

Article

Elaboration d'un protocole de préparation pour la numérisation des 3061 registres de la série des Pièces Originales du département des Manuscrits de la BnF à partir d'une étude par sondage du fonds.

Development of a conservation protocol to digitize the 3061 "Pièces Originales" registers thanks to a collection survey

Pauline Elie ^a, Caroline Laffont ^b, Isabelle Bonnard ^c, Serge Cohen ^d, Mathieu Lescuyer ^e

^a Chargée de la préservation, de la régie et de la restauration des collections, Saint-Denis, musée d'art et d'histoire Paul Eluard

^b Chargée d'études et de recherche en biologie, département de la Conservation, BnF

^c Experte en restauration, département de la Conservation, BnF

^d Chargé de recherche, IPANEMA, Cnrs

^e Adjoint au directeur du département des Manuscrits, BnF

Mots-clés: Pièces Originales
– Numérisation – Sondage
– Logiciel R – Etude de collection

Keywords: Pièces Originales
– Digitization – Poll – R
Software – Collection survey

Sommaire

Introduction : présentation du fonds et contexte historique	1
Contraintes imposées par le manque de visibilité sur un fonds volumineux	1
1. Réalisation d'une évaluation de l'état de la série par sondage	2
2. Exploitation des données collectées à l'aide du logiciel R	4
Bibliographie	7

Introduction : présentation du fonds et contexte historique

Le département des Manuscrits de la BnF conserve des pièces très précieuses pour les recherches généalogiques, notamment sur les familles françaises nobles. Une grande partie est réunie dans le *Cabinet des titres*, qui compte au total 6800 volumes.

Issu de l'ancien département des Titres et généalogies, le *Cabinet des titres* trouve son origine dans le legs du cabinet de François-Roger de Gaignières, en 1711, et de celui de Charles d'Hozier en 1717. Profitant de ces acquisitions volumineuses, l'abbé Bignon, Garde de la Bibliothèque royale, décide en 1720

de créer au sein de la bibliothèque un département spécialisé sur ces thèmes. Régulièrement enrichi au long du XVIII^e siècle, le département bénéficie de l'entrée exceptionnelle en 1785 de près de 600 quintaux de parchemins provenant de la Chambre des comptes de Paris. A la fin du XVIII^e s., il rassemble plus de 500 000 pièces de nature archivistique très variée, réunies dans des dossiers par nom de famille classés par ordre alphabétique. A la Révolution, le département des Titres et généalogies est rattaché au département des Manuscrits et prend le nom de *Cabinet des titres*. Sous l'impulsion de Léopold Delisle, alors administrateur de la bibliothèque, le fonds est classé entre 1876 et 1882 par Ulysse Robert. Celui-ci crée six séries homogènes et une septième plus disparate.

Parmi les six premières séries, la série des *Pièces Originales* se distingue par son volume : elle rassemble 3061 registres, soit près de la moitié du *Cabinet des titres*. Alors que les autres séries conservent essentiellement des travaux de généalogistes, celle des *Pièces Originales* contient des documents historiques qui peuvent renseigner sur l'histoire d'un individu (lettre d'anoblissement, lettre de charge, acte de mariage, etc.) Elle rassemble des documents établis entre le IX^e et le XVIII^es., sur papier et sur parchemin, parfois scellés et de formats divers. Ces documents sont classés par dossiers en fonction du nom de famille et ne sont pas inventoriés un à un. Lors du classement du fonds au XIX^e s., les documents ont été montés dans des registres pour former un ensemble apparemment homogène. Chaque feuillet composant les documents a été numéroté : ces éléments numérotés ont été nommés « pièces »¹. Chaque registre rassemble donc un nombre variable de dossiers, qui comprennent eux-mêmes un nombre variable de pièces.

1 : Un document peut donc être composé d'une pièce (document composé d'un seul feuillet recto-verso) ou de plusieurs pièces (documents composés de plus d'un feuillet recto-verso).

