

OECD Meeting of the Group of Experts on Business Registers

Genève, 30 septembre au 2 octobre 2019

Franck Cotton, Anne Hussein-Skalitz, Olivier HAAG

Insee

Session No. 5. Web portals for business respondents

De la métadonnées à la données : la nouvelle filière française pour collecter des enquêtes auprès des entreprises

Introduction

Un projet de refonte du système de collecte par internet des enquêtes auprès des entreprises a été mis en œuvre par Institut National de statistique français dans les années 2010.

La plate-forme de collecte résultant du projet Coltrane (Coltrane (COLlecte TRANsversale d'Enquêtes) a ainsi non seulement pris la place de quatre anciens dispositifs de collecte, mais elle a aussi été le premier maillon d'une nouvelle filière de l'Insee basée sur des métadonnées actives. Ainsi, à partir de la description du questionnaire par le concepteur, différents supports de collectes sont générés automatiquement et intégrés dans une infrastructure de collecte par Internet.

Plus qu'une simple plate-forme de collecte, Coltrane aujourd'hui c'est une offre de service complète qui permet aux concepteurs d'enquête de collecter les réponses par Internet bien entendu mais aussi de disposer d'un référentiel de contacts au sein des entreprises et de pouvoir envoyer des courriers, des mails et des questionnaires papier aux entreprises qui le souhaitent tout en disposant d'un système d'assistance dédié. Enfin, Coltrane c'est aussi une avancée notable pour les entreprises qui peuvent maintenant répondre à la plupart des enquêtes qui leur sont adressées par la statistique publique (aussi bien de l'Insee donc que des SSM) grâce à un unique couple « identifiant - mot de passe », et tout cela dans un cadre ergonomique commun.

Une nouvelle plate-forme basé sur les « métadonnées actives »

La principale originalité de ce projet et qui le distingue d'autres plates-formes existantes (Saraiva P. et al, 2016, Fazio et al 2015, Thieme et al 2015) était de reposer dès l'origine sur des liens forts avec Référentiel des Métadonnées Statistiques de l'institut et notamment d'utiliser les « métadonnées » décrivant les questionnaires d'enquête (liste et enchaînement des questions, modalités possibles des réponses, contrôles à la saisie, filtres, etc.) pour générer les outils de collecte directement à partir de cette description (d'où le nom de métadonnées « actives »). Cette description se base sur un standard

international de description des opérations statistiques et de leur documentation, progressivement adopté par les INS. L'Insee a ainsi adopté le modèle GSBPM¹ pour modéliser ses processus statistiques et le standard « DDI² » pour décrire formellement le cycle de vie des données, dont les objets « questionnaire ». Ce choix structurant a rendu impossible le recours à un progiciel de collecte du marché déjà existant car ceux-ci n'étaient pas encore adaptés à ces standards. La génération automatique des supports de collecte (c'est-à-dire des écrans et des programmes qui permettent de récupérer les réponses, à partir de leur description en DDI) est à la racine des gains de productivité escomptés.

Cette filière se base donc sur 4 principaux outils :

- Un référentiel de métadonnée du nom d'Rmès (voir figure 1 ci-dessous)
- Une interface « user friendly » de saisie des métadonnées du questionnaire d'enquête (libellés des questions, liste des modalités de réponses, contrôles, filtres, etc.). Ces informations sont ensuite traduites et enregistrées en DDI dans le répertoire des métadonnées. Cette application s'appelle Pogues dans la figure n°1.
- Un générateur de questionnaire qui traduit le DDI issu de Pogues en différents formats de support de collecte (Cotton 2013) :
 - o Transformation du DDI en Xforms pour le questionnaire au format web ;
 - o Transformation du DDI en Fo pour le questionnaire au format papier ;
 - o Transformation du DDI en Json pour le questionnaire au format de l'application de contrôle des données. Cette application est à destination des gestionnaires d'enquêtes de l'institut qui contrôlent et valident les réponses des entreprises aux enquêtes ;
 - o Transformation du DDI en fodt pour obtenir un document permettant de garder une trace des spécifications de l'enquête (contrôle notamment).

