

# Qu'est-ce qu'un répertoire ?

## De multiples exigences pour un système complexe



Pascal Rivière\*

*« Simple liste » d'objets (individus, entreprises) à laquelle on associe des caractéristiques stables, un répertoire n'a rien pour impressionner a priori, comparé à d'autres composantes d'un système d'information.*

*Pourtant, en tant que référence centrale voire opposable dans des processus de gestion, notamment administratifs, le répertoire charrie avec lui de multiples enjeux. Il invite à sa table des utilisateurs individuels ou institutionnels aux intérêts divergents, en attente d'un niveau de service élevé, mêlant qualité du contenu et interopérabilité. Pour atteindre ce degré d'exigence malgré l'hétérogénéité des usages, le répertoire doit posséder des propriétés particulières, et se fonder sur toute une organisation pour le construire et le faire fonctionner.*

*Il en résulte un système riche et dynamique, avec pour colonne vertébrale une sémantique rigoureusement définie, ainsi qu'une infrastructure juridique ou conventionnelle. La vie de cette structure complexe s'organise autour de processus d'alimentation bien identifiés et normalisés, d'une vaste palette de services, et d'une démarche de maîtrise de la qualité s'appuyant sur une équipe de gestionnaires dédiée. Le pilotage d'un tel système requiert des prises de décision à plusieurs niveaux, et par conséquent des instances de gouvernance adaptées. Nous sommes donc à des années-lumière de la « simple liste »...*

 *As a "simple list" of objects (individuals, businesses) with stable attributes, a register does not seem impressive at first glance, compared to other components of an information system.*

*However, as a central reference for operational processes, particularly administrative processes, a register raises multiple issues. Firstly, it often has numerous users who do not necessarily share the same objectives and agenda. Its expected level of service is high, as it combines data quality and interoperability needs. Reaching this level despite heterogeneity of uses requires specific properties, and a strong organization to ensure developments and day-to-day activities. Finally, a register can be seen as a powerful and dynamic system that relies on rigorous semantics, a legal architecture and a solid technical infrastructure. Its lifecycle is based on standardized data flows, a wide range of services, and quality control supported by a dedicated team. Managing such a system implies multi-level decision making, in other words a governance of the register. All this is very far from the "simple list"...*

\* Chef de l'Inspection générale, Insee,  
[pascal.riviere@insee.fr](mailto:pascal.riviere@insee.fr)

Dans nos sociétés, le besoin de références communes à visée opérationnelle se perd dans la nuit des temps. Dès l'Antiquité, les calendriers, nécessaires à l'organisation efficace des échanges et à toute forme de planification, apparurent et se perfectionnèrent, sans toutefois se coordonner. Dans la métrologie antérieure à la Révolution française, « *les unités [étaient] incohérentes, leurs mesures diverses et les procédés de mesure incertains : ni caractère national, ni doctrine d'ensemble* » (Lacombe, 1979). En 1789, on dénombrait en France presque autant d'unités de mesure que de lieux et de corps de métiers. À Paris, merciers, drapiers et marchands de toile avaient chacun leur aune, par exemple<sup>1</sup>. Muid, minot ou boisseau pouvaient représenter des mesures très diverses selon les situations. Elles pouvaient être liées à la matière : setiers de sel, mires d'huile, corbes de foin, mines de grains, penses de fromage, etc. Si les paysans, commerçant localement, s'en accommodaient peu ou prou, certains négociants n'y trouvaient pas leur compte<sup>2</sup>. Passer à une métrologie universelle pour la longueur, le temps, la masse, n'allait pas de soi, loin de là. On n'imagine pas aujourd'hui l'extraordinaire épopée que fut la création du système métrique (Débarbat et Quinn, 2019), ni le rôle décisif qu'eurent les circonstances politiques favorables des débuts pour lancer le processus.

Dans tous les domaines d'activité, on s'est ainsi attelé à construire des cadres de référence communs, partagés, reconnus, publiés : des **référentiels**. Ce ne fut pas sans douleur, sans conflits de pouvoir (Alder, 2005). À l'ère de l'informatisation et dans un monde de plus en plus complexe, les référentiels n'ont cessé de se développer, incluant de nouvelles dimensions applicatives, techniques, systémiques, sans que ne soit démentie la difficulté que représente leur construction et leur mise en œuvre.

Au sein des systèmes d'information, on peut distinguer plusieurs types de référentiels, mais deux en particulier méritent d'être isolés. En premier lieu, les **nomenclatures** (Guibert, Laganier et Volle, 1971), ou classifications, répondent à un besoin permanent de catégoriser (Bowker et Star, 2000)<sup>3</sup> : nomenclature de professions (Amossé, 2020), nomenclature d'infractions (Camus, 2022), classification internationale des maladies (CIM), etc. Mais il existe un second type de référentiel particulièrement crucial, d'apparence simple en première approche : les **répertoires**, répondant à un besoin de clarté sur une « population<sup>4</sup> » de référence. Mais en quoi consistent-ils ? Selon quels principes, avec quels outils, méthodes, contraintes la tenue d'un répertoire s'organise-t-elle ? Répondre à ces questions et donner ses lettres de noblesse à une brique majeure et pourtant mal connue de l'univers des données est l'objet de cet article.

## ► Un répertoire, qu'est-ce que c'est ?

Commençons par une définition générale, abstraite : *un répertoire, c'est une liste d'instances d'une même entité*. Par exemple, pour une entreprise, la liste de ses fournisseurs, celle de ses clients, la liste de ses produits, etc. Les « instances » ne sont pas des concepts, des catégories qu'on invente mais des objets précis, qui naissent et qui peuvent disparaître.

<sup>1</sup> L'instauration du système métrique décimal, 1790-1837 :

<https://unesco.delegfrance.org/L-instauration-du-systeme-metrique-decimal-1790-1837-Focus-Memoire-du-Monde-2639>.

<sup>2</sup> Ces préoccupations commerciales conduisirent à la revendication d'une uniformisation des mesures, dans les cahiers de doléances présentés aux États généraux en 1789.

<sup>3</sup> Le chapitre introductif de l'ouvrage s'intitule « *To classify is human* ».

<sup>4</sup> Terme à prendre au sens large : ce peut être une population de logements, par exemple.

Répertorier, c'est en quelque sorte inventorier, cataloguer, et le répertoire résulte de cette opération. Il se caractérise donc essentiellement par sa **structure** : c'est une liste « à plat », une liste d'objets avec une certaine pérennité, une certaine épaisseur temporelle (répertoire de véhicules, mais pas répertoire d'accidents, par exemple). La structure de répertoire n'implique pas systématiquement la dimension référentielle : un répertoire d'amis sur un *smartphone*, ou bien le répertoire d'une cantatrice (liste d'œuvres) n'ont pas nécessairement vocation à servir de référence.

Car « référentiel » signifie « qui a trait à la référence ». Le terme existe depuis longtemps dans de nombreux domaines comme la physique (référentiel galiléen), la linguistique, la psychologie ou l'éducation (*Cros et Raïsky, 2010*). Dans un système d'information, un référentiel est une *source d'information reconnue*, contenant des données « maître » (*master data*), et dans laquelle on peut puiser. Les répertoires sont une de ces sources. Dans ce qui suit, les répertoires désigneront implicitement des **répertoires avec un statut référentiel** (on utilisera indifféremment l'un ou l'autre terme). On peut citer par exemple des répertoires d'individus<sup>5</sup>, d'étudiants<sup>6</sup>, de professionnels de santé, d'électeurs<sup>7</sup> (*Demotes-Mainard, 2019*), d'entreprises (*Bernard, 1995*), de véhicules routiers, d'établissements d'enseignement<sup>8</sup>, etc.

Mais quelles sont les caractéristiques permettant d'affirmer qu'un répertoire possède un tel statut ? Pour cela, (*Bizingre et alii, 2013*) distinguent cinq propriétés fondamentales : centralité, qualité, stabilité, unité de sens et interopérabilité (**figure 1**).

