

ENO, UN GÉNÉRATEUR D'INSTRUMENTS DE COLLECTE

Heidi Koumarianos* et Éric Sigaud**

Eno, c'est l'histoire d'une innovation qui est devenue un maillon important du processus de collecte d'information de l'Insee. Partant d'un concept novateur, l'activation de métadonnées pour industrialiser un processus statistique, Eno en a apporté une application concrète à l'Insee : la génération d'instruments de collecte à partir des métadonnées. La réalisation d'un prototype a permis de valider le concept, ainsi que le choix des standards et de la filière technique. L'objectif de généricité a conduit à mettre en place une architecture souple et évolutive, aujourd'hui utilisée pour produire la quasi-totalité des questionnaires d'enquêtes auprès des entreprises, sous format web et papier. En portant une exigence accrue de formalisation et de standardisation, l'outil a également contribué à une harmonisation des pratiques et, de ce fait, des questionnaires. Le caractère transversal d'Eno a par ailleurs nécessité l'adoption d'une gouvernance inédite pour l'Insee, afin notamment de faciliter son extension à la sphère sociale. Pour cette dernière, de nouvelles fonctionnalités et une évolution technologique seront nécessaires, afin d'outiller la collecte multimodale auprès des ménages. Enfin, les méthodes de développement employées ouvrent des perspectives pour l'utilisation d'Eno par d'autres organismes statistiques.

 *ENO is the story of an innovation that has become an important component of INSEE's data collection process. Starting from an innovative concept, the activation of metadata to industrialize a statistical process, ENO has provided INSEE with a concrete application: the generation of collection instruments based on metadata. The realization of a prototype has made it possible to validate the concept, as well as the choice of the standards and the technical field. The objective of genericity has led to the implementation of a flexible and scalable architecture, which is now used to produce almost all business survey questionnaires, in both web and paper formats. By increasing the need of formalisation and standardisation, the tool has also contributed to the harmonisation of questionnaires and practices. Its transversal nature also required the adoption of a new form of governance for INSEE, in order to facilitate its extension to the social domain in particular. This extension will require the development of new functionalities and technological evolution, to enable mixed-mode data collection for households. The development methods used open the use of ENO to other statistical agencies.*

* Méthodologue, direction de la Méthodologie et de la coordination statistique et internationale, Insee, heidi.koumarianos@insee.fr

** Directeur adjoint délégué au programme Métallica (MÉTadonnées Actives, Logiciels Libres et Infrastructure pour une Collecte Assistée), direction des Statistiques démographiques et sociales, Insee, eric.sigaud@insee.fr

Les métadonnées associées à une enquête ne se limitent pas à des informations facilitant la lisibilité : on trouve aussi des métadonnées « actives », permettant la normalisation et l'industrialisation des processus statistiques (Bonnans, 2019). La collecte des données constitue un domaine propice à la mise en œuvre de cette approche : il s'agit d'un des postes de dépenses les plus importants pour un institut statistique et les processus, souvent répétés, y sont en général bien identifiés et formalisés. À partir de 2012, l'Insee a totalement refondu son système de collecte des enquêtes auprès des entreprises, avec la mise en œuvre du projet Coltrane¹. Ce contexte a constitué une opportunité pour une première mise en œuvre de métadonnées actives pour la génération d'instruments de collecte, à travers l'outil Eno.

ENO, UNE IDÉE

Pour décrire les fonctionnalités d'Eno², on prend un raccourci en parlant de « générer automatiquement des questionnaires ». Cette idée est séduisante en soi : elle permet de réduire les coûts de développement, entraîne de facto une standardisation des pratiques et les gains en qualité afférents, une puissance et une capacité à même de supporter un rythme et une fréquence d'enquêtes plus importants.



Mais derrière Eno, il y a une idée plus générale. Lorsqu'on sait capturer des métadonnées, qu'elles soient relatives au questionnaire, à la gestion ou à la qualité d'un processus, aux produits de diffusion, cela offre potentiellement bien plus que des gains en qualité et en maîtrise du processus statistique : on peut également imaginer de générer un jour jusqu'aux outils nécessaires au déroulé dudit processus.

Pour commencer à concrétiser ce concept, il fallait démontrer sa viabilité sur un cas « simple » : la spécification des questionnaires. Le principe fondateur d'Eno est d'automatiser la production des « instruments de collecte » (formulaires papier ou *web*, modules de collecte des enquêteurs, etc.) à partir d'une description formelle et abstraite du questionnaire sous forme de métadonnées standardisées. On retrouve cette idée dès la première spécification d'Eno, qui prit la forme du schéma en **figure 1**, tracé sur un tableau blanc.

« Le principe fondateur d'Eno est d'automatiser la production des « instruments de collecte ». »

tiser la production des « instruments de collecte » (formulaires papier ou *web*, modules de collecte des enquêteurs, etc.) à partir d'une description formelle et abstraite du questionnaire sous forme de métadonnées standardisées. On retrouve cette idée dès la première spécification d'Eno, qui prit la forme du schéma en **figure 1**, tracé sur un tableau blanc.

La vision d'ensemble était donc de produire automatiquement les supports associés aux différents modes de collecte à partir d'une description formelle³ et unique⁴ du modèle de questionnaire. Eno, figuré par la flèche verte dans la **figure 1**, serait au cœur de cette vision. Dès le début, il était clair également qu'une mise

1. Voir l'article de Jean-Marc Béguin sur la naissance d'une innovation en production statistique dans ce même numéro.
2. En référence à Brian Eno, musicien et producteur anglais qui popularisa notamment le concept de musique générative (https://en.wikipedia.org/wiki/Generative_music). Les outils associés à la plate-forme Coltrane sont généralement baptisés avec des noms de musiciens.
3. S'appuyant sur DDI (voir *infra*).
4. On verra plus loin que l'extension aux enquêtes multimodales de la sphère sociale a conduit par la suite à questionner l'unicité du modèle, sans toutefois la remettre totalement en cause.

en œuvre efficace de celle-ci imposait des choix d'implémentation forts et novateurs, en particulier une architecture de transformation de documents qui n'existait alors pas vraiment à l'Insee.