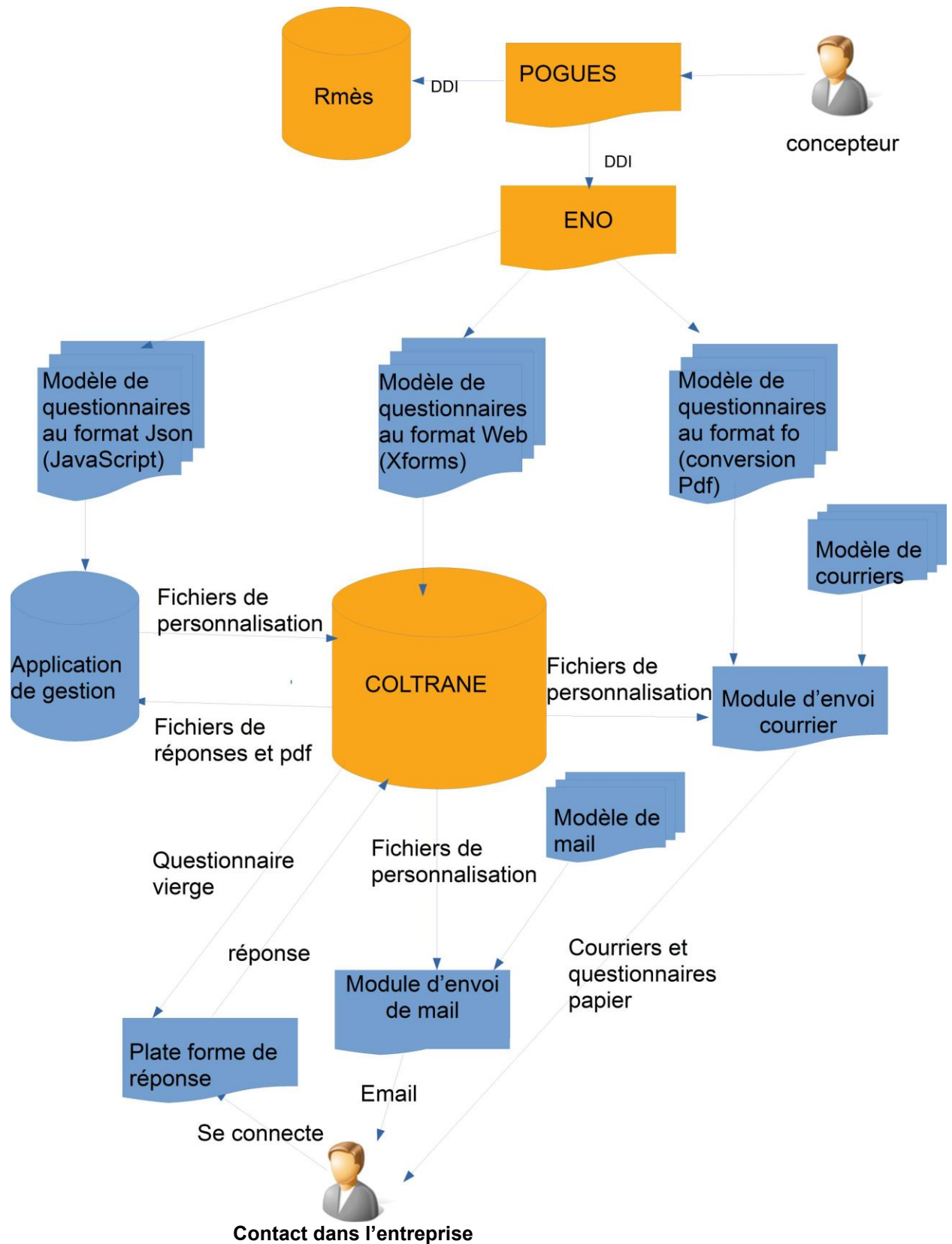
Ce générateur s'appelle ENO sur la figure n°1. Il est directement lié à Pogues, ce qui permet au concepteur d'enquête qui utilise Pogues de visualiser directement les questionnaires web et papier qui correspondent à sa spécification.

- Une plateforme de collecte, appelée Coltrane dans la figure 1, qui permet de collecter les réponses des entreprises échantillonnées dans l'enquête. Les questionnaires collectés par Coltrane sont la concaténation du modèle de questionnaire issu d'ENO « personnalisé » par les caractéristiques de l'unité enquêtée (adresse, nom du contact, activité principale etc.).