## ► Un répertoire : un positionnement central...

La centralité d'un répertoire signifie qu'il est « la référence » pour un ensemble d'acteurs, qui le reconnaissent en tant que tel, lui conférant ainsi sa légitimité. Ce serait en quelque sorte, et toutes proportions gardées, l'équivalent de *l'universalité* pour le système métrique. Cette légitimation implique un certain degré d'officialisation, qui se matérialise souvent par un arsenal juridique : par exemple, l'arrêté du 13 novembre 2013 relatif à la mise en place d'un répertoire national des établissements sanitaires donne un statut officiel à Finess<sup>9</sup>, en décrivant notamment ses finalités, son contenu et le service qui le gère.

Dans certains cas, cette forte légitimité donne un poids important au répertoire, allant parfois jusqu'à le rendre opposable pour des actes administratifs. Cette position centrale est facilitée par la neutralité de l'organisme en charge de sa gestion. Mais cette centralité ne s'exprime pas dans l'absolu, pour tous usages : elle s'inscrit dans le cadre des finalités attribuées au répertoire<sup>10</sup>, des enjeux auxquels il est censé répondre. Il existe ainsi des répertoires dont la finalité est essentiellement statistique : ainsi, pour le répertoire statistique des véhicules routiers (RSVERO), la finalité principale est « *la production de statistiques, d'études ou de rapports d'évaluation des dispositifs de politiques publiques* ». Pour d'autres,

<sup>5</sup> Cf. l'article de Lionel Espinasse et Valérie Roux sur le RNIPP dans ce même numéro.

<sup>6</sup> Avec l'INE (identifiant national étudiant).

<sup>7</sup> On constate au passage que l'objet est lié à un rôle : un même individu peut être électeur, étudiant, client, etc. et ce sont là des objets bien distincts, associés à des systèmes d'information totalement différents.

<sup>8</sup> Répertoire académique et ministériel sur les établissements du système éducatif (RAMSESE).

<sup>9</sup> Fichier national des établissements sanitaires et sociaux.

<sup>10</sup> Au minimum, on peut dire que les finalités d'un répertoire sont transverses, et non spécifiques à un métier.



**L'encadrement juridique ou conventionnel imposé à un répertoire peut s'accompagner de contraintes, comme l'obligation d'alimenter le répertoire... qui offrent en contrepartie des garanties de qualité, fort utiles.**



les finalités, parfois nombreuses, sont administratives : ainsi, l'une des finalités de Finess est d'être l'autorité de référence pour les établissements qui demandent des accès numériques aux différents systèmes de santé.

Faire jouer au répertoire un rôle central requiert d'associer de nombreux acteurs à travers des instances de gouvernance : les organismes concernés (producteurs

ou utilisateurs) s'expriment avec chacun leurs propres enjeux, qui ne sont pas toujours cohérents avec ceux du répertoire. Des arbitrages sont donc indispensables.

► **Figure 1 - Cinq propriétés d'un répertoire avec statut de référentiel**

1		<b>CENTRALITÉ</b>	<b>UN RÉPERTOIRE EST CENTRAL, C'EST UNE RÉFÉRENCE LÉGITIME</b>
2		<b>QUALITÉ</b>	<ul style="list-style-type: none"><li>• DES FLUX D'ENTRÉE EXHAUSTIFS ET FIAIBLES</li><li>• DES DONNÉES FRAÎCHES ET TRAÇABLES</li><li>• UNE DOCUMENTATION ET DES SERVICES DE QUALITÉ</li></ul>
3		<b>STABILITÉ</b>	<b>DES DONNÉES STRUCTURANTES, STABLES &amp; FIAIBLES...</b> <b>... DONC EN RELATIF PETIT NOMBRE</b>
4		<b>UNITÉ DE SENS</b>	<b>CE N'EST PAS UN FOURRE-TOUT...</b> <b>... IL POSSÈDE UNE HOMOGENÉITÉ SÉMANTIQUE</b>
5		<b>INTEROPÉRABILITÉ</b>	<ul style="list-style-type: none"><li>• TECHNIQUE (DISPONIBILITÉ, SÉCURITÉ),</li><li>• SYNTAXIQUE (FORMAT)</li><li>• SÉMANTIQUE (IDENTIFIANTS, CONCEPTS, NOMENCLATURES)</li></ul>

Enfin, l'encadrement juridique ou conventionnel imposé à un répertoire peut s'accompagner de contraintes, comme l'obligation d'alimenter le répertoire... qui offrent en contrepartie des garanties de qualité<sup>11</sup>, fort utiles. Réciproquement, le fait que les données soient reconnues comme fiables lui confère une plus grande légitimité, facilitant sa position de pivot : en effet, décréter qu'un répertoire est central, sans se préoccuper de la réalité de son contenu est insuffisant ; il faut qu'il soit de qualité acceptable... Comment appréhender cette qualité ?

## ► ... une exigence de qualité...

La qualité des données revêt en pratique des aspects divers (*Di Ruocco, Scheiwiler et Sotnykova, 2012*) : pertinence, exactitude, complétude, consistance, accessibilité, etc. Vis-à-vis de cette grille d'analyse générale, le cas d'un répertoire présente des spécificités. Ainsi, de par sa position de référence ultime, il est par construction plus difficile d'analyser sa qualité en le comparant avec d'autres sources de données... car un référentiel, c'est justement ce à quoi les autres sources se comparent (*Bizingre et alii, 2013*)<sup>12</sup>. Il soulève ainsi des difficultés particulières pour élaborer des indicateurs (*Rivière, 2005*), par exemple pour mesurer le taux de « faux actifs »<sup>13</sup>.

En dehors des délicates questions d'exactitude, on peut insister sur :

- l'exhaustivité, sur un champ préalablement défini (*Wallgren et Wallgren, 2016*) ; deux défauts symétriques sont possibles, la sur-couverture (unités présentes à tort, par exemple des doublons), ou la sous-couverture (des unités qui devraient être présentes, ne le sont pas). La sous-couverture est toujours plus délicate à traiter, car on ne sait pas où sont les manques ;
- la fiabilité des flux d'entrée, ce qui requiert le développement d'une batterie de contrôles avant mise à jour du répertoire, tests de diverses natures pour vérifier que les données sont conformes et s'assurer que le risque d'introduire des erreurs est limité (*Sureau et Merlen, 2021*) ;
- la fraîcheur, c'est-à-dire la rapidité de prise en compte des événements de mise à jour : cela peut varier d'un répertoire à l'autre, voire d'une variable à l'autre, en fonction des besoins, des usages. Pour un répertoire administratif comme Sirene, l'exigence est plus importante : le délai de prise en compte d'une création d'entreprise y est très court, en raison des conséquences directes pour les entreprises. À l'inverse, les répertoires dits « statistiques », sans enjeu administratif individuel, peuvent avoir une exigence de fraîcheur moins forte ;
- ... bien d'autres aspects tels que la traçabilité des événements, voire l'historicisation des données, la documentation, ainsi qu'une autre dimension de la qualité, la qualité de service (voir *infra* la partie sur l'*interopérabilité*).

Les deux propriétés suivantes, stabilité et unité de sens, renvoient à la sémantique des données.

<sup>11</sup> Le répertoire national commun de la protection sociale (RNCPS) constitue un bon contre-exemple : il est moins contraignant que d'autres, ce qui peut avoir des impacts sur la qualité, du fait de l'absence de déclaration de certains événements.

<sup>12</sup> Cf. chapitre 6, consacré à la qualité.

<sup>13</sup> Unités que le répertoire suppose « vivantes » mais qui ne le sont pas en réalité.