UN PROTOTYPE POUR VALIDER LE PRINCIPE DE GÉNÉRATION

Le projet Coltrane, plateforme pour la collecte internet auprès des entreprises⁵, offrait les circonstances favorables, les moyens humains et techniques, et un cas concret relativement urgent pour le développement d'un prototype de générateur : la dématérialisation de l'enquête structurelle annuelle auprès des entreprises (ESA). Celle-ci s'appuyait alors sur 200 modèles de questionnaires⁶ dont la réalisation était jusque-là confiée à un prestataire externe.

Le choix de cette enquête était doublement intéressant : d'une part sa complexité était à même de prouver la validité du principe, au-delà de son cas particulier ; et d'autre part, la maîtrise retrouvée du processus de réalisation des questionnaires était susceptible d'offrir un gain financier et en qualité.

Plus largement dans le contexte métier de l'époque, la production de questionnaires pour les enquêtes auprès des entreprises se faisait « à façon », sans capitalisation des développements d'une enquête à l'autre et avec peu de documentation. Il s'agissait d'un processus long (autour de six mois de délai), coûteux (de deux à trois développeurs mobilisés sur la période) et risqué (le caractère « artisanal » pouvait générer des bugs difficiles à identifier).

Figure 1. Eno, une idée : un modèle unique, plusieurs supports de collecte

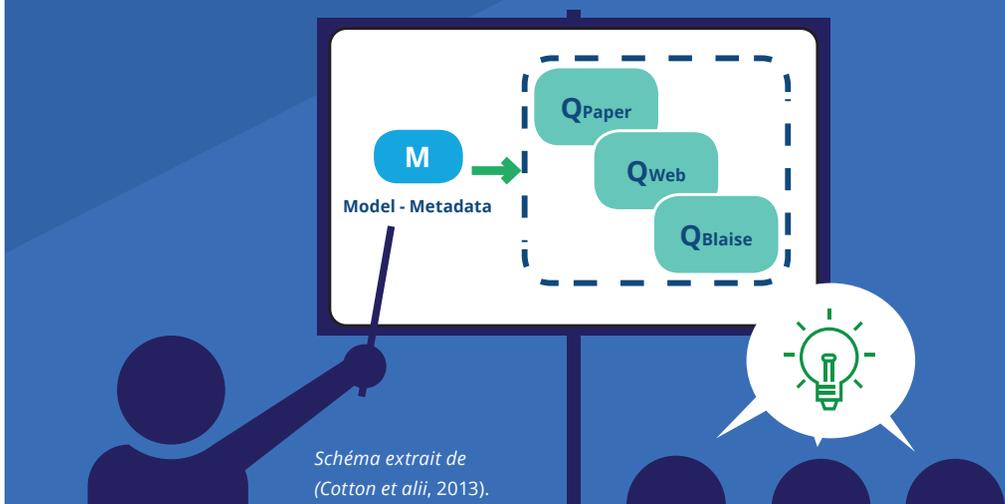


Schéma extrait de
(Cotton et alii, 2013).

5. Voir l'article de d'Olivier Haag et Anne Husseini-Skalitz dans ce même numéro.

6. Le nombre élevé de questionnaires était notamment hérité de l'empilement des spécificités sectorielles, lorsque l'Insee a fondu en un seul processus la quasi-totalité des enquêtes structurelles auprès des entreprises. Depuis, les ESA ne s'appuient plus que sur... 48 modèles.

Il y avait ainsi un intérêt certain à la mise en œuvre d'un générateur permettant d'industrialiser la production des questionnaires : gains en coût et en temps (automatisation), gain en qualité (standardisation, traçabilité), récupération de la maîtrise des spécifications (métadonnées comme « input » du générateur).

« La transversalité de l'outil était présente dès sa conception. »

Après le prototype pour l'enquête structurelle annuelle auprès des entreprises, une première version d'Eno a été développée afin de générer le questionnaire *web* d'une enquête du ministère du Travail⁷, puis enrichie progressivement pour répondre à l'ensemble des besoins des enquêtes intégrant la plateforme Coltrane. Ces enquêtes émanaient de différentes maîtrises d'ouvrage appartenant au Service statistique public⁸. La transversalité de l'outil était présente dès sa conception.

Le générateur a été développé par étapes, et le graphique simpliste des origines, s'il restait le cadre directeur, a évolué au cours des années pour devenir un système d'information riche, couvrant notamment la production de différents formats pour les supports de collecte.

❶ DÉCRIRE FORMELLEMENT LA COMPLEXITÉ SÉMANTIQUE DES QUESTIONNAIRES

Eno a été conçu dès l'origine dans le respect des principes qui fondaient également le développement concomitant du référentiel de métadonnées statistiques RMÉS (Bonnans, 2019) : utilisation des standards internationaux et activation des métadonnées statistiques.

Le questionnaire d'enquête, objet métier central du générateur Eno, est un objet complexe⁹. C'est pour le modéliser que le recours à un standard international a été étudié. En dehors de normes purement techniques, comme XForms¹⁰, seules deux options étaient véritablement à même de couvrir la complexité des questionnaires rencontrés dans la statistique publique : DDI (*Data Documentation Initiative*)¹¹ et SDMX (*Statistical Data and Metadata eXchange*)¹². Curieusement, ces deux standards répondaient au problème de façon complètement différente (**encadré 1**).

On avait donc schématiquement, avec DDI et SDMX, le choix entre d'un côté un standard complexe et sémantiquement précis et de l'autre un standard simple mais sémantiquement faible. Tous deux sont spécifiés par des schémas XML (*Extensible Markup Language*), et peuvent donc être utilisés directement avec les langages de programmation du monde XML, en particulier le langage de transformation de documents XSLT¹³. Tout modèle de

7. Il s'agit de l'enquête Activité et conditions d'emploi de la main-d'œuvre (Acemo) dont la maîtrise d'ouvrage est la direction de l'Animation de la recherche, des études et des statistiques (Dares).

8. Outre la Dares déjà citée, il s'agissait de trois directions de l'Insee : la direction des Statistiques d'entreprises, la direction des Études et synthèses économiques, et de la direction des Statistiques démographiques et sociales.

9. Voir notamment l'article sur l'outil de conception de questionnaires Pogues de Franck Cotton et Thomas Dubois, dans ce même numéro.

10. Recommandation du W3C, l'organisme de normalisation du *web*, pour la représentation et l'utilisation des formulaires *web*. XForms est en fait utilisée par Coltrane pour mettre à disposition les questionnaires des enquêtes par Internet.