¹ The Generic Statistical Business Process Model (GSBPM)

²Data Description Initiative

Figure 1 : Présentation de la filière de collecte



Pogues un outil de conception de support de collecte

Le DDI n'est pas un langage facile à écrire. Il était donc nécessaire de mettre en place une interface permettant au concepteur de décrire simplement son questionnaire puis de convertir cette description en DDI. C'est le rôle de Pogues.

Pogues permet en effet au concepteur de structurer son questionnaire en séquences (modules et sous-modules) regroupant des questions et des éléments textuels, ainsi que de définir la logique de parcours (enchaînement des questions, etc.) et les contraintes d'ensemble pour le remplissage. Différentes sortes de questions sont supportées : simple à choix unique ou multiple, tableau etc., ainsi que différents types pour les réponses : texte, date, nombre ou booléen, les valeurs pouvant être contraintes par des bornes ou des listes de codes prédéfinies.

Outre les séquences et les questions, Pogues permet de spécifier différents éléments textuels : commentaires, consignes ou aide pour le remplissage, messages d'avertissement issues de contrôles, etc. Selon le type de support de collecte qui sera produit, ces éléments apparaîtront (ou pas) sous des formes variées, par exemple des infobulles sur les formulaires web.

Il faut noter que la structure formelle de questionnaire qui est ainsi spécifiée, avec les questions et les variables associées, peut être réutilisée plus en aval dans le déroulement des opérations statistiques, et en particulier lors de la phase de contrôle et de reprise des données collectées

La figure 2 ci-dessous est une copie d'écran de Pogues.

Dans cet exemple le concepteur a créé :

- une question « In which city do the Simpsons reside » à choix unique et avec une liste de modalité « TOWN » qu'il a créé par ailleurs ;
- une consigne « One possible answer » saisi dans l'onglet « déclarations » ;
- une précision sur le mot réside « choose the city of the main residence » ;
- un filtre : « If you do not know the answer please go to the question 5 », saisi dans l'onglet "redirections".

Figure 2 : Copie d'écran de la définition d'une question dans Pogues

The screenshot displays the 'MODIFIER LA QUESTION' (Edit Question) interface in the Pogues application. The browser window shows the URL <https://conception-questionnaires.insee.fr/#/questionnaire/i6vw21>. The question being edited is 'In which city do the Simpsons reside?'. The identifier field contains 'CITY'. Under 'Mode de collecte' (Collection mode), 'CAWI' is selected. The 'Format des réponses' (Response format) tab is active, showing 'Réponse à choix unique' (Unique choice response). The 'Type de question*' (Question type) is set to 'Réponse à choix unique'. The 'Obligatoire' (Mandatory) checkbox is unchecked. The 'Type de saisie*' (Input type) is set to 'Bouton radio' (Radio button). The 'Spécifier la liste de codes*' (Specify the list of codes) section shows 'Retrouver dans le questionnaire' (Retrieve from questionnaire) selected. The 'Spécifier la liste de codes*' field contains 'TOWN'. At the bottom, there are 'ANNULER' (Cancel) and 'VALIDER' (Validate) buttons. The taskbar at the bottom shows various application icons and the system clock indicating 15:31 on 18/07/2019.

Par ailleurs Pogues est interfacé avec le répertoire de métadonnées ce qui permet au concepteur de réutiliser des questions déjà existantes. De fait, le concepteur a, à sa disposition, une banque de questions qu'il peut enrichir le cas échéant.

Pogues propose donc une interface graphique pratique et intuitive pour décrire le questionnaire et ses composants. À tout moment, l'utilisateur peut visualiser son questionnaire sous la forme de son choix (web ou PDF), l'exporter en DDI pour le stocker dans le référentiel de métadonnées ou produire le document de spécification détaillée qui lui correspond. Ces transformations sont faites par le biais d'ENO.

Eno : où comment donner vie aux métadonnées

Eno a pour but de produire automatiquement les supports associés aux différents modes de collecte à partir de la description formelle unique du questionnaire sous-jacent (en DDI) faites dans Pogues. En

pratique, Eno autorise une grande variété de formats en entrée et en sortie, à partir du moment où ces formats s'expriment en XML.