## ► ... une nécessaire stabilité...

La **stabilité** est une propriété très particulière d'un répertoire (*Régnier-Pécastaing, Gabassi et Finet, 2008*)<sup>14</sup>, qui le distingue fondamentalement d'autres bases de données et en particulier des bases de gestion. Il doit en effet contenir des données structurantes, qui ne changent pas en permanence. À cette aune, l'adresse du siège<sup>15</sup> d'une entreprise est une information de référence, mais pas le montant de ses investissements. *Pour chaque entité d'un répertoire*, le principe doit être que les données changent peu dans le temps<sup>16</sup>. « Stable » ne signifie pas que les données sont « immuables » : par exemple, une facture est une information figée, dont le contenu ne changera plus, mais qui n'a rien à voir avec les référentiels.

Il est difficile de fonder la notion de stabilité sur la fréquence de mise à jour, trop dépendante de phénomènes métier et difficilement accessible. Plus simplement, la stabilité des données d'un répertoire peut être caractérisée par le fait que leurs évolutions doivent être indépendantes des processus métier qui les utilisent. Par exemple, dans un référentiel d'hôpitaux, le nombre de lits pourrait avoir une dimension référentielle<sup>17</sup>, alors que le nombre de lits *occupés*, lié aux processus métier (accueil et départ des patients) et qui peut changer tous les jours, n'est en rien de nature référentielle.

La conception d'un répertoire suppose donc de porter une grande attention au choix des variables et de veiller à se limiter à celles qui sont essentielles, structurantes, stables et caractérisant l'objet : ainsi, lorsqu'une entreprise gère un répertoire de clients, on y trouvera par exemple des données de contact (numéro de téléphone, adresse mail), le canal de contact préférentiel (téléphone, courriel, site *web*, courrier, etc.), voire le type d'achat, mais certainement pas les achats eux-mêmes.

## ► ... le respect d'une unité de sens...

Avec la notion d'**unité de sens**, on souligne que le répertoire doit posséder une certaine homogénéité sémantique : on ne peut donc pas y mettre n'importe quoi pour céder aux



**Le répertoire doit posséder une certaine homogénéité sémantique : on ne peut donc pas y mettre n'importe quoi pour céder aux demandes.**



demandes. L'idée est d'éviter que le répertoire devienne un fourre-tout dans lequel on ajoute des informations sous prétexte qu'elles sont utiles à certains utilisateurs. Cela vaut d'abord pour les entités : les responsables Sirene ont ainsi été sollicités pour immatriculer des éoliennes, ou des ruches par exemple... ce qui n'a rien à voir avec la notion d'entreprise ou d'établissement. De telles demandes se comprennent : les utilisateurs

<sup>14</sup> Pp. 45-47.

<sup>15</sup> La variable adresse est un sujet délicat ; c'est volontairement que l'on a pris l'exemple de l'adresse du siège. En effet, une adresse est liée à un rôle : adresse administrative, de l'accueil, du siège, de livraison, etc. On peut aussi parler d'adresse géographique, d'adresse postale, ce qui complique encore les choses...

<sup>16</sup> Ne pas se méprendre ici : quand le répertoire contient des millions d'entités, de nombreuses mises à jour seront effectuées sur l'ensemble du répertoire... mais très peu entité par entité. On peut penser par exemple à la rareté des événements affectant un individu (naissance, décès, mariage, changement de nom, etc.).

<sup>17</sup> On simplifie beaucoup en écrivant les choses ainsi. Le nombre de lits renvoie plus généralement à la notion de capacité, mais il peut y avoir plusieurs capacités, comme la capacité administrative, par exemple.

du répertoire peuvent voir ce dernier comme un outil utile pour faciliter leur travail, sans se préoccuper du « bien commun », perçu comme théorique. La gouvernance d'un répertoire donne donc naturellement lieu à des tensions entre besoins des utilisateurs et nécessité de cohérence au niveau global.

L'unité de sens s'applique également aux données : dans un répertoire d'électeurs (*Demotes-Mainard, 2019*), on va se limiter aux variables utiles pour le rôle d'électeur (commune de rattachement, bureau de vote), et ne pas ajouter des données sans rapport avec ce rôle (par exemple, nombre d'enfants). Plus généralement, un répertoire doit contenir « relativement peu » de variables en principe, car il est coûteux d'en assurer la qualité, et en particulier la fraîcheur ou encore la minimisation des données à caractère personnel, au titre du Règlement général sur la protection des données.

## ► ... et la garantie d'une interopérabilité

Cette dernière propriété aborde un aspect plus technique, qui n'est pas spécifique aux référentiels : le service rendu aux utilisateurs par le « système » répertoire, ainsi que son insertion technique dans l'ensemble du système d'information. En effet, si le répertoire possède toutes les propriétés évoquées précédemment, mais qu'il n'est pas pratique à utiliser et à intégrer, il perd tout son intérêt.

L'interopérabilité<sup>18</sup> revient à considérer le référentiel non pas en termes de contenu, mais de services offerts, au sens informatique du terme<sup>19</sup> (*Régner-Pécastaing, Gabassi et Finet, 2008*), et d'interconnexion efficace avec les systèmes d'information qui l'environnent, partageant avec ceux-ci un même langage, en quelque sorte. Dire qu'il est interopérable, c'est dire qu'il est ouvert et accessible (par exemple *via* des API ou *Application Programming Interface*)<sup>20</sup>.

On distingue traditionnellement interopérabilité technique (exigences de performance, de disponibilité, de sécurité, d'utilisation de standards techniques, etc.), interopérabilité syntaxique (format utilisé, par exemple XML<sup>21</sup>, JSON<sup>22</sup>), et interopérabilité sémantique (identifiants, concepts, nomenclatures, etc.)<sup>23</sup>.

Les répertoires se caractérisent souvent par un grand nombre d'utilisateurs, sans pouvoir connaître et encore moins maîtriser leurs usages ; d'où ce fort besoin qu'ils soient interopérables. À titre d'exemple, la présence d'identifiants reconnus (NIR, numéro Siren, etc.) est un facteur d'interopérabilité, car elle permet de faire le lien avec d'autres processus (par exemple l'utilisation du numéro Siren dans la déclaration sociale nominative ou DSN<sup>24</sup>).

<sup>18</sup> Cf. la présentation du RGI (Référentiel général d'interopérabilité) dans (*DINSIC, 2015*).

<sup>19</sup> Cf. chapitres 5 et 6, sur les spécificités des référentiels en la matière.

<sup>20</sup> Voir Définition de la CNIL : <https://www.cnil.fr/fr/definition/interface-de-programmation-dapplication-api>.

<sup>21</sup> *Extensible Markup Language*.

<sup>22</sup> Le *JavaScript Object Notation* (JSON) est un format standard utilisé pour représenter des données structurées comme les objets *JavaScript*.

<sup>23</sup> On a cité plus haut le RGI, on peut aussi évoquer le cadre d'interopérabilité des systèmes d'information de santé (*Agence du numérique en santé, 2022*).

<sup>24</sup> (*Humbert-Bottin, 2018*).

Derrière cette notion, c'est aussi tout le sujet de l'accostage du référentiel aux systèmes d'information qui est en jeu : qui doit s'accoster, comment... Cela comporte une dimension technique (par exemple une API à utiliser), mais aussi des règles de fonctionnement auxquelles on s'astreint : obligation pour telle application de s'accoster à tel référentiel, obligation de ne s'alimenter qu'à tel répertoire, etc. *Last but not least*, il ne faut pas oublier le sujet de la cohérence des répertoires entre eux, de leurs interactions, et les enjeux associés (par exemple, les liens entre Finess et Sirene, ou bien la mise en cohérence mutuelle SNGI – RNIPP<sup>25</sup>), qui relèvent clairement de l'interopérabilité.

## ► Le contenu d'un répertoire : des données d'identification...

Un répertoire est un ensemble d'entités de même type, et pour chacune d'elles, on trouve toujours peu ou prou le même type de données, liées au statut référentiel.