11. Voir <https://www.ddialliance.org/>.

12. Voir <https://sdmx.org/>.

13. XSLT (<https://www.w3.org/TR/xslt-30/>) est un langage déclaratif (lui-même exprimé en XML) permettant de transformer des documents XML en d'autres documents XML. Eno est écrit en XSLT.

questionnaire exprimé en DDI ou SDMX est donc « activable » au sens où tout questionnaire conforme à ce modèle peut être passé sans modification en entrée d'une chaîne de traitement. On est donc bien dans les deux cas sur des exemples d'activation de métadonnées : le modèle de questionnaire, en l'occurrence.

Et de fait, en poussant la logique, on peut instruire très naturellement le choix entre DDI et SDMX. L'utilisation active des métadonnées est d'autant plus féconde que ces dernières sont sémantiquement riches, ce qui a conduit clairement à privilégier DDI.

Il est à noter qu'à l'époque, des travaux exploratoires similaires étaient conduits du côté de l'institut statistique australien (ABS) visant à développer un prototype de générateur

Encadré 1. DDI vs SDMX : deux standards, deux approches



DDI, issu essentiellement du monde de la recherche en sciences sociales et des bibliothèques, s'est développé depuis 1995 sous l'impulsion principale de l'Université du Michigan. Initialement axé sur la documentation des études scientifiques et de leurs résultats, DDI s'est étendu à partir de 2007 à la modélisation du cycle de

vie complet des données, de leur collecte à leur publication et leur archivage. Cette version, dite « *DDI Lifecycle* », couvre notamment d'une façon très précise la représentation des questionnaires statistiques, mais au prix d'une verbosité qui peut parfois en rendre l'utilisation difficile. En effet, ce standard, s'appuyant sur le profil d'archiviste ou de chercheur de ses concepteurs et qui se veut un standard de documentation, privilégie l'exhaustivité, la précision et les possibilités de réutilisation et de mutualisation des objets décrits, à la praticité d'utilisation et à la simplicité de représentation.



SDMX, de son côté, est une norme d'origine plus statistique et internationale, dont on peut faire remonter la naissance à une réunion organisée par le Fonds Monétaire International à Washington en septembre 2001. SDMX prenait la relève de travaux précédents sur le message EDIFACT-GESMES*, dont il constituait essentiellement une traduction en

XML (voir (Eurostat, Unece, 2002) pour plus d'information sur l'historique de SDMX). SDMX bénéficie de l'appui de sponsors influents (Fonds monétaire international, Banque centrale européenne, Banque mondiale, Banque des règlements internationaux, Eurostat, etc.) ; en particulier, Eurostat œuvre activement pour en imposer l'utilisation au sein du système statistique européen. SDMX est plus réputé pour l'échange et la diffusion de données tabulaires et des métadonnées structurelles associées. Il comporte également un modèle moins connu dédié aux métadonnées de référence. Ce modèle spécifie pour l'essentiel une structure d'annotations hiérarchiques et reste très générique. Il peut toutefois, au prix d'extensions dédiées (*plug-in* concernant les types, codes, etc.), représenter un questionnaire d'enquête, ou tout autre objet, si on l'adapte différemment. Comparativement à DDI, ce standard, dont les concepteurs avaient un profil plus opérationnel et technique et qui se veut un standard pour de l'échange d'informations, privilégie quant à lui le caractère pratique et générique d'utilisation à la précision et complexité des objets représentés (Arofan et Heus, 2007 ; Heus, Thomas et Vardigan, 2008).

* EDIFACT est un cadre développé par l'Unece et standardisé par l'ISO pour l'échange de données informatisé. GESMES (*Generic statistical message*) est un message EDIFACT pour l'échange de données et métadonnées statistiques.

de questionnaires *web* (XForms) à partir d'une description en DDI¹⁴. Les opportunités de collaboration avec les collègues australiens ont d'ailleurs encouragé l'équipe projet dans le choix du standard DDI et de la technologie XForms.

HARMONISATION INDUITE DES SUPPORTS DE COLLECTE

Le choix d'un standard pour la modélisation des questionnaires, en l'occurrence DDI, était un pré-requis à l'automatisation. La standardisation permet à terme des gains en coût de réalisation et en délai. Mais elle a également des impacts sur le fond et la forme des questionnaires conçus dans ce cadre.

L'utilisation d'un standard formel et sémantiquement précis comme DDI a provoqué des réflexions plus poussées sur la représentation de nos questionnaires : la formulation des questions, la distinction entre questions et instructions, etc.

La phase de conception était auparavant réalisée à l'aide d'outils bureautiques, et autorisait une construction linéaire du questionnement. Certaines parties comportaient parfois plusieurs questions, des éléments d'instruction, des champs de réponse, sans les distinguer formellement. La conceptualisation à l'œuvre dans le standard dévoile ces confusions.

Si le cadre formel n'est pas exempt de contraintes, l'exercice de conceptualisation qui en découle est souvent une opportunité de clarification. Pour exemple, la standardisation a conduit à identifier formellement chaque question, et à numéroter l'ensemble des questions de façon homogène, créant parfois une rupture avec la numérotation « historique » du questionnaire. Or celle-ci est souvent réutilisée dans les applications de gestion d'enquête, et sa modification a nécessité des évolutions applicatives, voire des changements dans les habitudes des gestionnaires d'enquêtes.

“*La conceptualisation plus poussée et la standardisation de la forme concourent toutes deux à une meilleure qualité des questionnaires.*”

La standardisation permet également lors de la phase de conception de normaliser les questionnaires générés : les choix de mise en forme (police, couleurs, espacements, disposition) sont génériques. Il en découle une présentation homogène des questionnaires des différentes enquêtes et une identité visuelle plus reconnaissable pour l'Insee et pour chaque SSM, que renforce par ailleurs la mise en place d'un portail de collecte unique (*figure 2*).

La conceptualisation plus poussée et la standardisation de la forme concourent toutes deux à une meilleure qualité des questionnaires, car elles permettent d'intégrer dans le processus de génération le respect de bonnes pratiques de conception de questionnaires¹⁵. La standardisation impose un changement de pratiques, et une rupture avec le passé, en particulier lorsque les choix de conception antérieurs visaient à optimiser la mise en page pour diminuer le coût de la collecte papier (par l'utilisation de tableaux notamment).