Les figure 3 et 4 ci-dessous montre les sorties au format web et papier générés par ENO et correspondant à la définition faite dans Pogues de la figure 2.

On constate que dans le format web que :

- la consigne (« One possible answer ») se place sous le libellé de la question ;
- la précision (« choose the city of the main residence ») est matérialisée par une infobulle ;
- le filtre est géré par la cinématique du questionnaire web et n'est donc pas visible sur la copie d'écran.

Figure 3 : Questionnaire web correspondant à la question de la figure 2 spécifiée dans Pogues

The screenshot displays a web browser window with the URL <https://visualisation-questionnaires.insee.fr/messtromae/fr/esa-dc-2018/simpsons/new?unite-enquete=123456789>. The page header includes the Insee logo and the title 'Questionnaire SIMPSONS'. A progress bar labeled 'Avancement' is located at the top right. The main content area is titled 'I - SIMPSONS' CITY' and contains the text 'This module asks about your general knowledge of the Simpsons city'. The question '1. In which city do the Simpsons reside?' is followed by a tooltip that reads 'choose the city of their main residence'. Below the question, there are three radio button options: Springfield, Shelbyville, and Seinfeld. At the bottom right, there are two buttons: 'Retour' and 'Enregistrer et continuer'.

On constate que dans le format papier :

- la consigne se place sous le libellé de la question ;
- la précision est matérialisée par une astérisque. En revanche le libellé de cette précision ne figure pas dans le questionnaire par convention mais est présent dans la notice qui l'accompagne.
- Le filtre (« If you do not know the answer please go to the question 5 ») apparaît bien en clair.

Figure 4 : Questionnaire papier correspondant à la question de la figure 2 spécifiée dans Pogues

II - Simpsons' city

This module asks about your general knowledge of the Simpsons city

➔ **1. In which city do the Simpsons reside*?**

One possible answer

☐ Springfield

☐ Shelbyville

☐ Seinfeld

➔ **If you do not know the answer please go to the question 5**

2 / 7

Eno : Un progiciel en open source

Dès sa conception, Eno a eu une dimension internationale (les premiers éléments de sa conception proviennent d'échange avec les experts d'ABS). Durant son développement, cette dimension a été étendue.

L'accrochage d'Eno au standard DDI a donné l'opportunité d'effectuer de multiples présentations de l'outil dans les colloques consacrés à DDI.

Eno était l'un des outils choisis dans le cadre de l'ESSnet Sharing Common Functionalities et fait l'objet d'un cas de réutilisation par SURS, INS slovène.

COLTRANE : la plateforme qui permet de recueillir les réponses aux questionnaires

Une fois que le concepteur a validé la version définitive de son questionnaire vierge, Coltrane génère et héberge tous les questionnaires web des unités à interroger, à partir de ce modèle de questionnaire vierge et non personnalisé en quelque sorte fourni par Eno et de l'échantillon de l'enquête avec ses données de personnalisation mises en forme ³ (au format XML) par l'**application de gestion de l'enquête**⁴ (cf. figure 1). Au moment du lancement de la collecte, Coltrane a donc généré autant de questionnaires vierges (mais personnalisés) qu'il y a d'unités interrogées.

Une fois la collecte d'une enquête ouverte, Coltrane extrait les réponses qui ont été collectées depuis la dernière extraction, au rythme choisi par le concepteur d'enquête (actuellement dans les faits jusqu'à 4 fois par jour) et les transmet (au format XML) à l'application de gestion de l'enquête. Certaines applications ont également choisi de récupérer les PDF preuves de dépôt.

Une offre de service aux entreprises enquêtées

L'entreprise (ou contact sur la figure 1) reçoit un courrier (voire un courriel) lui annonçant sa sélection dans un échantillon d'enquête et donc la nécessité pour elle de répondre à un questionnaire. Ce courrier, en plus de présenter succinctement l'enquête, contient des éléments pour se connecter à la plate-forme (identifiant voire mot de passe).

Le contact⁵ s'authentifie alors sur le **portail d'authentification** via la saisie de son identifiant et d'un mot de passe. À la première authentification, le contact doit personnaliser son mot de passe. En cas de perte du mot de passe, le contact peut demander son renouvellement en ligne.