En premier lieu, l'identifiant : par exemple, le numéro Siret pour un établissement, le numéro RPPS<sup>26</sup> pour un professionnel de santé, le numéro INE<sup>27</sup> pour un élève, étudiant ou apprenti. Il doit être présent pour tout répertoire, puisque c'est ce qui désigne l'objet et l'officialise. Attribuer un identifiant à un objet se nomme *immatriculation* et revient à créer l'objet dans le répertoire. Le processus d'immatriculation est une opération majeure qui demande la plus grande rigueur ; en effet, ce processus ouvre accès à des services, ce qui poserait problème au cas où l'immatriculation est faite avec une fausse identité. L'identifiant est ainsi la clé d'accès à l'entité, clé qui permet aussi de faciliter les appariements entre fichiers portant sur le même objet. Une bonne pratique consiste à faire en sorte que cet identifiant soit non significatif, c'est-à-dire qu'il ne contienne aucune information porteuse de sens ; cette pratique n'est d'ailleurs pas toujours respectée (voir le NIR).



**Les traits d'identité sont distincts de l'identifiant et ne le contiennent pas.**



Mais l'identifiant ne peut être créé à partir de rien ; il faut savoir « de qui on parle ». Ainsi, on doit disposer de *traits d'identité*, grâce auxquels on peut repérer sans ambiguïté l'objet : par exemple, nom-prénom-date de naissance pour un individu, raison sociale-adresse pour un établissement. Les traits d'identité sont distincts de l'identifiant et ne le contiennent pas.

Une précision importante : les traits d'identité ne doivent être présents (avec statut référentiel) que pour certains types de répertoires, dits **répertoires-socles** : ce sont eux qui formalisent le lien entre identifiant et traits d'identité. Les autres répertoires sont dits **répertoires adossés**. Par exemple, le REU, répertoire d'électeurs, est adossé au RNIPP, répertoire-socle d'individus ; autre cas, le RGCU, répertoire de carrières, est adossé au SNGI. Ainsi, dans un répertoire adossé, les traits d'identité, s'ils sont présents, ne constituent pas une référence : le lien entre traits et identifiant dépend complètement du répertoire-socle auquel il est adossé.

<sup>25</sup> Cf. les articles de Joseph Préveraud de Vaumas sur le système national de gestion des identités (SNGI) et de Lionel Espinasse et Valérie Roux sur le répertoire national d'identification des personnes physiques (RNIPP) dans ce même numéro.

<sup>26</sup> Répertoire partagé des professionnels de santé.

<sup>27</sup> Identifiant national étudiant.

## ► ... et d'autres données, référentielles ou non

Au-delà de ce noyau dur, un répertoire contient systématiquement des données de catégorisation, car les entités qui le composent ne sont pas toutes équivalentes. Par exemple, pour des établissements d'enseignement, on utilise un découpage du type : établissements du premier degré, du second degré, de l'enseignement supérieur ou de formation continue, etc. Un répertoire d'entreprises comme Sirene s'appuie quant à lui sur des nomenclatures officielles, ayant une existence et une gestion en dehors du répertoire : ainsi, pour caractériser l'activité principale, on se fonde sur la nomenclature d'activités française (NAF).

On trouve également, mais de façon moins systématique<sup>28</sup>, des données pour positionner l'objet dans l'espace : adresse, géolocalisation, code lié à un zonage. Il faut également le situer dans le temps, à l'aide de dates-clés : dates d'ouverture, ou de création, et réciproquement de fermeture, ou de cessation. Plus généralement, la connaissance de l'historique des objets (étapes de création, cessation-fermeture, mais aussi fusions ou scissions, dans le cas des entreprises par exemple) peut se révéler indispensable.

Il est souvent nécessaire de compter parmi les données les identifiants d'autres objets liés<sup>29</sup>, ou des références à des conventions<sup>30</sup>.

À toutes ces données de fond, il faut ajouter des données dites *de gestion* ; ces données ne sont pas de nature référentielle, mais facilitent la gestion courante du répertoire. Par exemple, les données de contact<sup>31</sup> telles que l'adresse mail, l'adresse physique ou le numéro de téléphone associées à l'entité sont importantes lorsqu'il faut justement la contacter, pour effectuer des vérifications notamment.

Enfin, il est fort utile de disposer, pour certaines données, des métadonnées : provenance de la donnée, ou date de mise à jour, par exemple. Les conventions peuvent aussi se présenter comme des métadonnées.

Ainsi, construire une base contenant toutes les données nécessaires est une chose. Faire en sorte que cet ensemble de données soit de nature référentielle, et donc respecter les cinq propriétés citées précédemment, en est une autre. Toute une organisation doit être mise en place pour obtenir un tel résultat. Comment fait-on, concrètement ?

**28** Ainsi, l'adresse va être référentielle pour une entreprise (sachant que, pour pimenter le tout, il peut y avoir plusieurs adresses), et ne le sera en général pas pour un individu.

**29** Ainsi, pour un répertoire de professionnels de tel ou tel domaine (professionnels de santé, par exemple), il est important d'avoir, parmi les variables, l'identifiant de l'établissement auquel la personne est rattachée.

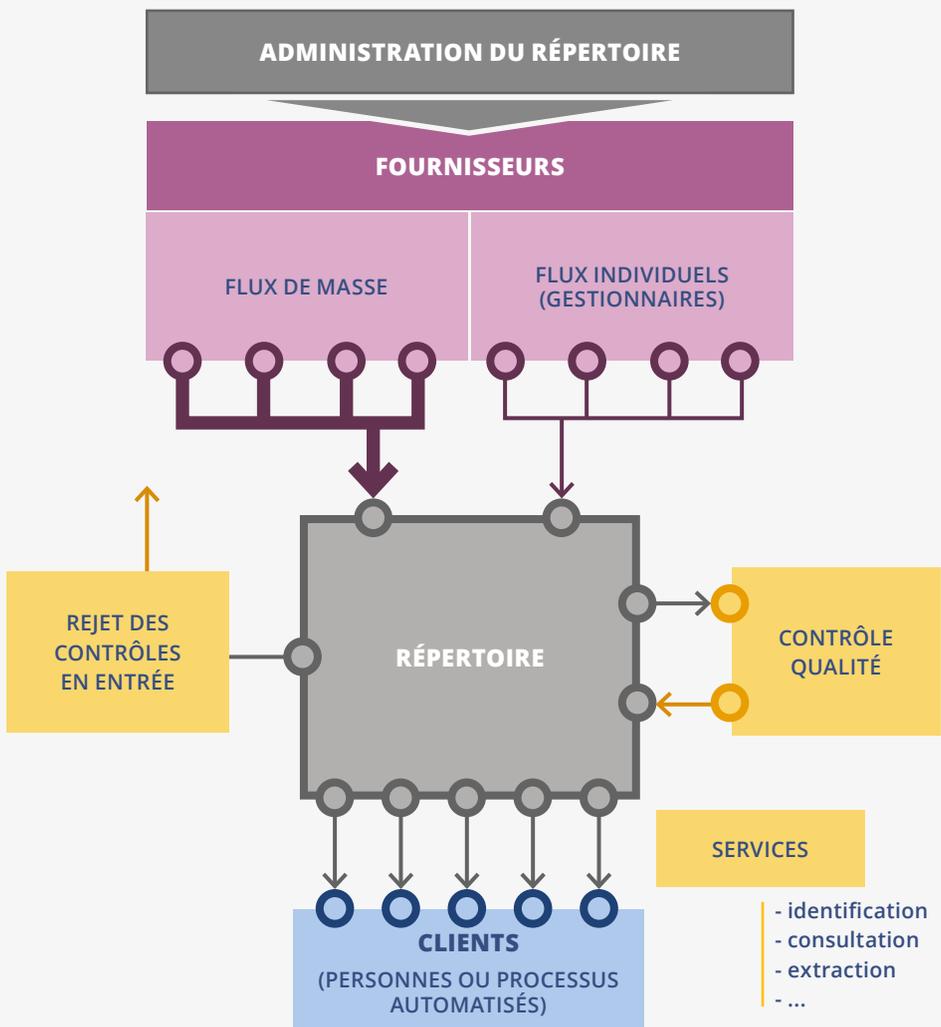
**30** Lorsqu'un répertoire inclut une catégorisation d'activités (par exemple pour les répertoires d'entreprises, d'établissements de santé, d'établissements d'enseignement), certaines peuvent être encadrées par des conventions, dont il faut conserver la référence.