14. Voir le générateur Ramona dont les derniers développements datent de 2013 : <https://github.com/LegoStormtroopr/ramona-ddi-instrument-creator/>.

15. Voir à ce sujet (Haraldsen, Jones, Snijkers et Willimack, 2013), (Christian, Dillman et Morrison, 2010) et (Christian, Dillman et Smyth, 2014).

Elle peut aussi parfois susciter des difficultés d'acceptation, mais *in fine* elle décharge les concepteurs des choix de mise en forme, et leur libère ainsi du temps potentiel pour des tâches de conception à plus forte valeur ajoutée.

Figure 2. Une proximité forte dans l'identité visuelle des questionnaires

The figure displays three screenshots of Insee questionnaires, illustrating visual consistency. The top two screenshots are from the 'Capacité à Innover et Stratégie (CIS) 2018' survey, and the bottom one is from the 'Acemo sur le dialogue social en entreprise 2018' survey. A blue callout box on the right contains the text 'Deux enquêtes Insee au visuel identique' with arrows pointing to the top two screenshots. A red callout box at the bottom left contains the text 'Un visuel pour la Dares qui ne diffère que par la couleur et le logo' with an arrow pointing to the bottom screenshot.

Top Screenshot (CIS 2018): Shows the 'II - INNOVATION' section. The Insee logo is in the top left. A navigation menu on the right includes 'Notice en ligne', 'Exemple de questionnaire', 'Informations de diffusion', 'Contacter l'assistance', and 'Déconnexion'. The progress bar shows 33% completion.

Middle Screenshot (Sectorielle Annuelle 2018): Shows the 'III - EMPLOI' section. It features the same Insee logo and navigation menu as the top screenshot. The progress bar shows 27% completion.

Bottom Screenshot (Acemo 2018): Shows the 'VI : LES ARRÊTS COLLECTIFS DE TRAVAIL EN 2018' section. The Dares logo is in the top left. The navigation menu on the right is colored red and includes 'Notice en ligne', 'Exemple de questionnaire', 'Informations de diffusion', 'Contacter l'assistance', and 'Déconnexion'. The progress bar shows 67% completion.

① UNE ARCHITECTURE CONÇUE POUR UN OUTIL MODULAIRE, ÉVOLUTIF ET EXTENSIBLE

Si l'objectif principal et initial d'Eno était la production automatique de questionnaires *web*, rappelons que la mise en œuvre de métadonnées actives en était la ligne directrice depuis le départ. En particulier, le développement s'est toujours orienté sur des objectifs de généralité, de ré-utilisabilité et d'extensibilité maximales. Ainsi :

- ① Eno permet de produire différents modèles de questionnaire (pour les collectes *web* ou papier) basés sur le standard de modélisation du cycle de vie des données DDI, donc indépendant de la sphère métier « ménage » ou « entreprise » (généricité).
- ① Eno permet d'ajouter des nouveaux formats de sortie et d'entrée, non identifiés au départ, comme un format Open Document de spécification des questionnaires, un format pour la collecte par enquêteur ou pour des besoins post-collecte de reprise des données (ré-utilisabilité et extensibilité).

Cette architecture (**encadré 2**) offre une gestion « douce » des opérations de migration technique. En effet, en cas d'obsolescence d'une technologie d'un format d'entrée ou de sortie, il s'agit de prendre en compte ce nouveau format et d'organiser ensuite la migra-

tion de l'ancien vers le nouveau format. Ainsi par exemple, la montée de version du standard DDI a été réalisée de façon transparente pour les interfaces de sortie. Il est également envisageable de remplacer le format XForms de la filière *web* par des technologies plus récentes comme JavaScript et ce, sans affecter la spécification DDI en entrée.

« Cette architecture offre une gestion « douce » des opérations de migration technique. »

① UN OUTIL UTILISÉ POUR LES ENQUÊTES AUPRÈS DES ENTREPRISES...

Si Eno a vu le jour à la direction des Statistiques d'entreprises (DSE), ce n'est pas un hasard. La statistique d'entreprises offrait en effet un terrain favorable à la naissance, puis à la maturation d'un générateur de supports de collecte. La longue expérience d'enquêtes auto-administrées, et le caractère central du répertoire d'entreprises dans le système d'information ont conduit, au fil des années, à élaborer des processus de production statistique plus structurés et des questionnaires moins complexes que dans la sphère sociale.

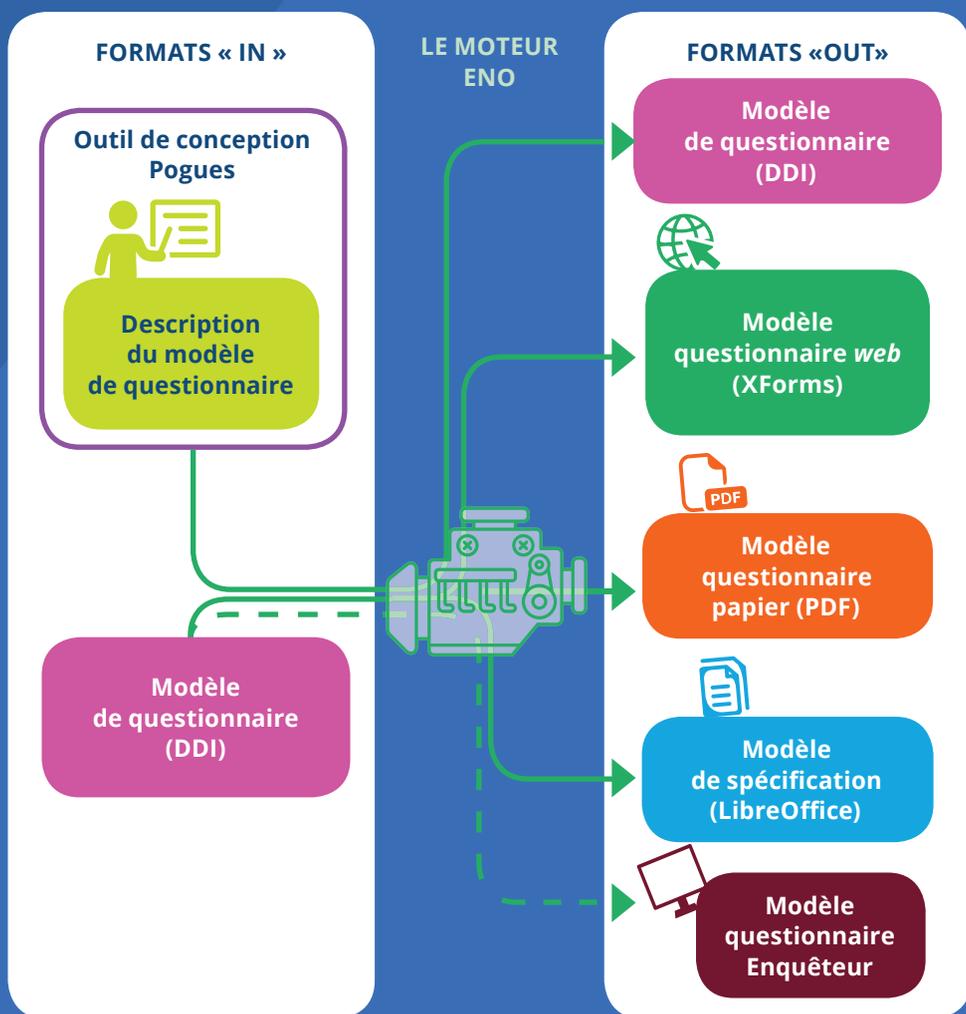
C'est dans ce contexte que le générateur s'est développé. Il est formellement aujourd'hui dans sa version 1. Celle-ci a déjà connu un certain nombre d'évolutions fonctionnelles qui lui permettent notamment de générer la grande majorité des questionnaires *web* des enquêtes entreprises de la filière Coltrane, ainsi que leur équivalent PDF.