Une fois authentifié, le contact est dirigé sur le portail « **Mes Enquêtes** ». C'est le point d'entrée unique pour répondre à toutes les enquêtes gérées par Coltrane (enquête de l'Insee mais aussi du service statistique du ministère en charge de l'emploi (Dares)). Il y voit les enquêtes en cours de collecte pour lesquelles il a les droits de réponse, ainsi que les dates butoir de retour souhaitées, voire les dates de

³ Ce fichier dit « de personnalisation » contient pour chaque unité enquêtée :

- l'ensemble des données d'identification des entreprises tirées (identifiant, raison sociale, Siren, etc.)
- éventuellement les contacts (identifiant (s'il existe déjà dans Coltrane et qu'on veut qu'il soit mutualisé), adresse, nom-prénom, fonction, téléphone, mail, etc.)
- les données dites de personnalisation du questionnaire. Il peut s'agir par exemple de réponses fournies par l'unité enquêtée lors d'une précédente campagne de collecte.

⁴Cette application est en dehors du projet Coltrane et différente pour chaque enquête aujourd'hui. Elle est de la responsabilité de chaque maîtrise d'ouvrage d'enquête et permet de gérer la collecte et de contrôler les réponses aux questionnaires.

⁵Il s'agit d'une personne physique qui va répondre à l'enquête pour une unité enquêtée via le portail de collecte. Une unité enquêtée a au moins un contact (dont les noms et prénoms peuvent ne pas être connus au lancement de l'enquête) et peut en avoir plusieurs. Un même contact peut être associé à plusieurs unités enquêtées. Le contact peut ne pas appartenir à l'unité enquêtée (cas du comptable extérieur qui répond aux enquêtes Insee pour ses clients). À un contact est associé un identifiant et un mot de passe.

réponse si le questionnaire a déjà été renseigné. Le contact accède au questionnaire de son choix parmi la liste des enquêtes soit pour répondre à l'enquête, soit pour consulter des questionnaires déjà saisis (tant que la collecte n'est pas fermée).

Lorsqu'il navigue dans **un questionnaire**, ses réponses sont automatiquement enregistrées à chaque changement de page. Il peut ainsi quitter le questionnaire à tout moment et y revenir quand il le souhaite : il lui est alors proposé de retourner directement sur la dernière page consultée (ou de recommencer depuis le début). En fin de questionnaire il peut choisir d'envoyer sa réponse ou de simplement l'enregistrer (pour y revenir ultérieurement ou la soumettre à quelqu'un d'autre). En cas d'envoi, un téléchargement au format « PDF » récapitulant sa réponse lui est proposé.

En outre, depuis ce portail de collecte, le contact a bien sûr la possibilité de renseigner ou de mettre à jour ses informations personnelles (nom, prénom, adresse mél, numéro de téléphone, fonction, adresse postale, etc.) par le biais de l'onglet « **Mon Compte** ». Sur ce dernier il peut également personnaliser son mot de passe et surtout, dans le cas où il aurait reçu plusieurs identifiants de connexion (un par enquête par exemple), il peut « regrouper » ses identifiants et disposer d'un unique compte pour accéder et répondre aux questionnaires de plusieurs unités enquêtées et/ou plusieurs enquêtes.

Par ailleurs, tout au long de son chemin sur la plate-forme de collecte, il peut accéder à une foire aux questions et à une **assistance en ligne**, via un formulaire. Cette assistance est contextualisée, c'est-à-dire que selon l'endroit d'où le formulaire est appelé, il est plus ou moins pré rempli et peut être transmis à deux acteurs :

- Insee Contact (gestionnaires de l'Insee en charge de répondre aux internautes), pour répondre aux questions dites « techniques » (modalité d'accès à la plate-forme de collecte, réalité de l'enquête, objectifs, perte de code d'accès, etc.) ;
- les équipes de gestionnaires d'enquêtes, pour répondre aux questions « métier » directement en lien avec l'enquête.

Un module courrier pour gérer les relations avec les entreprises

Pour mener à bien la collecte d'une enquête, il est indispensable de contacter régulièrement les entreprises afin de leur rappeler leur obligation de répondre aux enquêtes labellisées de la statistique publique. Traditionnellement, quatre courriers sont prévus :

- courrier d'ouverture de la collecte ;
- rappel ;
- mise en demeure ;
- constat de non-réponse.

