre sécurité et libertés individuelles et publiques n'est envisageable. C'est pourquoi, les fournisseurs de solutions technologiques se doivent de respecter les principes suivants : garantir que toutes les décisions reposant sur des données soient prises exclusivement par des agents et non des algorithmes – les technologies doivent responsabiliser et non remplacer les agents ; fournir un journal d'audit robuste ainsi que des capacités d'analyse de ce journal ; s'appuyer sur les autorités de protection des données, seules à posséder l'expertise pour protéger les données – ainsi pour en arriver aux systèmes logiciels « *privacy and legal by design* », c'est-à-dire intégrant les principes protecteurs des libertés individuelles

et du cadre légal (proportionnalité, subsidiarité, contrôle) dès leurs conceptions, les fournisseurs technologiques doivent pouvoir interagir avec ces autorités à toutes les étapes du développement.

Tous les principes décrits ci-dessus ne peuvent être considérés isolément, comme autant d'outils déconnectés, car cela conduirait à aggraver les risques de sécurité et de conformité tout en réduisant les possibilités d'interopérabilité et de partage de l'information <sup>(15)</sup>.

### La « donnée » monde ou la nécessaire ambition warburgienne du CNR-LT

L'historien de l'art Warburg, en cartographiant par des séries d'images similaires un certain territoire de leur histoire, donne à voir le récit iconique fondé sur la spatialisation des phénomènes, organisé autour de relations de proximité et de distance. La donnée, comme l'image, devient alors « champ de forces » qui suscite la mise en place de parcours cognitif.

Les SR rêvent depuis longtemps de cet Atlas rendant le monde accessible dans leur désir panoptique de connaissance. L'innovation numérique et la mise en donnée du monde portent l'ambition d'augmenter l'analyste, qui précédemment était seul face à cette ambition de connaître le monde, d'appréhender une zone géographique, de pressentir le cours des événements. La donnée permettrait alors de faire défiler l'État du monde, rêve d'analyste courant après ce globe virtuel rendant accessible le savoir en un clic droit.

Face à cette masse inédite d'informations numérisées, la capacité de la CNR, comme celle des échelons de synthèse des services de renseignement, à jouer un rôle de fédérateur de l'information, s'en trouve affaiblie. Face à la chimère d'un agent de renseignement rendu obsolète par les capacités d'un supercalculateur, le politique doit au contraire doter les SR des moyens pour se concentrer sur l'essentiel : la réflexion et la prise de décisions. Cela ne sera possible sans une fondation de donnée, mise en œuvre par des ressources humaines internes travaillant sur une plateforme d'intégration, d'agrégation et d'exploitation de données massives conforme au cadre légal.

Comme jadis le DNI au profit de l'intégration de donnée, le CNR-LT est l'acteur interstitiel en mesure de transformer « le *pharmakôn* toxique en une thérapeutique. Seul ce qui est toxique peut remédier à la toxicité » <sup>(16)</sup>, posant les préceptes d'un couple inédit encore, entre mémoire et IA, mêlant le génie de Simonide de Céos et le regard herméneutique de Warburg. ♦

<sup>(15)</sup> Jean Guisnel : « Services secrets étrangers : le double discours de la France », *Le Point*, juin 2021.

<sup>(16)</sup> Bernard Stiegler : « Le numérique empêche-t-il de penser ? », *Esprit*, janvier 2014.