Eno permet également de produire des spécifications dans un format bureautique de questionnaires décrits en DDI (**encadré 2**). Si ces spécifications ne concernent pas encore l'intégralité des éléments composant les questionnaires, la couverture actuelle permet déjà de disposer de premières spécifications pour les étapes de conception (relecture de la spécification) et de développement (pour des filières techniques tierces, telle que la filière Blaise).

Encadré 2. L'architecture du moteur Eno facilite la génération de supports de collecte sous plusieurs formats

Plus qu'un simple générateur de questionnaires *web*, Eno est un outil permettant de produire un ou plusieurs formats de sortie, génériquement désigné par « Out », à partir d'un format en entrée dit « In ». Cela est réalisé en s'appuyant sur une configuration déclarative, avec une technique dite de découplage fort (appelée « *separation of concerns* ») entre les formats d'entrée et de sortie, qui consiste à lier les éléments issus du format d'entrée avec ceux du format de sortie dans un fichier tableur, plutôt que sous la forme de code ou de fichiers de configuration technique réservés aux seuls spécialistes. C'est un choix de conception, intégré depuis le début dans le cœur même du moteur Eno, qui garantit une meilleure lisibilité et maintenabilité des fonctionnalités les plus complexes de l'outil.

En poussant cette logique jusqu'au bout, cela permet d'avoir des développeurs experts du format « In » et des développeurs experts du format « Out » (Une évolution du générateur est prévue, afin d'avoir un découplage total des formats d'entrée et de sortie).



📍 ... AU SEIN D'UN ÉCOSYSTÈME BASÉ SUR LES MÉTADONNÉES

La plateforme Coltrane offre les services nécessaires à la collecte *web* et papier, dont certains sont étroitement intégrés avec Eno (*figure 3*) : la plateforme de collecte permet au concepteur de visualiser les supports de collecte *web*, lesquels ont été produits par l'entremise d'Eno sous forme de descriptions techniques ; Coltrane permet ensuite de personnaliser les différents modèles de questionnaires pour les enquêtés.

Il en va de même pour les questionnaires papier : ceux-ci sont produits par un module de génération des courriers (opéré par un groupe dédié aux activités éditiques). À l'instar de la plateforme de collecte *web*, ce module crée l'ensemble des supports de collecte papier à partir du modèle de questionnaires issu d'Eno et du fichier de personnalisation.

📍 DES ÉVOLUTIONS FONCTIONNELLES ENGAGÉES POUR LES MÉNAGES ET AU-DELÀ

Le développement du multimode dans les enquêtes auprès des ménages impose de mutualiser au maximum les développements nécessaires à la conception des supports de collecte, dans la mesure où le nombre de modes mobilisés est plus important que dans le cas des entreprises. Il implique des évolutions des protocoles de collecte et des questionnaires, mais également des outils dédiés à la collecte¹⁶.

L'utilisation d'Eno pour les questionnaires *web* des enquêtes ménages, va conduire à étendre ses fonctionnalités, notamment pour parvenir à générer les boucles de questions fréquemment utilisées dans ce type de questionnaires (par exemple l'articulation ménage/individus).

Ces évolutions sont toutefois mineures au regard de celles requises par l'intégration des enquêtes par enquêteur (face-à-face ou téléphone), qui exige un changement d'architecture et donc de format de sortie. (*figure 3*). En effet, il s'agit notamment de produire des questionnaires embarqués sur des postes enquêteurs, devant pouvoir fonctionner