Un outil dit « module courrier » (cf. figure 1) permettant de créer des courriers « prêts à être imprimés » a été développé dans le cadre de Coltrane, et il est aujourd'hui mis à disposition de toute autre application ayant besoin d'envoyer des courriers papier. Le contenu générique de ces courriers a été standardisé. Le contenu spécifique à chaque enquête (nom de l'enquête, objectif, etc.) est composé à l'aide de

métadonnées stockées dans les bases Coltrane et alimentées par le concepteur d'enquête (toujours selon le même principe de métadonnées actives).

À partir d'un fichier XML de personnalisation (qui contient la liste des unités pour lesquelles le courrier doit être envoyé, le type de courrier à envoyer et les éléments de personnalisation de l'unité) fourni par l'application de gestion de l'enquête, Coltrane produit un fichier transmis au module courrier : celui-ci créera alors, à partir des modèles de courrier qui y sont décrits (ouverture de la collecte par exemple) des courriers au format PDF.

L'envoi de mail, un plus pour encourager la réponse des entreprises

En plus de l'envoi de courriers, Coltrane permet d'envoyer, lors des ouvertures de collecte ou des rappels, des messages électroniques aux contacts disposant d'une adresse mail, les invitant à répondre à l'enquête. La logique en est la même que pour l'envoi des courriers (cf. figure 1) : à partir de modèles de mail prédéfinis et d'un fichier XML de personnalisation (qui contient la liste des unités pour lesquelles le mail doit être envoyé et le type de mail à envoyer) fourni par l'application de gestion de l'enquête, Coltrane génère un mail qui est envoyé via l'outil d'envoi de mail automatique de l'Insee. Dans ce mail figurent des éléments sur l'enquête concernée (nom, objectifs, date butoir de retour attendu, etc.), des extraits des textes légaux relatifs aux enquêtes mais surtout un rappel de l'identifiant du contact, le lien pour accéder au portail d'authentification et un lien en cas d'oubli de mot de passe, afin de réduire les demandes d'assistance. Par convention, l'émetteur de ces mails est le service collecteur de l'enquête, ce qui permet à l'enquêté d'avoir une assistance facile à contacter.

En outre, à terme, il sera également possible d'envoyer le questionnaire au format PDF par mail à une entreprise qui le souhaite.

Une application interne pour gérer les contacts

En plus d'être une plate-forme de collecte conviviale pour les répondants et d'offrir des outils pour les concepteurs d'enquêtes, Coltrane est aussi un référentiel de contacts pour les enquêtes entreprises.

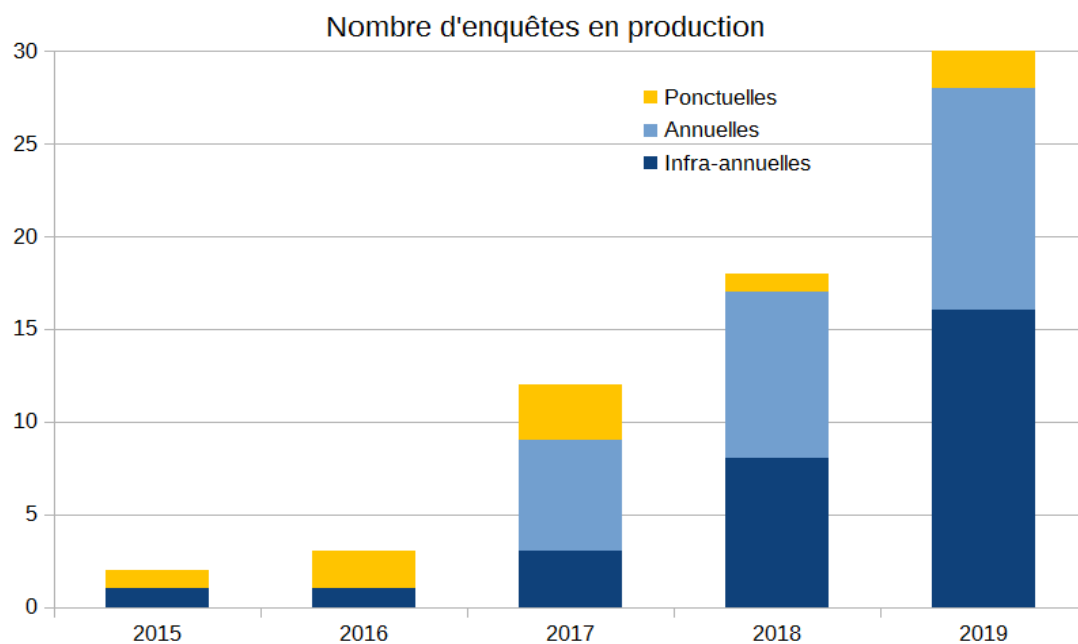
Pour ce faire, lorsqu'une enquête intègre le dispositif, Coltrane importe massivement les contacts connus de l'enquête (en particulier lorsqu'elle recourait déjà à la collecte par Internet). Ensuite, afin de faire vivre ce référentiel, une application interne de « **gestion des contacts** » est mise à disposition des gestionnaires d'enquête qui sont au contact des entreprises lors de la collecte des enquêtes. Avec cette application, ils peuvent par exemple créer un contact, modifier ses caractéristiques, lui donner ou lui enlever le droit de répondre à une enquête, etc. Lorsqu'un gestionnaire modifie une caractéristique d'un contact répondant à plusieurs enquêtes, les responsables des autres enquêtes sont prévenus par mail. L'application est également accessible par l'équipe Insee Contact pour aider l'enquêté (perte de code d'accès par exemple).

