

Jean-Paul Oddos

Bibliothèque nationale de France

LA NUMÉRISATION DES ARCHIVES GÉNÉRALES DES INDES À SÉVILLE

LE VISITEUR qui, par un soir d'hiver (mais l'hiver existe-t-il à Séville ?), vient flâner entre la cathédrale (volée de cloches du haut de la Giralda, le minaret de briques roses, diseuses de bonne aventure) et les murs crénelés qui cernent les jardins royaux de l'Alcazar (quels murs pourraient enfermer les odeurs d'oranger ?) peut découvrir un bâtiment de « *fort belle apparence* », et qui appelle irrésistiblement en lui les mots de « palais colonial » : ce quadrilatère bas (toit en terrasse que signale une balustrade de pierre, ombre des palmes qui bougent sur les façades), posé sur un socle de marches lissées, protégé de grosses chaînes, c'est l'ancienne Casa Lonja, un ancêtre de l'actuelle Chambre de commerce. Là, au cœur de Séville, sur 5 000 m², 9 km de rayonnages, en 43 000 liasses bien denses, reposent les fonds de deux institutions disparues, la Casa de Contratacion et le Conseil des Indes, c'est-à-dire la

mémoire de la découverte et de la colonisation du Nouveau Monde. L'Archivo general de Indias (AGI) – c'est le nom de l'institution qui gère ces fonds rassemblés ici depuis la fin du XVIII^e siècle – porte aussi en elle un rêve nouveau, depuis quelques

années. Et si le visiteur gravit les marches, s'aventure au long des vastes galeries, c'est aujourd'hui pour découvrir un projet exceptionnel : un programme de protection et de mise en valeur des fonds basé sur la technique de transfert numérique. Dans une salle de lecture, ce visiteur va découvrir, côte à côte, un chercheur feuilletant des liasses manuscrites et un autre, installé devant deux écrans (l'un pour les informations textuelles, l'autre pour les images), manipulant

une souris : ce contraste-là résume l'expérience menée à Séville, où, pour la première fois sans doute, la numérisation a été appliquée de façon systématique à des collections anciennes. Il y a là quelque chose d'exemplaire qui méritait bien une escale et quelques notes de voyage¹. Ces notes sont brèves, incomplètes sans doute, et trahissent peut-être l'admiration ressentie : au lecteur donc de se montrer plus sévère dans son jugement, ou de prendre à son tour la route de Séville.

Patrimoine de l'humanité

Les Archives générales des Indes conservent un fonds pratiquement clos, rassemblé à la fin du XVIII^e

¹ Ces notes sont le compte rendu d'une mission à Séville, les 9 et 10 décembre 1993, pour l'Etablissement public de la Bibliothèque de France

siècle à partir de séries de diverses provenances. Cependant, ce fonds se caractérise par son homogénéité et sa forte structuration.

Homogénéité : il s'agit d'un fonds historiquement ciblé (histoire de la découverte et des relations ibéro-américaines du XV^e au XIX^e siècle), constitué de documents d'un format quasi identique (feuillet de format A3 environ), sur un papier de bonne qualité, dont l'état de conservation est pour l'essentiel très satisfaisant². Seuls 7 à 8 000 cartes et plans forment un ensemble à part : formats plus grands (autrefois pliés en 2 ou 4), et souvent rehaussés de couleur.

Structuration : ces archives ont une organisation hiérarchique très simple. L'ensemble est constitué de quelques grandes provenances (archives de la Casa de Contratacion par exemple), chaque fonds est constitué de séries, chaque série de liasses, chaque liasse a environ 2 000 pages³.

Pour les besoins de repérage (préparation, contrôle qualité...), chaque liasse est divisée en « blocs » correspondant à une unité logique (*expediente* ou document), puis en feuillet.

A chaque niveau hiérarchique correspond un niveau de description historique (du fonds, de la série, du contenu de la liasse) plus ou moins développée. Mais chaque feuillet est parfaitement situé dans un contexte hiérarchique.

Ce fonds a une importance historique exceptionnelle : il a été classé par l'Unesco « Patrimoine de l'humanité ». Il est fortement sollicité par les chercheurs, ce qui présente des risques de dégradation, malgré sa qualité originelle. Le cinquième centenaire de la Découverte du Nouveau Monde a, de plus, renouvelé l'intérêt des historiens pour ces documents.

D'autre part, son niveau de description était hétérogène et les outils d'ac-

cès très divers (inventaires imprimés, inventaires manuscrits, fichiers, etc.), ce qui induisait une importante manipulation de documents pour une accessibilité relativement faible.

Ce sont ces deux données qui justifient les objectifs du projet : assurer la protection de documents uniques, très précieux, permettre leur meilleure accessibilité et une meilleure exploitation des informations, par la reproduction massive de documents originaux et leur constitution en bibliothèque électronique.

Un système intégré de gestion

Le projet est cependant né d'une volonté extérieure aux Archives : celle de deux mécènes (la fondation Ramon Aceres⁴ (FRA) et IBM-Espagne) de rechercher une nouvelle application aux technologies de l'image. Les circonstances historiques (l'approche de 1992, l'Exposition universelle), le partenariat effec-

(notamment les moyens mis en œuvre, pour un budget de 1 milliard de pesetas), il n'est pas seulement une expérience. Il est conçu pour être développé sur place puis élargi à l'ensemble des archives espagnoles, et ses éléments de base pourront être réutilisés pour d'autres programmes internationaux (comme par exemple pour le projet des archives du Komintern de Moscou, sous l'égide du Conseil international des archives et du Conseil de l'Europe).