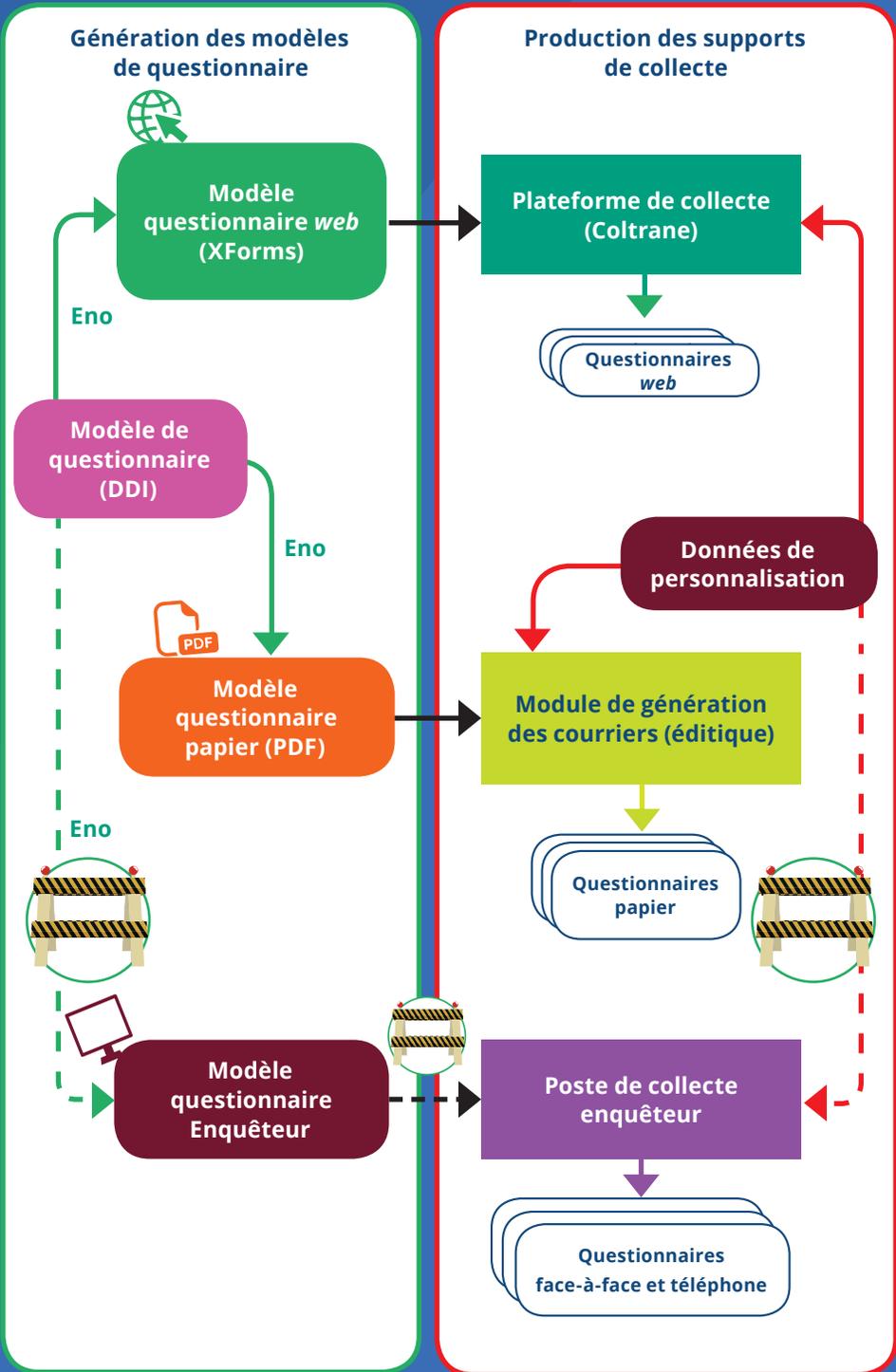
“ Le questionnaire est vu comme un ensemble de composants graphiques modulaires, réutilisables et mutualisables. ”

hors-ligne, quand jusqu'à présent Eno générerait des questionnaires pour une plateforme *web* classique. Ce changement de paradigme technique entraîne nécessairement un changement de technologies et constitue donc l'occasion de proposer une évolution ambitieuse : le questionnaire est vu comme un ensemble de composants graphiques modulaires, réutilisables et mutualisables. Ces composants hébergés dans une application dédiée

constitueront le questionnaire affiché sur le poste de l'enquêteur. Cette approche offre la possibilité de partager les composants avec d'autres outils ou même de les utiliser pour afficher autre chose qu'un questionnaire de collecte à proprement parler (par exemple les questionnaires dans une forme adaptée à l'activité de reprise par un gestionnaire).

16. L'Insee a lancé le programme Métallica (MÉTadonnées Actives, Logiciels Libres et Infrastructure pour une Collecte Assistée) afin d'optimiser le processus de conception des enquêtes ménages, et en particulier des questionnaires.

Figure 3. Eno dans la construction des supports de collecte



Les développements correspondants s'appuient sur une bibliothèque¹⁷ de composants *web* (par exemple un libellé de question, une liste déroulante, un champ de saisie pour les dates, etc.), conforme à l'état de l'art en matière de développement *web*¹⁸, et qui peut être utilisée dans d'autres applications. Une telle réutilisation est déjà engagée dans deux projets :

- ① la refonte de l'indice des prix, qui mettra en œuvre le partage des composants avec une autre application ;
- ① le développement d'un poste générique de traitement des enquêtes : les composants seront utilisés pour afficher des écrans de contrôle de questionnaires.

Ainsi, comme ce dernier cas l'illustre, la description des modèles de questionnaires en DDI, déjà utilisée pour produire les supports de collecte, servira également à générer les écrans relatifs au questionnaire dans l'application de reprise (*figure 4*).

Ces évolutions illustrent la faculté d'Eno à s'adapter à des architectures radicalement différentes, qu'elles soient centralisées (serveur de questionnaires) ou complètement distribuées (poste enquêteur embarqué), à des supports variés (tablettes des enquêteurs prix) et à des contextes applicatifs différents (collecte, traitements aval, etc.).

Enfin, des évolutions techniques sont par ailleurs prévues pour rendre le moteur Eno plus puissant et plus facilement paramétrable.

① UNE GOUVERNANCE TRANSVERSALE POUR FACILITER LA RÉUTILISATION

Une innovation n'émerge pas toujours dans le cadre le plus adapté à son bon fonctionnement en régime courant : le contexte « métier » initial est particulier, il a ses contraintes propres, ses objectifs et son savoir-faire, au sein duquel l'innovation s'apparente d'ailleurs au lancement de projet avec ressources dédiées sans l'être formellement.

Eno a été développé à l'Insee sous la gouvernance de la direction des Statistiques d'entreprises (DSE). Les conditions étaient alors réunies pour permettre le développement d'un générateur de questionnaires destiné en premier lieu à satisfaire un besoin pour des enquêtes auprès des entreprises.