Par ailleurs, les gestionnaires d'une enquête peuvent visualiser les questionnaires de leurs enquêtés quel que soit leur statut (déjà transmis à l'Insee ou pas). Ceci peut notamment permettre d'aider les entreprises qui ont des difficultés à répondre ou vérifier l'envoi ou le simple enregistrement d'un questionnaire.

Coltrane : un millionnaire en devenir

Coltrane a atteint en 2019 son régime de croisière avec l'intégration de toutes les enquêtes entreprises conduites par l'Insee et par la Dares, proposant une collecte par Internet. Comme le montre le graphique ci-dessous, le nombre d'enquêtes collectées par Coltrane ne cesse de croître au fil des ans. En outre, des contacts sont en cours avec deux autres services statistiques ministériels. L'objectif est d'essayer de collecter via Coltrane un maximum d'enquêtes du système statistique public de façon que les entreprises qui répondent à plusieurs enquêtes n'aient qu'une seule infrastructure à connaître. Le million annuel de questionnaires⁶ mis en collecte dans Coltrane sera alors dépassé. En 2019, il y aura en effet un peu plus de 900 000 questionnaires accessibles sur la plate-forme et on devrait dépasser pour la première fois la barre des 500 000 questionnaires retournés. Depuis 2015, plus de 615 000 contacts ont été enregistrés dans le référentiel.