Contraintes imposées par le manque de visibilité sur un fonds volumineux

Aujourd'hui, la série des *Pièces Originales* est conservée sur le site de Bussy-Saint-Georges (fig.1). Elle est fréquemment demandée en consultation par les chercheurs et les amateurs de généalogie. Mais les registres, réalisés avec des matériaux de qualité médiocre, se sont largement dégradés : plus de 20% d'entre eux ne peuvent plus être communiqués. La numérisation pourrait rendre le fonds plus accessible, mais

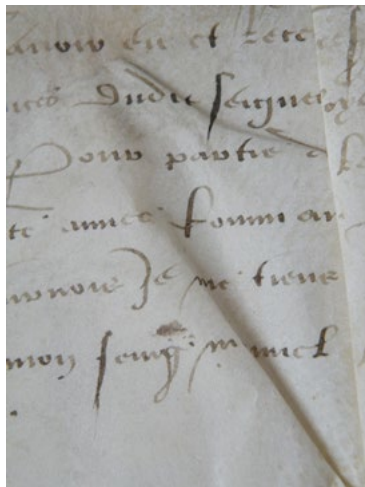


Figure 2. Détail d'une pièce dont les déformations empêchent la numérisation en l'état (© INP/ Pauline Elie)

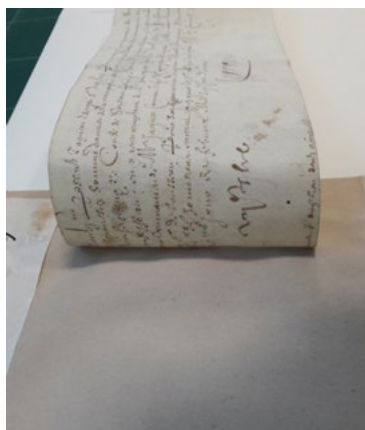


Figure 3. Vue d'une pièce dont le verso est inaccessible pour numérisation (© INP/Pauline Elie)

l'état des registres et de certaines pièces (fig.2) rend ce projet impossible à l'heure actuelle. L'objectif est aussi d'améliorer l'état de conservation du fonds pour assurer sa bonne conservation à long terme.

Le département des Manuscrits souhaite donc mettre en place un programme de conservation-restauration du fonds étalé sur plusieurs années, afin d'améliorer l'état de conservation de la série tout en permettant sa numérisation.

Cependant, le volume conséquent de la série et l'absence de visibilité sur la nature et l'état de conservation des pièces rend impossible la planification d'un tel projet. L'évaluation de l'état de la série est complexifiée par la différence d'échelle : faut-il regarder le registre ou le document ? De plus, si les reliures des registres sont homogènes, les documents contenus le sont moins : ils peuvent être composés d'un ou de plusieurs feuillets, sur papier ou sur parchemin, de formats variés. Les systèmes de montage dans les registres sont également différents d'une pièce à l'autre. Ils rendent parfois le verso d'une pièce inaccessible pour numérisation sans risquer d'aggraver l'état de conservation, comme lorsqu'une pièce est collée par l'un de ses bords sur le papier de montage (fig.3).

Dès lors, comment définir un protocole d'intervention pour ce type d'ensemble, afin d'améliorer l'état de conservation de la série tout en permettant sa numérisation ?

Cette question a fait l'objet de mon mémoire de fin d'études au département des restaurateurs de l'Institut national du patrimoine lors de l'année scolaire 2017-2018. Un registre représentatif, coté PO 12 (fig.4), a été étudié et traité et une étude globale de la série a été réalisée. Le changement d'échelle constant entre l'étude du cas particulier et l'étude de l'ensemble de la série a permis de définir une proposition d'une démarche de traitement adaptée.



Figure 1. Vue de la série des Pièces Originales (© INP/ Pauline Elie)



Figure 4. Vue du registre Mss Français 26496, PO 12 avant traitements (© INP/Pauline Elie)

1. Réalisation d'une évaluation de l'état de la série par sondage

Les études de collections se sont largement développées depuis les années 1970. L'émergence de l'informatique a contribué à faciliter leur exécution et à en faire un outil indispensable en conservation. Dans le cas de collections ou de fonds particulièrement volumineux, comme celui des *Pièces Originales*, il est impossible d'observer exhaustivement chaque pièce contenue dans les registres. Il faut alors recourir à des méthodes empruntées au domaine des statistiques,

comme le sondage. Le principe du sondage est d'observer un échantillon, c'est-à-dire une partie de l'ensemble (appelé « population »), afin de tirer des conclusions. Pour que l'étude soit fiable, elle doit répondre à certaines exigences méthodologiques. Celles-ci sont décrites dans plusieurs guides ou articles, et notamment dans la norme AFNOR NF Z40-011. La préparation de l'étude se fait en deux étapes parallèles, la préparation du questionnaire et la préparation du **plan de sondage**.