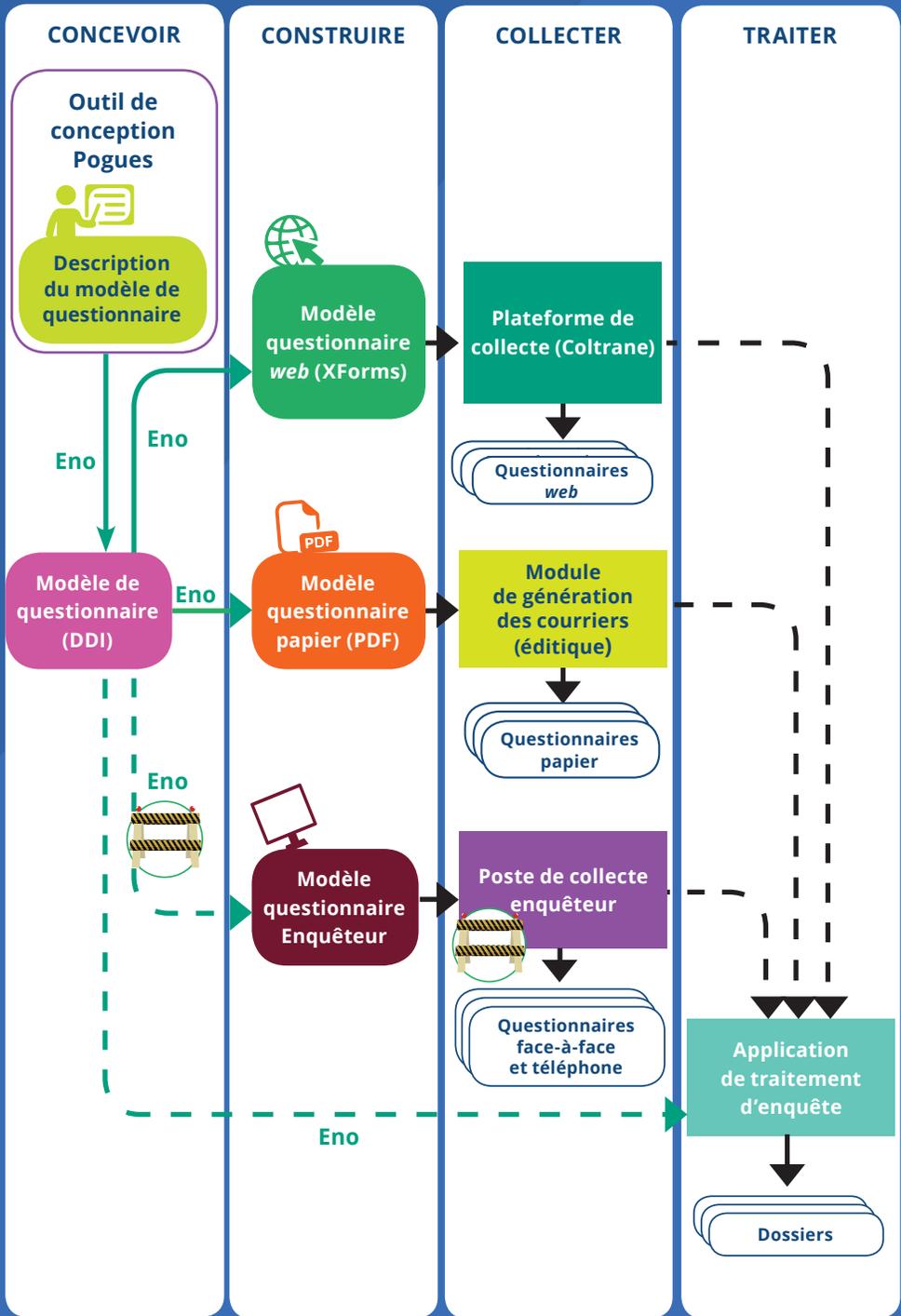
Mais dès le départ, Eno avait été conçu pour être générique, réutilisable et extensible et s'appuyant sur un standard. La question de la gouvernance d'un tel outil, à vocation transversale, s'est donc rapidement posée.

Les conditions initiales n'étaient plus adaptées pour développer un outil transversal dont la portée pouvait aller au-delà du périmètre de la direction métier « mère ». Assez naturellement, la gouvernance d'Eno est donc passée d'une direction métier à une direction plus orientée « conseil/expertise » et donc plus transversale : la direction de la Méthodologie et de la coordination statistique et internationale (DMCSI), déjà maître d'ouvrage du référentiel de métadonnées statistiques RMÉS.

17. La bibliothèque Lunatic est publiée en *open source*, voir <https://github.com/InseeFr/Lunatic>.

18. Par état de l'art, on parle ici des exigences liées à JavaScript, en termes d'architecture et de code.

Figure 4. Eno demain, de la Conception au Traitement ?



Note de lecture : aujourd'hui Eno contribue aux phases de Conception et de Construction (au sens du GSBPM) des supports. Son principe de base le rend éligible à contribuer au-delà, à la phase de Traitement.

Ce changement de gouvernance, assez inédit dans le contexte de l'institut, peut être vu comme une innovation en soi. C'est également une réussite, car il offre aujourd'hui de nouvelles perspectives d'évolutions à l'outil Eno. Ainsi, tout en continuant d'étendre l'offre

« Eno devient peu à peu le moteur de génération des modèles de questionnaires pour la collecte d'enquêtes auprès des ménages voire la collecte des prix. »

de services de la filière des enquêtes auprès des entreprises (par exemple l'ajout du format de sortie pour la collecte papier), Eno devient peu à peu le moteur de génération des modèles de questionnaires pour la collecte d'enquêtes auprès des ménages voire la collecte des prix. Il est aussi utilisé en aval de la collecte, pour des représentations de questionnaires de reprise de données post-collecte (voir *supra*). Mais, même confiée à une unité transversale, la gouvernance d'outils transverses pose de redoutables questions dont les réponses, encore empiriques aujourd'hui, sont à consolider :

comment arbitrer entre des demandes d'évolution émises par des utilisateurs différents ? Comment gérer les montées de version de ces composants transverses, lorsqu'elles ont un impact sur les applications clientes ? Comment financer la maintenance ? On voit que l'innovation technique amène à questionner nos organisations.

🌐 OUVERTURE À L'INTERNATIONAL

Eno a été pensé pour acquérir une dimension internationale, dès l'origine, comme en témoignent les échanges avec les experts de l'office statistique australien ABS, qui ont fourni les premiers éléments de sa conception, puis au fil de ses développements successifs.