**31** Cette différence entre données référentielles et de gestion est liée au rôle du répertoire, à ses enjeux. En particulier, les « données de contact » seront de nature référentielle dans le cas d'un répertoire de clients.

## ► Un répertoire est un système vivant

L'image courante, lorsqu'on évoque une information de référence, consiste à dire qu'elle est « gravée dans le marbre ». Mais cette image est tout à fait inadaptée dans le cas d'un répertoire. En effet, ce dernier ne peut pas être figé, car le monde réel qu'il est censé décrire ne l'est pas : de nouvelles instances apparaissent (naissances, créations, ouvertures), d'autres disparaissent (cessations, fermetures, décès), et les données descriptives peuvent évoluer (par exemple la capacité d'accueil d'un établissement de santé). Ainsi, les grands répertoires sont affectés quotidiennement par de nombreux événements de mise à jour.

► Figure 2 - Le répertoire, au centre d'un échange de flux



Un répertoire se caractérise également par les forts liens qu'il entretient avec de nombreux systèmes d'information : en amont pour le nourrir, en aval pour les services rendus aux utilisateurs. Ainsi, une image plus appropriée d'un répertoire serait celle



*Une image appropriée d'un répertoire serait celle d'une pompe aspirante refoulante d'informations, d'un cœur battant, régulièrement alimenté par des événements de mise à jour, et irriguant en permanence d'autres systèmes.*



d'une pompe aspirante refoulante d'informations, d'un cœur battant, régulièrement alimenté par des événements de mise à jour, et irriguant en permanence d'autres systèmes (figure 2). Car il n'a pas d'intérêt pris isolément, et ne prend sa signification que vis-à-vis de ses utilisateurs (fournisseurs ou consommateurs), variés et nombreux.

Plus qu'une simple base de données, un répertoire est un véritable système dont les données de référence constituent le centre. À ce titre, on peut le caractériser selon plusieurs dimensions (entrée, sortie, contenu, boucle de rétroaction<sup>32</sup>) :

- les **flux d'alimentation** (unitaires, ou de masse), avec les différents niveaux de contrôle automatique de ces flux entrants ;
- les **services** rendus aux utilisateurs, aux autres systèmes d'information, qui engendrent des flux sortants ;
- la **structure** des informations qu'il contient (schéma de données, pour l'essentiel), donc le contenu, déjà évoqué ;
- les procédures de **contrôle qualité** internes, mêlant contrôles automatiques et vérifications manuelles effectuées par les gestionnaires.

## ► Maîtriser les flux d'entrée : le besoin de normaliser les échanges

Le système vivant ainsi décrit doit être alimenté régulièrement, avec des informations fraîches (autant que possible), à travers des « événements » qui requièrent une mise à jour : création, fermeture, changement d'adresse ou d'activité, toute forme de changement de statut, etc.

Ces modifications ne relèvent pas d'une mise à jour classique dans une base de données : le statut référentiel du répertoire entraîne des exigences particulières de maîtrise de la qualité, de traçabilité des modifications. Il faut donc en particulier mettre en place des contrôles automatisés permettant de filtrer les entrées dans le répertoire.

Plus précisément, on vérifie que le flux d'entrée transmis, le « message », est conforme à un standard : on peut imaginer cela en disant qu'on « branche » le flux de données sur le répertoire et on vérifie que la prise est bien conforme. Dans notre cas, le standard se présente sous la forme de règles à respecter pour toutes les données transmises, que l'on appelle **norme d'échange**.

<sup>32</sup> Au sens systémique (Wiener, 1948) : entrée, sortie, boucle de rétroaction.

Quelles sont ces règles ? Tout d'abord, l'ordre dans lequel se présentent les données du message : par exemple nom, prénoms, date de naissance, département, commune de naissance. Plus généralement, on vérifie que la structure du message est conforme à une structure attendue. On effectue également des contrôles de type (numérique, alphanumérique, date, chaîne de caractères) et de domaine d'appartenance. Par exemple, une date doit respecter un certain format (comme JJMMAAAA), et des règles spécifiques (mois  $\leq 12$ , jour  $\leq 31$ , etc.). Un code de département, ou un code d'activité doivent appartenir à une liste précise et prédéfinie.

En amont, la normalisation est également sémantique : il faut avoir défini le sens de chaque donnée, de chaque valeur dans une liste d'items ou une nomenclature, pour maîtriser l'information transmise. Enfin, standardiser les échanges, c'est formaliser la cinématique des flux, un processus complet dans lequel sont décrites les modalités d'envoi entre émetteur et récepteur ou les notifications transmises, et ce selon le type d'événement.

Normer les échanges est indispensable dans le cas de grands répertoires largement utilisés : norme EDI-CFE pour Sirene, norme A pour le SNGI<sup>33</sup>, norme R pour le RGPU<sup>34</sup>, ce qui donne lieu à de vastes documentations techniques<sup>35</sup>.

Au-delà des référentiels, la standardisation des flux de données est une pratique essentielle dans les systèmes d'information alimentés de façon industrielle : par exemple, la norme GTF (General Transit Feed Specification) est un format standardisé pour communiquer, entre autres, des horaires de transports en commun et les informations géographiques associées ; la norme Odette fait référence dans l'automobile ; le monde de la protection sociale fait l'objet d'une forte rationalisation des échanges (Gratieux et Le Gall, 2016), et utilise en particulier la norme NEODES pour la déclaration sociale nominative (Humbert-Bottin, 2018).

## ► Définir les services rendus par le répertoire

Le répertoire, ainsi nourri et même rassasié de flux d'information réguliers et standardisés, ne demande désormais qu'à être utilisé. Point d'articulation entre systèmes d'information, il se comporte comme un fournisseur de services envers des utilisateurs variés. Dans le cas du référentiel de métrologie (seconde, kilogramme, mètre), des « services » avaient en quelque sorte été développés en dehors du référentiel : respectivement, le chronomètre, la balance, le mètre ruban. C'est très différent pour un répertoire, évoluant dans un système d'information : les services lui sont intégrés, en sont une composante à part entière.



**Il faut isoler en premier lieu les services spécifiques d'un répertoire-socle : l'immatriculation et l'identification.**



Quels sont ces services ? Il faut isoler en premier lieu les services spécifiques d'un répertoire-socle : l'immatriculation et l'identification. Immatriculer un objet, cela consiste à le créer dans le répertoire<sup>36</sup>,

<sup>33</sup> Cf. l'article sur le SNGI dans ce même numéro.

<sup>34</sup> (Sureau et Merlen, 2021), déjà cité.

<sup>35</sup> Sur la difficulté de l'écriture des normes en général, et ce qu'elles impliquent en termes de gouvernance, on consultera avec profit (Mallard, 2000).

<sup>36</sup> Dans le cas du service d'immatriculation, l'utilisateur est en même temps fournisseur ; la frontière amont-aval n'est pas si simple...

à lui donner naissance dans le système d'information. Pour cela, à partir des traits d'identité de l'objet (nom-prénoms-date et lieu de naissance pour un individu, raison sociale – adresse pour un établissement), on vérifie automatiquement qu'il n'existe pas déjà, puis on lui attribue un identifiant (NIR, numéro Siret, etc.)<sup>37</sup>.

La fonction d'identification est tout à fait différente, car elle ne crée aucun objet. À partir de traits d'identité connus (par exemple, « *Leyla Garcia, née le 2 novembre 1999 à Lille* »), on cherche s'il existe une personne ayant des caractéristiques identiques ou proches (par exemple « *Leila* » et non « *Leyla* »), pour obtenir son identifiant (ici le NIR)<sup>38</sup>. Lorsque l'information est imparfaite, incomplète (par exemple en absence de l'année de naissance) ou qu'il reste des ambiguïtés, l'algorithme d'identification doit proposer plusieurs possibilités, plusieurs échos, et les classer par pertinence.