1.1 Préparation du questionnaire à partir de l'étude technique et du constat d'état du PO 12

Un outil de collecte est nécessaire pour enregistrer les informations relevées au cours de l'étude. Bien que des fiches papier puissent être utilisées, l'outil numérique est plus adapté car il permet d'exploiter plus facilement les données et de multiplier les copies de sauvegarde. Plusieurs logiciels permettent de réaliser des bases de données, dont certains dédiés comme Filemaker Pro® ou Access®. Dans le cadre des *Pièces Originales*, la création d'une base de données relationnelle ne se justifiait pas car l'objectif était d'avoir une vision du fonds dans son état actuel et non de son évolution. Le logiciel Excel®, qui présentait aussi des atouts pour l'exploitation des données, a donc été privilégié.

Les différents champs de la base ont été définis à partir de l'étude technique et du constat d'état réalisés sur le registre PO 12. La base s'articulait en trois parties : la première était dévolue à la description et à l'état du registre observé, la deuxième à la description matérielle de la pièce (**fig.5**) et la troisième à l'état de conservation de la pièce. L'objectif était de collecter des informations suffisamment diverses pour avoir une étude précise de la série sur le plan de sa composition et de son état.

N° de dossier	N° de pièce	Support	Nombre de sceaux	Format de la pièce	Pièce recto-verso ²	Montage 1	Montage 2
31856	7	Papier vergé	0	1- Petit format	oui	4- Montage sur onglet	0- Pas de double montage
31856	8	Papier vergé	0	1- Petit format	oui	4- Montage sur onglet	0- Pas de double montage
31856	9	Parchemin	0	1- Petit format	oui	4- Montage sur onglet	0- Pas de double montage
31856	10	Papier vergé	0	1- Petit format	non	1- Collage en plein	0- Pas de double montage
31856	11	Papier vergé	0	1- Petit format	non	1- Collage en plein	0- Pas de double montage
31856	12	Papier vergé	0	1- Petit format	non	1- Collage en plein	0- Pas de double montage

Figure 5. Extrait de la base de données Excel® : partie consacrée à la description des pièces (© INP/Pauline Elie)

La quasi-totalité des champs ont été conçus pour être renseignés de façon contrainte, via une liste déroulante ou sous forme numérique, afin d'harmoniser les réponses et ainsi de faciliter l'exploitation des résultats.

1.2 Préparation du plan de sondage

En parallèle, un plan de sondage a été préparé. Celui-ci devait permettre de définir précisément les variables, c'est-à-dire ce qui allait être observé. Le choix s'est porté sur la pièce, c'est-à-dire chaque feuillet recto-verso numéroté², car cela était plus pertinent pour l'objectif du traitement de numérisation. Cependant, pour disposer de données statistiques sur le contenu des registres, toutes les pièces contenues dans chaque registre sélectionné ont été évaluées.

Afin de choisir les registres à observer, le **tirage aléatoire** a été privilégié au tirage systématique qui est la méthode proposée dans la norme. Le tirage aléatoire est la méthode la plus fiable en statistique, mais elle n'est que rarement appliquée en conservation, le tirage systématique lui étant souvent préféré pour des questions logistiques. Le tirage aléatoire repose sur un tirage au sort : il faut donc posséder la liste complète des unités pour ensuite en sélectionner « au hasard ». Or, dans le cadre des études de collections, il est rare d'avoir un catalogue à jour, sans pièce manquante, qui puisse être facilement interrogé par un outil informatique. Dans le cadre des *Pièces Originales*, chaque registre porte un nombre compris entre 1 et 3061. La liste des registres à observer a donc été générée informatiquement, en tirant aléatoirement une liste de 200 nombres compris entre 1 et 3061. Cette suite irrégulière de nombres donnait la liste des registres à prélever en respectant l'ordre de tirage. L'objectif était d'observer un maximum de registres de la liste lors du temps imparti.

Ce plan de sondage a été élaboré lors d'une visite-pilote réalisée sur le site de Bussy-Saint-Georges. Cette visite préalable à l'étape de collecte des données a également permis de tester la base et d'apporter les corrections finales.