Le choix du standard DDI a donné l'opportunité d'effectuer de multiples présentations du générateur dans les colloques consacrés à DDI.

Eno fait partie des outils choisis pour être mutualisés dans le cadre d'un projet européen (ESSnet *Sharing Common Functionalities*). Il a ainsi été réutilisé par SURS, l'INS slovène : cela reste un cas d'école qui n'a pas été jusqu'à une mise en production, mais des contacts existent avec d'autres utilisateurs potentiels.

En parallèle, des travaux d'internationalisation du code se sont poursuivis (traduction en anglais des éléments de commentaires et de documentation de ce code). Il est également publié et partagé en *open source* sur le dépôt GitHub¹⁹ de l'Insee. La communauté des utilisateurs d'Eno devra encore se développer pour que l'institut tire véritablement profit de cette ouverture. Un certain nombre de travaux reste d'ailleurs à conduire pour faciliter la réutilisation d'Eno, indépendamment d'un environnement technique : amélioration du *packaging*, développement de services simplifiant son utilisation, etc.

19. GitHub est une plateforme qui héberge le code d'un logiciel, d'une application, etc. Ce code peut être partagé, s'il est *open source*, et enrichi par des contributions volontaires, les auteurs du code source conservant les droits de validation. Voir <https://github.com/InseeFr> et l'article dans ce même numéro de Franck Cotton et Thomas Dubois sur Pogues.

📍 EVERYTHING THAT HAPPENS WILL HAPPEN TODAY²⁰ (AND FURTHER...)

Eno est désormais la solution de référence à l'Insee pour la création de tous les supports de collecte. Fin 2019, Eno produit presque tous les questionnaires auprès des entreprises, ainsi qu'un premier questionnaire ménages. La mise en œuvre du générateur dans la plate-forme Coltrane a contribué à une industrialisation poussée de la collecte des enquêtes d'entreprises. Cette industrialisation s'est accompagnée d'une large unification des pratiques et d'une certaine « normalisation » des outils. On citera en particulier l'ergonomie des supports de collecte (et plus généralement des documents échangés avec les enquêtés, par exemple les mails ou les courriers) : celle-ci a connu une harmonisation qui a pu conduire parfois à brider la créativité des concepteurs d'enquête²¹, mais qui a largement contribué à professionnaliser la démarche de l'Insee auprès des entreprises.

Eno s'engage maintenant, avec son utilisation dans les enquêtes en multimode auprès des ménages, dans une nouvelle période qui permettra d'enrichir encore ses fonctionnalités. S'il reste une insatisfaction à ce stade, elle se situe au niveau de la réutilisation d'Eno en dehors de l'Insee. Malgré de nombreuses manifestations d'intérêt (différents services statistiques ministériels, la Banque de France, des instituts statistiques étrangers), et malgré les efforts faits pour documenter l'outil, le conditionner sous des formes simplement réutilisables²² et le présenter en de nombreuses occasions, aucun cas de réutilisation n'a été poussé jusqu'en production. On touche là les limites d'une communication peu développée et encore très centrée sur la ré-utilisation interne de nos innovations et nos applications : il n'y a pas à l'Insee d'approche professionnelle en matière de « marketing » des développements informatiques, contrairement à ce que l'on peut observer dans d'autres organisations statistiques.

20. « Tout ce qui peut arriver arrivera aujourd'hui », dernier clin d'œil des auteurs à Brian Eno.

21. Voir sur ce point l'article d'Olivier Haag et Anne Husseini-Skalitz dans ce même numéro.

22. Eno fut le premier développement de l'Insee (et reste à ce jour le seul) à être publié sur le répertoire global de librairies Java (voir <https://mvnrepository.com/artifact/fr.insee.eno>).

▣ BIBLIOGRAPHIE

AROFAN, Gregory et HEUS, Pascal, 2007. DDI and SDMX : Complementary, Not Competing, Standards. In : *site de l'Open Data Foundation*. [en ligne]. Juillet 2007. [Consulté le 16 octobre 2019]. Disponible à l'adresse : http://www.opendatafoundation.org/papers/DDI_and_SDMX.pdf

BONNANS, Dominique, 2019. RMÉS, le référentiel de métadonnées statistiques de l'Insee. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. N°N2, pp.46-55. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4168396/courstat-2-6.pdf>

CHRISTIAN, Leah M., DILLMAN, Don A. et MORRISON, Rebecca, 2010. Questionnaire Design Guidelines for Establishment Surveys. In : *Journal of Official Statistics*. [en ligne]. Vol. 26, N°1, pp. 43-85. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/questionnaire-design-guidelines-for-establishment-surveys.pdf>

CHRISTIAN, Leah M., DILLMAN, Don A. et SMYTH, Jolene D., 2014. *Internet, Phone, Mail, and Mixed-Mode surveys – The tailored design method – 4th edition*. Août 2014. Wiley Online Edition. ISBN 978-1-11-845614-9

COTTON, Franck, SIGAUD, Éric, TAILHURAT, Romain et VAN DER VLIST, Éric, 2013. *XForms generation: a real world example*. [en ligne]. 5 août 2013. International Symposium on Native XML User Interfaces. Balisage Series on Markup Technologies, Vol. 11. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://www.balisage.net/Proceedings/vol11/html/Cotton01/BalisageVol11-Cotton01.html>

EUROSTAT, UNECE, 2002. *Common Open standards for the Exchange and Sharing of Socio-economic Data and Metadata : the SDMX Initiative*. [en ligne]. 6-8 mars 2002. Work Session on Statistical Metadata, Working Paper N°11. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <https://sdmx.org/wp-content/uploads/wp111.pdf>

HARALDSEN, Gustav, JONES, Jacqui, SNIJKERS, Ger et WILLIMACK, Diane K., 2013. *Designing and conducting business surveys*. 26 juillet 2013. Wiley Online Edition. ISBN 978-0-47-090304-9

HEUS Pascal, THOMAS, Wendy et VARDIGAN, Mary, 2008. Data Documentation Initiative : Toward a Standard for the Social Sciences. In : *The International Journal of Digital Curation*. [en ligne]. N°1, Vol. 3, pp. 108-113. [Consulté le 16 octobre 2019]. Disponible à l'adresse : <http://www.ijdc.net/article/view/66/45>