D'autres fonctions viennent naturellement à l'esprit : consulter les données d'une entité (par exemple, toute l'information disponible dans Finess sur tel établissement de santé), ou bien effectuer une sélection (par exemple, la liste, au 31/12/2021, des établissements de santé de l'Oise, créés après 2010). Le fait de pouvoir sélectionner tout ou partie des objets du répertoire à un instant t, de « prendre une photographie », sont des services particulièrement précieux pour les statisticiens, car cela permet de constituer une base de sondage, et plus généralement d'avoir une population de référence pour toute statistique.

Des services peuvent aussi être déclenchés<sup>39</sup> par un mécanisme d'abonnement : par exemple, une caisse de retraite demande à être informée automatiquement de mises à jour concernant les personnes affiliées à son régime.

Les répertoires ont aussi vocation à être utilisés au sein de processus automatisés, sans intervention humaine, en proposant des API intégrables à des processus de production (par exemple, le RGCU fait appel au moteur d'identification de Sirene, pour déterminer les établissements dans lesquels une personne a travaillé pendant sa carrière).

Au total, un répertoire doit proposer une palette de services standard, et différentes modalités d'appel de ces services. Attention cependant, un répertoire n'est certainement pas là pour offrir des services « sur mesure » à chaque utilisateur : par exemple, fournir toutes les « photographies » possibles (contenu, fréquence, filtre, etc.). Ce n'est pas son rôle. La gestion courante en pâtirait, paralysée par la multiplicité des demandes, avec le risque de perdre de vue l'objectif principal, la vocation de référentiel<sup>40</sup>.

Pour finir, soulignons qu'il ne faut pas aborder la notion de service en se limitant aux seuls services numériques : dans les répertoires importants, des équipes de gestionnaires assurent en général un travail de réponse aux sollicitations (téléphone, courriel). Cette activité est essentielle pour la qualité du répertoire, et elle fait bien sûr, partie intégrante du service rendu aux utilisateurs.

**37** ... en s'assurant qu'on ne réutilise pas un identifiant existant.

**38** Cf. l'article sur le SNGI, déjà cité.

**39** Une petite subtilité : de façon générale, les services sont déclenchés par une demande. Et cette demande est un flux entrant, un message, auquel on va appliquer des contrôles figurant dans la norme d'échange.

**40** Une telle limite, auto-imposée, est l'équivalent de l'unité de sens, appliquée aux services.

## ► Le contrôle de la qualité : des traitements automatisés... —

Il en va d'un répertoire comme de tout processus d'approvisionnement : il est toujours plus pratique et plus efficace de contrôler la marchandise à son arrivée, immédiatement. Ainsi, avant l'entrée dans le répertoire, on vérifie la provenance et la conformité à des standards, et cela se fait automatiquement (contrôle d'habilitation, application de la norme d'échange). Ces contrôles sont bloquants : si le moindre d'entre eux n'est pas respecté, il y a rejet. Des contrôles non-bloquants peuvent compléter le dispositif : contrôle de cohérence entre données du flux, ou bien comparaison entre la donnée du flux et celle présente dans le référentiel, pour déterminer si l'évolution est plausible. L'élaboration de ces contrôles non-bloquants et leur mise en œuvre sont un sujet en soi<sup>41</sup> non détaillé ici.

On ne peut se limiter à vérifier le flux d'entrée : il faut aussi analyser le répertoire dans son ensemble, de façon macroscopique. Par exemple, repérer des doublons, pour éviter le phénomène de sur-couverture du répertoire, ou bien repérer des données manquantes, ou plus généralement vérifier que les données du référentiel, « en stock », respectent la structure attendue (en termes de typage et de domaine d'appartenance notamment). C'est ce que proposent les systèmes avancés de contrôle, généralement réunis sous le vocable de *data quality tools* (Boydens, Hamiti et Van Eeckhout, 2021), avec des techniques dites de *profiling* (Olson, 2003).

## ► ... et un vaste travail de vérification humaine —

Il serait illusoire de penser que la démarche qualité d'un répertoire se résume à une approche purement mécaniste et algorithmique, en grande partie automatisée. Qui dit répertoire dit équipe de gestionnaires, chargés de vérifications plus complètes et plus fines qui échappent à la machine, et aussi d'échanges directs avec les entités concernées (téléphone, courriel), avec le terrain en quelque sorte, pour une meilleure compréhension en remontant à la source (Denis, 2018)<sup>42</sup>. L'organisation de ce travail, souvent intitulée « administration de référentiel », prévaut dans tous les grands répertoires administratifs.

Entre autres exemples :

- le travail des gestionnaires Sirene fait l'objet d'une riche documentation composée de fiches pour chaque cas de figure, indiquant les vérifications à effectuer ;
- pour le SNGI, l'équipe du Sandia vérifie la conformité des pièces justificatives (passeports étrangers) ;
- les gestionnaires RNIPP échangent régulièrement avec les mairies ;
- dans l'organisation du RGCU, figure une cellule d'administration de référentiel, en complément du travail de vérification de carrière faits par les techniciens retraite.

<sup>41</sup> Cf. le cas du RGCU, déjà cité.

<sup>42</sup> Jérôme Denis explicite dans le détail toutes les composantes du travail de vérification manuelle des données, activité souvent méconnue mais qui est absolument essentielle pour obtenir des données de qualité.

Administrer le répertoire, choisir ses cibles d'intervention, requiert une vision d'ensemble, quantifiée, sous forme de tableaux de bord, notamment pour fixer des priorités et ainsi orienter le travail des gestionnaires de répertoire. Cela comporte des tâches récurrentes (comme le suivi du traitement des rejets) et des tâches projet (par exemple un projet d'amélioration de la qualité du stock). Cette administration inclut également, de façon plus macroscopique, un suivi de production, qui peut utilement mettre en évidence des anomalies : par exemple, lors d'un suivi mensuel du nombre de mises à jour du répertoire, une chute brutale un mois donné alerte sur un possible dysfonctionnement dans les flux d'entrée.

## ► De l'importance de la gouvernance du répertoire

Pour qu'un système aussi complexe qu'un répertoire fonctionne, il faut le piloter, et en particulier prendre des décisions concernant ses évolutions, dans le respect permanent de ses finalités. Pour cela, si les tableaux de bord sont très utiles, il faut également que des instances de décision soient mises en place, à savoir une véritable gouvernance du référentiel, associant les différentes parties prenantes.

Celle-ci est en fait de deux natures : la gouvernance structurelle (sa définition) et la gouvernance « instancielle » du répertoire (son contenu, son peuplement). Dans la première, on trouve par exemple le modèle de données (même si les modalités de validation sont difficiles), la norme d'échange<sup>43</sup>, le suivi des réglementations concernées, et dans la seconde, la qualité des données et les conséquences concrètes de la non-qualité (retours d'usage), ou bien l'accostage du répertoire sur les systèmes d'information métier.

Dans ces instances, il faut parvenir à un consensus sur les évolutions de contenu : nouveaux types d'objets à incorporer (« intègre-t-on les associations dans le répertoire d'établissements ? »), nouvelles données, nouvelles nomenclatures, etc., en gardant en tête l'unité de sens, pour éviter d'ajouter des objets ou des variables non pertinents.

Les sujets à aborder portent aussi sur les services offerts par le répertoire : extension de la palette de services et du niveau de service (ergonomie, temps de réponse, etc.).

En pratique, « gouverner un répertoire » revient souvent à « rester raisonnable dans ses ambitions ». Ajouter ne fût-ce qu'une seule variable dans un répertoire représente un coût : introduire une nouvelle variable requiert en effet

d'organiser les flux d'alimentation associés, donc de trouver les fournisseurs, de s'assurer de la fraîcheur des informations fournies, de développer les contrôles correspondants, de faire évoluer les services, etc. Cela ne se limite donc pas à une simple modification de schéma de base de données.