2 : Un document peut être composé d'une ou de plusieurs pièces

1.3 Collecte des données

La collecte des données a ensuite été réalisée en trois sessions, sur dix journées de sept heures, par Isabelle Bonnard et Pauline Elie. A tour de rôle, l'une était chargée de la saisie informatique tandis que l'autre réalisait le constat des pièces et dictait ses observations.

Au terme de la collecte des données, 11 719 pièces ont été évaluées, soit 1 114 dossiers rassemblés dans 53 registres. Bien que cet échantillon ne représente qu'1,73% de la série, le nombre d'unités observées est suffisamment important au regard de l'ensemble de la série pour que l'étude soit fiable. En effet, dans le cas d'un sondage, plus la population est importante, moins le pourcentage de la série observé peut être important.

2. Exploitation des données collectées à l'aide du logiciel R

Une fois les données collectées, il faut parvenir à en tirer des conclusions. Pour mener l'analyse des données, il est préférable de savoir précisément quelles informations l'on recherche, afin de questionner la base en ce sens.

2.1 Choix du logiciel R pour l'exploitation des résultats

Bien que la base de données ait été réalisée sur Excel®, un autre logiciel a été utilisé pour réaliser l'exploitation des résultats, le logiciel R.

R est à la fois un **langage de programmation** et un **logiciel libre** et gratuit, qui peut être installé sur n'importe quel ordinateur. Ce logiciel est couramment utilisé par les statisticiens, mais encore peu employé dans le domaine patrimonial. L'un des principaux avantages de R est de ne pas manipuler directement le fichier dans lequel sont stockées les données, contrairement à Excel®. Il n'y a donc quasiment aucun risque de modifier accidentellement les données lors de l'exploitation des résultats. A l'ouverture, une console s'affiche. Des fonctions doivent alors être utilisées pour importer les données et les manipuler. En parallèle, un script est rédigé. Récapitulatif des fonctions utilisées, le script permet d'exécuter à nouveau l'analyse et de garder une traçabilité des opérations effectuées.

Certaines fonctions sont très pratiques et permettent d'aller plus vite que sur Excel : la fonction « summary » par exemple, permet d'obtenir en quelques secondes un résumé de l'ensemble des données.

2.2 Principaux axes d'analyse des données et résultats

Pour l'étude des *Pièces Originales*, trois axes d'analyse principaux ont été définis en vue de la définition et de la planification des traitements en termes de temps et de compétences à mobiliser :

- Déterminer si les registres doivent être déreliés. Une analyse des données registre par registre a montré qu'un pourcentage significatif de pièces dans chaque registre n'est pas correctement numérisable uniquement à cause de la structure reliée. L'analyse a également montré que 71,5% des pièces peuvent être numérisées en l'état sans retirer la reliure. Ce nombre passe à 86,0% en retirant la reliure. La reliure doit donc être retirée car elle empêche systématiquement la numérisation d'une partie de chaque registre. Le retrait de la reliure donne également accès à davantage de pièces qui peuvent être numérisées sans traitement.

- Répartir les pièces en différentes catégories en fonction des traitements nécessaires pour la numérisation, afin de déterminer par la suite des temps de traitement pour la série. Quatre typologies de traitement ont été définies : décollage, remise à plat, consolidation du support et consolidation des sceaux et cachets. Parmi les 14,0% de pièces qui nécessitent des traitements, la plupart doivent être décollées de leur support (7,0%) ou consolidées (4,9%). Seules 1,7% des pièces requièrent une remise à plat avant numérisation. La quasi-totalité des pièces à traiter n'a besoin que d'un seul type de traitement.

- Obtenir des connaissances supplémentaires sur la série, notamment du point de vue de la répartition des supports et des sceaux. Le type de support a un impact important sur le temps de traitement : parmi les pièces observées, 56,0% ont été réalisées sur papier vergé et 41,0% sur parchemin (fig.6). Les 3,0% restants sont des pièces sur papier vélin ajoutées lors de la constitution

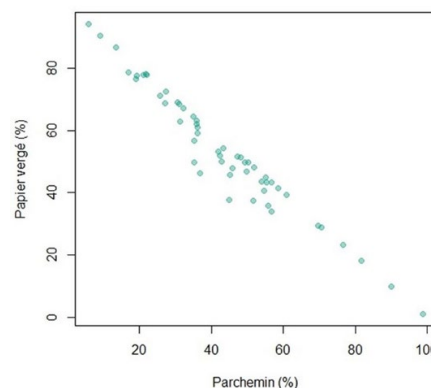


Figure 6. Graphique présentant la fraction parchemin/papier vergé par registre (un point = un registre) (© INP/Pauline Elie)

des registres. Concernant les sceaux, seules 5,0% des pièces en comportent, mais plus de la moitié d'entre eux sont en mauvais état (45,8%).