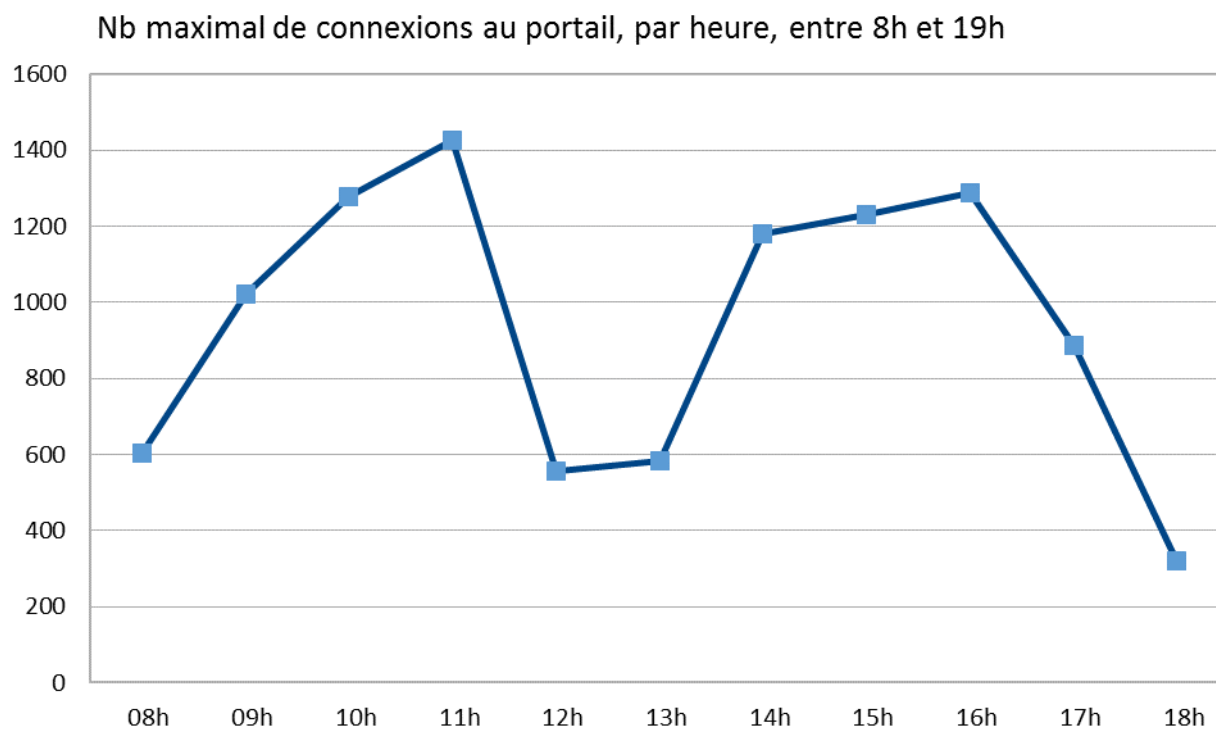
Figure 5 : Nombre d'enquêtes (par type) collectées dans Coltrane depuis 2015 :



L'activité du site n'est pas uniforme sur une journée. On constate des pics de collecte quotidiens en fin de matinée et en fin d'après-midi comme le montre la figure ci-dessous.

⁶En raison de la fréquence de certaines enquêtes.

Figure 6 : Nombre maximal de connexions par heure entre 8 h et 19 h en 2019



Lecture : Le maximum de connexions quotidiennes observé dans le courant de l'année 2019 entre 11 h et midi est de 1400. C'est aussi le maximum jamais atteint sur une heure dans une journée en 2019

En guise de conclusion

La filière mise en œuvre à l'Insee permet aujourd'hui de standardiser et industrialiser la collecte par internet des enquêtes entreprise. Elle facilite en outre grandement la tâche du concepteur d'enquête en lui permettant d'éviter du travail en double voire triple qu'il faisait par le passé. Là où il devait spécifier son questionnaire, papier, puis son questionnaire papier puis les écrans de son application de gestion de l'enquête, il lui suffit aujourd'hui de saisir dans Pogues la description des métadonnées de son enquête qui permettent ensuite par le biais d'ENO de disposer des 3 éléments décrits ci-dessus.

Par ailleurs, l'ensemble de services offert par Coltrane permet à l'Insee d'améliorer son identité visuelle vis-à-vis des grandes entreprises qui reçoivent un nombre important de questionnaires d'enquêtes statistiques. Ces dernières voient d'une part se réduire le coût d'appropriation du remplissage des questionnaires et de la navigation sur la plate-forme de collecte. D'autre part elles peuvent être incitées à répondre aux enquêtes qui figurent dans leurs en-cours et dont elles avaient oublié l'existence. Si l'on n'a pas encore observé d'accroissement des taux de réponse, au moins le système permet-il de maintenir les taux existants à des niveaux plutôt élevés par rapport aux autres pays, dans un contexte global où ils ont plutôt tendance à s'éroder.

Bibliographie

COTTON F and al. "XForms generation: a real world example", International Symposium on Native XML User Interfaces 2013, <https://www.balisage.net/Proceedings/vol11/html/Cotton01/BalisageVol11-Cotton01.html>

SARAIVA P., MOREIRA A. « Managing an integrated respondent communication : statistics Portugal experience » Q2016 Madrid

FAZIO Natale Renato and MOTTURA Giulia « The Business Portal - an integrated system to collect data and communicate (Italy) », Workshop on Statistical Data Collection: Riding the Data Deluge, 2015

THIEME, Michael T., MATHUR Anup « Designing and Architecting a Shared Platform for Adaptive Data Collection in Surveys and Censuses », US Census, 2015