**En pratique,  
« gouverner un répertoire »  
revient souvent à  
« rester raisonnable  
dans ses ambitions ».**

<sup>43</sup> Par exemple, décider des organismes habilités à transmettre des mises à jour, du mécanisme des contrôles de flux, de la cinématique des flux. La gouvernance d'une norme d'échange est un sujet en soi, qui engendre des réunions empreintes de technicité : c'est le cas pour la norme EDI-CFE (alimentation de Sirene), pour la norme A (SNGI), pour NEODES (norme de la Déclaration Sociale Nominative).

Un répertoire joue un rôle pivot : c'est un objet-frontière, qui « articule des perspectives d'acteurs appartenant à des mondes sociaux hétérogènes » (*Trompette et Vinck, 2009*)<sup>44</sup>. À ce titre, il est naturellement confronté à de multiples acteurs ayant leurs propres intérêts, leurs propres besoins, qui vont aller à l'encontre de la stabilité et de l'unité de sens. Les instances décisionnelles sont ainsi, naturellement, un lieu de rapport de forces entre d'une part des utilisateurs qui en souhaitent toujours plus, d'autre part les responsables du répertoire cherchant à garantir sa généralité, sa cohérence et sa solidité pour des usages futurs.

## ► Il faut toujours se méfier des impressions superficielles de simplicité

Comme on l'a souligné en introduction, l'entité « répertoire » présente tous les attributs de la simplicité : on imagine une liste d'objets, une liste de courses à faire, etc. et on n'y associe pas *a priori* une quelconque complexité. Dès lors, le fait d'assurer la gestion d'un tel objet peut donner l'illusion de l'évidence. Cet *a priori* est source d'erreur et conduit parfois à de grandes difficultés, voire des échecs : (*Bizingre et alii, 2013*) consacrent ainsi un chapitre au cas du référentiel d'une organisation, décrivant « simplement » la liste de ses unités, et montre pour quelles raisons de tels référentiels sont très difficiles à élaborer. Il apparaît, en particulier, que les unités organisationnelles ne sont pas les mêmes selon les usages, et qu'il s'avère impossible pour un référentiel de satisfaire une trop grande variété de besoins, exprimés ou implicites<sup>45</sup>.



***Ce que l'on stocke, échange et met à disposition, ce ne sont pas des choses inertes, mais de véritables concentrés de sens, mouvants, diversement interprétables selon les acteurs et ayant des impacts colossaux sur des processus métiers.***



Car dans un monde où la donnée ne cesse de prendre de l'importance, les répertoires s'inscrivent dans des systèmes référentiels plus larges<sup>46</sup>, véritables infrastructures de connaissances (*Borgman, 2015*), incluant des nomenclatures ou d'autres répertoires, une standardisation des échanges, un système documentaire, une comitologie (groupe de travail, comité décisionnel), un sous-bassement juridique, le tout étant associé à une « culture référentielle » de l'organisation<sup>47</sup> dont il ne faut pas sous-estimer l'importance. Pourquoi tout cela ?

Ce que l'on stocke, échange et met à disposition, ce ne sont pas des choses inertes, mais de véritables concentrés de sens, mouvants, diversement interprétables selon les acteurs et ayant des impacts colossaux sur des processus métiers.

<sup>44</sup> Cf. page 8 : il y est rappelé que l'article à l'origine du concept fécond d'objet-frontière (*Star et Griesemer, 1989*) en distingue quatre types, l'un d'entre eux étant... le répertoire, un autre étant... le format d'échange.

<sup>45</sup> Le chapitre 5 porte explicitement sur ce sujet.

<sup>46</sup> On peut citer, à titre d'exemple, le système d'information sur le milieu marin (SIMM). L'annexe de l'arrêté du 8 juillet 2019 approuvant le schéma national des données sur le milieu marin (SNDMM) définit le système d'information, la nature des données, explicite la gouvernance, les référentiels, les besoins fonctionnels, etc.

<sup>47</sup> On peut aussi citer le Sandre pour les systèmes d'information de l'eau.

C'est une matière soumise à une double dynamique :

- sur le temps court, à haute fréquence, les événements qui incessamment la modifient (créations, cessations, mises à jour) ;
- sur le temps long, l'évolution de la sémantique, qui en change la texture ; ce que les données portent en elles n'est pas figé, absolu, hors contexte (Boydens, 2000)<sup>48</sup>. Les répertoires, véritables « hubs » de données, portent sur un réel qui n'est ni déterministe, ni facile à appréhender.

Dès lors, rien de plus normal que des réunions techniques ou décisionnelles parfois interminables, avec des consensus difficiles à trouver. Que les projets de répertoire<sup>49</sup> soient longs et coûteux ne doit pas non plus surprendre au vu des sujets à traiter : structuration de la base de données, tour d'horizon des sources d'alimentation, reprise des données passées, transition et bascule vers le nouveau système, élaboration et maintenance de la norme d'échanges, transformation des anciens flux d'entrée, poste de travail du gestionnaire, gamme de services à mettre en œuvre, optimisation du fonctionnement en production... sont le lot commun de ces projets ((Alviset, 2020) et (Demotes-Mainard, 2019)).

Plus généralement, et même si la dimension informatique y joue un rôle majeur en raison du haut degré d'interopérabilité généralement attendu, un projet de réalisation de répertoire ne doit pas être vu en premier lieu comme une opération technique. Car le résultat est avant toute chose un système hautement sensible caractérisé par des finalités générales certes partagées, mais qui se heurtent dans l'opératif à des enjeux individuels qui eux ne le sont pas. C'est donc une exigence permanente de compromis entre les acteurs du répertoire, avec la flexibilité suffisante pour satisfaire des utilisateurs, tout en ne baissant jamais la garde sur la rigueur de conception, les finalités transverses. C'est aussi naturellement un reflet d'enjeux de pouvoir, un « lieu » de confrontation d'intérêts divergents... comme ce fut le cas durant la passionnante aventure qui a conduit à élaborer le système métrique<sup>50</sup>.

<sup>48</sup> Pour reprendre les termes de l'auteur, on ne peut faire « l'hypothèse du monde clos ».

<sup>49</sup> On parle ici de répertoire à statut référentiel, bien entendu.

<sup>50</sup> (Alder, 2005), par exemple p. 28 ou p. 30 : « L'erreur fondamentale des utopistes est de supposer que tout le monde partage la même utopie. ».