D'autres informations ont pu être extraites à partir des données collectées, comme le format, l'empoussièrement ou l'état des reliures des registres.

2.3 Vérification de la possibilité d'extrapoler à l'ensemble de la série

Pour extrapoler les observations réalisées sur l'échantillon à l'ensemble des pièces, il faut vérifier grâce à des tests statistiques que les données suivent une loi de probabilité. Une loi de probabilité est un modèle qui permet de décrire une expérience aléatoire. En vérifiant que les données de l'échantillon suivent un certain modèle, on peut extrapoler les résultats au reste de la population en suivant ce modèle. Le modèle de la loi normale (aussi appelée loi de Gauss)³ est le modèle le plus courant et le plus pratique d'utilisation. Il a donc été testé en premier à l'aide de R et s'est avéré satisfaisant. Il a ensuite été utilisé pour réaliser l'extrapolation des résultats observés sur l'échantillon à l'ensemble de la population.

Cette extrapolation, couplée au calcul de l'intervalle de confiance et de la marge d'erreur, a permis de déterminer le nombre de pièces probable par registre. En moyenne, il y a 221 pièces par registre. Avec un intervalle de confiance de 95% (soit une marge d'erreur de 5%), un registre pris au hasard a 95% de chances d'avoir un nombre de pièces compris entre 117 et 325.

2.4 Définition d'un protocole de traitement et application au PO 12

A partir des informations obtenues lors de l'étude historique, du constat d'état du PO 12 mais aussi de l'étude statistique du fonds, un protocole de traitement a été défini.

2.5 Elaboration d'un protocole en trois phases

En raison du très grand nombre des pièces qui constituent la série des *Pièces Originales*, il n'est pas envisageable de traiter l'ensemble dans un objectif optimal de conservation. En effet, la plupart des pièces étant écrites à l'encre ferrogallique, tout traitement aqueux doit être suivi d'un traitement complémentaire afin d'éviter que les encres ne se dégradent. Une solution optimisée a donc été proposée, avec un protocole de traitement en trois phases :

Lors de la phase 1, la reliure est retirée (**fig.7A**) et les pièces dont le verso est inaccessible car collé au support de montage sont démontées, sans retirer les résidus de papier encore maintenus par la colle. Les traitements (**fig.7B**) consistent en un dépoussiérage rapide des pièces, associé à une consolidation des altérations mécaniques qui peuvent conduire à une perte d'informations. Au cours de cette phase, ne sont mis à plat que les documents dont les déformations empêchent la numérisation en l'état, par des méthodes compatibles avec la présence d'encres ferrogalliques. Les pièces sont ensuite reconditionnées dans un classeur qui permet d'y accéder facilement pour numérisation ou pour traitement complémentaire (**fig.7C**). Cette étape permet de préparer les pièces pour la numérisation et de stabiliser les altérations qui menacent à court terme la bonne conservation des documents. Elle est la seule nécessaire pour rendre les documents accessibles au plus vite via la numérisation.



Figure 7. A) Retrait de l'apprêt et séparation des cahiers B) Traitement en série des pièces C) Vue du remontage et du reconditionnement des pièces en classeur (© INP/Pauline Elie)

3 : La loi normale est la loi statistique la plus connue, dont le graphique ressemble à une cloche.

La phase 2 permet de stabiliser dans un second temps l'ensemble des altérations non urgentes mais qui peuvent être traitées rapidement. Elle consiste en un dépoussiérage approfondi des documents. Les consolidations non urgentes qui ne risquent pas de conduire à une perte d'information sont traitées à cette occasion.

Enfin, la phase 3 permet de traiter l'ensemble des altérations dans un objectif optimal de conservation. Les documents sont tous entièrement décollés du papier de montage du XIX^e s. et remis à plat.