## ► Bibliographie

---

- AGENCE DU NUMÉRIQUE EN SANTÉ, 2022. *CI-SIS. Cadre d'Interopérabilité des Systèmes d'Information de Santé*. [Consulté le 7 septembre 2022]. Disponible à l'adresse suivante : <https://esante.gouv.fr/produits-services/ci-sis>.
- ALDER, Ken, 2005. *Mesurer le monde : 1792-1799, l'incroyable histoire de l'invention du mètre*. 24 avril 2005. Trad. M. Devillers-Argouarc'h. Flammarion. ISBN 978-2082103282.
- ALVISET, Christophe, 2020. La troisième refonte du répertoire Sirene : trop ambitieuse ou pas assez. In : *Courrier des statistiques*. [en ligne]. 29 juin 2020. Insee, N° N4, pp. 101-121. [Consulté le 7 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4497083/courstat-4-7.pdf>.
- AMOSSÉ, Thomas, 2020. La nomenclature socioprofessionnelle 2020 : Continuité et innovation, pour des usages renforcés. In : *Courrier des statistiques*. [en ligne]. 29 juin 2020. Insee. N° N4, pp. 62-80. [Consulté le 7 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4497076/courstat-4-5.pdf>.
- BERNARD, Catherine, 1995. Le répertoire Sirene. In : *Courrier des statistiques*. Décembre 1995. Insee, n° 75-76.
- BIZINGRE, Joël, PAUMIER, Joseph et RIVIÈRE, Pascal, 2013. *Les référentiels du système d'information*. Juillet 2013. Dunod. Collection InfoPro. ISBN 978-2100598748.
- BORGMAN, Christine L., 2015. *Big data, little data, no data : scholarship in the networked world*. The MIT Press, 2015.
- BOWKER, Geoffrey C. et STAR, Susan Leigh, 2000. *Sorting things out. Classification and its consequences*. 25 août 2000. The MIT Press. ISBN 978-0262522953.
- BOYDENS, Isabelle, 2000. *Informatique, normes et temps*. Février 2000. Éditions Bruylant. ISBN 978-2802712688.
- BOYDENS, Isabelle, HAMITI, Gani et VAN EECKHOUT, Rudy, 2021. Un service au cœur de la qualité des bases de données. Présentation d'un prototype d'ATMS. In : *Courrier des statistiques*. [en ligne]. 8 juillet 2021. Insee. N° N6, pp. 100-122. [Consulté le 7 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/5398691/courstat-6-art-6.pdf>.
- CAMUS, Benjamin, 2022. Le défi de l'élaboration d'une nomenclature statistique des infractions. In : *Courrier des statistiques*. [en ligne]. 20 janvier 2022. Insee. N° N7, pp. 146-161. [Consulté le 7 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/6035948/courstat-7-art-8.pdf>.
- CROS, Françoise et RAISKY, Claude, 2010. Autour des mots de la formation « Référentiel ». In : *Recherche et Formation*. [en ligne]. ENS de Lyon. N°64-2010, pp. 105-116. [Consulté le 7 septembre 2022]. Disponible à l'adresse : <https://journals.openedition.org/rechercheformation/215#tocto1n7>.
- DÉBARBAT, Suzanne et QUINN, Terry, 2019. Les origines du système métrique en France et la Convention du mètre de 1875, qui a ouvert la voie au Système d'international d'unités et sa révision de 2018. In : *Comptes Rendus Physique*. [en ligne]. Janvier-février 2019. Elsevier. Volume 20, n° 1-2, pp. 6-21. [Consulté le 7 septembre 2022]. Disponible à l'adresse : <https://doi.org/10.1016/j.crhy.2018.12.002>.
- DEMOTES-MAINARD, Magali, 2019. Élire, un projet ambitieux au service du Répertoire électoral unique. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. Insee. N° N2, pp. 58-71. [Consulté le 7 septembre 2022]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4168399/courstat-2-7.pdf>.

- DENIS, Jérôme, 2018. *Le travail invisible des données. Éléments pour une sociologie des infrastructures scripturales*. [en ligne]. Août 2018. Presses des Mines, Collection Sciences Sociales. [Consulté le 7 septembre 2022]. Disponible à l'adresse : <https://doi.org/10.4000/books.pressesmines.3934>.
- DI RUOCCO, Nunzio, SCHEIWILER, Jean-Michel et SOTNYKOVA, Anastasiya, 2012. La qualité des données : concepts de base et techniques d'amélioration. In : *BERTI-ÉQUILLE, Laure, 2012. La qualité et la gouvernance des données au service de la performance des entreprises*. 18 septembre 2012. Hermes Science Publications. PP. 25-54. ISBN 978-2-7462-2510-7.
- DINSIC, 2015. *Référentiel général d'interopérabilité. Standardiser, s'aligner et se focaliser pour échanger efficacement*. [en ligne]. Décembre 2015. Direction Interministérielle du Numérique et du Système d'Information et de Communication de l'État. Version 2.0. [Consulté le 7 septembre 2022]. Disponible à l'adresse : [https://www.numerique.gouv.fr/uploads/Referentiel\\_General\\_Interoperabilite\\_V2.pdf](https://www.numerique.gouv.fr/uploads/Referentiel_General_Interoperabilite_V2.pdf).
- GRATIEUX, Laurent et LE GALL, Olivier, 2016. L'optimisation des échanges de données entre organismes de protection sociale. Rapport IGAS – IGF, février 2016. Disponible à l'adresse : <https://www.igas.gouv.fr/IMG/pdf/2015-090R1.pdf>.
- GUIBERT, Bernard, LAGANIER, Jean et VOLLE, Michel, 1971. Essai sur les nomenclatures industrielles . In : *Économie et statistique*. [en ligne]. Février 1971. Insee. N° 20, pp. 23-36. [Consulté le 7 septembre 2022]. Disponible à l'adresse : <https://doi.org/10.3406/estat.1971.6122>.
- HUMBERT-BOTTIN, Élisabeth, 2018. La déclaration sociale nominative. Nouvelle référence pour les échanges de données sociales des entreprises vers les administrations. In : *Courrier des statistiques*. [en ligne]. 6 décembre 2018. Insee. N° N1, pp. 25-34. [Consulté le 7 septembre]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/3647025/courstat-1-6.pdf>.
- LACOMBE, Anne, 1979. Histoire de l'invention et de la mise en application du système métrique. In : *The French Review*. [en ligne]. Décembre 1979. American Association of Teachers of French. Vol. 53, n° 2, pp. 246-254. [Consulté le 7 septembre]. Disponible à l'adresse : <https://www.jstor.org/stable/390566>.
- MALLARD, Alexandre, 2000. L'écriture des normes. In : *Réseaux*. [en ligne]. Volume 18, n°102, 2000. La fabrication des normes. pp. 37-61. [Consulté le 7 septembre]. Disponible à l'adresse : <https://doi.org/10.3406/reso.2000.2257>.
- OLSON, Jack E., 2003. *Data Quality – The Accuracy Dimension*. [en ligne]. Janvier 2003. Morgan Kaufmann. ISBN (1-1-55860-891-5, 978-1-55860-891-7).
- RÉGNIER-PÉCASTAING, Franck, GABASSI, Michel et FINET Jacques, 2008. *MDM – Enjeux et méthodes de la gestion des données*. Novembre 2008. Dunod. Collection InfoPro. ISBN 978-2100519101.
- RIVIÈRE, Pascal, 2005. Indicateurs de qualité en matière de production de données : quelques éléments de réflexion. In : *Courrier des statistiques*. Septembre 2005. Insee. N° 115, pp. 35-40.
- STAR, Susan Leigh et GRIESEMER, James R., 1989. Institutional Ecology, 'Translations', and Boundary Objects: Amateurs and Professionals on Berkeley's Museum of Vertebrate Zoology, 1907-39. In : *Social Studies of Science*. [en ligne]. 1<sup>er</sup> août 1989. Volume 19, n°3, pp. 387-420. [Consulté le 7 septembre]. Disponible à l'adresse : <https://doi.org/10.1177%2F030631289019003001>.

- SUREAU, Christian et MERLEN, Richard, 2021. Le Répertoire de gestion des carrières unique (RGCU). Un nouveau référentiel ouvrant des perspectives pour l'analyse sociale. In : *Courrier des statistiques*. [en ligne]. 8 juillet 2021. Insee. N° N6, pp. 64-81. [Consulté le 7 septembre]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/5398687/courstat-6-art-4.pdf>.
- TROMPETTE, Pascale et VINCK, Dominique, 2009. Retour sur la notion d'objet-frontière. In : *Revue d'anthropologie des connaissances*. [en ligne]. 2009/1 Vol. 3, n° 1, pp. 5-27. [Consulté le 7 septembre]. Disponible à l'adresse : <https://doi.org/10.3917/rac.006.0005>.
- WALLGREN, Anders et WALLGREN, Britt, 2016. Frames and Populations in a Register-based National Statistical System. In : *Journal of Mathematics and Statistical Science*. Volume 2016, pp. 208-216. [Consulté le 7 septembre]. Disponible à l'adresse : <http://www.ss-pub.org/wp-content/uploads/2016/04/JMSS15121601.pdf>.
- WIENER, Norbert, 1948. *Cybernetics - Or Control and Communication in the Animal and the Machine*. 1961, 2<sup>e</sup> édition. The MIT Press, Cambridge, Massachusetts. ISBN 978-0-262-73009-9.