Les traitements de la phase 1 et de la phase 2 ont été appliqués au registre PO 12.

2.6 Estimation du temps de traitement

Les temps de traitement du PO 12 ont été mesurés pour chaque opération. Au total 178 heures ont été nécessaires pour le traitement du PO 12, dont 106 heures pour la phase 2, non obligatoire pour l'objectif de numérisation.

Les durées de chaque typologie de traitement ont été calculées à partir des temps mesurés lors du traitement du PO 12 et rapportées à l'échelle de l'échantillon, lui-même extrapolé à l'ensemble de la série. La simple multiplication du temps de traitement du PO 12 par le nombre de registres restant à traiter aurait donné une estimation faussée, car se baser sur un seul registre ne donne pas des données représentatives.

L'application de la phase 1 à l'ensemble des registres requiert un temps de traitement total de 12 ans à raison d'une équipe de 5 personnes à temps plein. La plus grande partie du temps de traitement concerne le retrait de la reliure (environ 24h pour le PO 12) et le retrait des pièces du support de montage (environ 30 h pour le PO 12). Il est toutefois possible d'améliorer encore ce temps de traitement. Travailler sur plusieurs registres en parallèle permet par exemple d'optimiser les temps d'attente.

Conclusion

Le projet des *Pièces Originales* a requis une approche d'ensemble, nécessaire dans le cadre d'un traitement de masse. L'étude par sondage du fonds est une étape indispensable pour mieux le comprendre dans sa globalité et ainsi construire un projet de conservation-restauration approprié. Les informations obtenues constituent un véritable outil d'aide à la décision pour le département des Manuscrits, qui a une meilleure visibilité sur l'état de la série et sur les temps de traitement nécessaires.

Bibliographie

ELIE, P. (2018). « 12/3061 : Etude et conservation-restauration des 207 pièces d'archives rassemblées dans le registre BnF, MSS, Pièces Originales (PO) 12 (XIV^e-XVIII^e siècles ; Paris, Bibliothèque nationale de France). Proposition d'une démarche de préparation à la numérisation réalisable sur les 3061 registres PO. Evaluation statistique de l'état du fonds : mise en place d'une méthode par sondage », Aubervilliers, INP, département des restaurateurs du patrimoine, 293 p.

Association Française de Normalisation. (2005). « Norme AFNOR NF Z40-011 : Méthode d'évaluation de l'état physique des fonds d'archives et de bibliothèques », Saint-Denis-la-Plaine, 33 p.

Glossaire

Langage de programmation : Langage destiné à formuler des algorithmes et à produire des programmes informatiques qui appliquent lesdits algorithmes.

Logiciel libre : Logiciel qui peut être utilisé, modifié et redistribué par la personne qui l'utilise.

Plan de sondage : Document qui décrit la méthodologie employée lors de la réalisation du sondage.

Tirage systématique : Prélèvement qui consiste à évaluer la population totale et à la diviser par le nombre d'unités qui doivent être observées de façon à obtenir un « pas » régulier.

Remerciements

Les auteurs remercient l'ensemble des personnes ayant permis de mener ce projet à bien : le personnel de l'INP, en particulier Maroussia Duranton et Jane Echinard, ainsi que l'équipe de Nathalie Auffredou, BnF, département de la Conservation (Bussy-Saint-Georges).

Résumé :

En 2017-2018, une étude préalable à la numérisation des 3061 registres de la série des Pièces Originales (département des Manuscrits, BnF) a été menée. Elle était l'objet de mon mémoire de fin d'études à l'Institut national du patrimoine, dirigé par Isabelle Bonnard et co-encadré par Mathieu Lescuyer, Caroline Laffont et Serge Cohen. Le travail a notamment porté sur le développement d'une démarche de préparation à la numérisation, à partir d'une évaluation de l'ensemble du fonds et du traitement d'un registre représentatif, coté PO 12.

Abstract :

In 2017-2018, a study was carried out prior to the digitisation of the 3061 registers of the "Pièces originales", (Manuscripts Department, BnF). It was the subject of my thesis at the Institut national du patrimoine, directed by Isabelle Bonnard and co-supervised by Mathieu Lescuyer, Caroline Laffont and Serge Cohen. The work focused in particular on the development of an approach to prepare for digitization, based on an evaluation of the entire collection and the processing of a representative register (ref. PO 12).