Les responsables du projet, dans leurs présentations écrites ou orales, insistent beaucoup sur l'aspect intégré du système mis en place. Celui-ci comporte en effet trois volets :

- la gestion d'une base de données catalographiques, celle-ci étant issue, d'une part, d'une conversion des données héritées et, d'autre part, d'une immense remise à niveau des outils d'accès ;
- la gestion des principaux services au public (accréditations, communications, réservations, statistiques, reproductions...);

Le programme est conçu pour être développé sur place puis élargi à l'ensemble des archives espagnoles

tif d'IBM-Espagne, du ministère de la Culture, et de la fondation Ramon Aceres ont permis la concrétisation du projet. Celui-ci est devenu une réussite parce que les Archives présentaient, comme on l'a dit, des conditions particulièrement favorables (homogénéité, forte structuration du fonds) et aussi parce que ses promoteurs ont su replacer l'aspect particulier - la reproduction numérique des documents - au sein d'un programme plus vaste : la gestion informatisée de toutes les fonctions archivistiques.

Ainsi, si ce programme présente beaucoup d'aspects exceptionnels

- la gestion d'une bibliothèque d'images sur disques optiques, aujourd'hui riche de 9 millions d'images sur 9 000 disques de type Panasonic.

Chacun de ces modules a connu un niveau de réalisation minimum en fin de phase 1⁵, permettant le fonctionnement intégré (accréditation, gestion d'une place, consultation du catalogue, commande d'une série d'images numérisées ou de liasses-papier, consultation, demande de reproduction), même si tous les développements possibles n'ont pas été réalisés.

2. Quelques liasses ont été mouillées lors d'un incendie en 1924 et sont en cours de restauration. Certaines archives concernant les Philippines sont sur papier de riz assez friable. Le problème le plus fréquent est l'usure des bords de feuillet, ou feuillets « roulés » (conditionnement en liasse), qui peuvent faire perdre des parties de textes à la scannérisation.

3. La moyenne de pages à reproduire par liasse s'établit à 1 850.

4. Ramon ACERES était, avant sa mort en 1990, le fondateur et le propriétaire d'une des plus grandes chaînes de diffusion européenne (El Corte Inglés).

5. Un premier accord, proposé en 1985, a été signé en 1986 par les trois partenaires IBM Espagne, FRA et ministère de la Culture. Un deuxième accord, plus réduit, a été signé par les mêmes en 1993 et 1994 pour permettre un certain nombre de développements, avant de passer à une situation de « fonctionnement ordinaire » (avec d'autres financements).

Par exemple, la recherche des disques optiques dans le stock est, elle encore, manuelle, alors que la robotisation de cette fonction aurait pu être visée : le délai de service manuel (moins d'une minute) a paru, pour l'instant, acceptable, en comparaison du coût de la robotisation.

Choix, coût et accueil de la numérisation

Tout le projet n'a donc pas été orienté vers la simple production d'images numérisées ; son axe est plutôt l'amélioration d'ensemble de la gestion et de l'usage du fonds par la mise à disposition d'une partie des documents sous forme numérique. C'est donc en terme d'impact général que l'on peut mesurer son intérêt.

Très rapidement, l'idée de reproduire tout le fonds, qui avait été à l'origine même du projet, a été remplacée par la perspective d'un système global d'exploitation, et par la recherche d'un seuil minimum de documents à reproduire, pour permettre une amélioration sensible de la gestion d'ensemble.

L'étude des documents les plus consultés (ceci allant jusqu'à 50 communications par an pour une liasse) a montré que 10 % des documents supportaient 25 à 40 % des demandes. C'est ce seuil de 10 %, soit 8 millions d'images (sur les 80 millions de pages que compte le fonds), qui a paru réalisable pour l'année 1992⁶.

Cet ensemble d'images numériques est désormais un élément décisif dans la consultation du fonds. On observe par exemple que les demandes de reproduction se répartissent, dès 1992 – alors qu'il n'y a pas de numérisation à la demande –, de la façon suivante : 12 000 pages reproduites sur microfilms (demandes extérieures), 13 000 pages par photocopies (demandes sur

place), 9 000 pages par tirage papier à partir d'images numériques.

Même si la fréquence de consultation d'une liasse n'a pas été le seul critère de sélection pour son transfert numérique, les 10 % de départ ont été assez bien ciblés, puisqu'ils représentent de façon continue plus de 25 % des demandes de reproduction.

Sans vouloir aborder ici une analyse détaillée des coûts, on peut indiquer que le budget d'ensemble s'est réparti sur plusieurs postes :

- la conception et la réalisation du système de gestion informatisée (base de données, gestion des services, gestion des images) par une équipe ayant compté jusqu'à 20 informaticiens pendant plusieurs années ;

Le lecteur peut renverser l'image, augmenter le contraste, effacer les fonds, ainsi que les taches et les transparences parasites, agrandir tout ou partie de l'image...

- la préparation et la saisie des données textuelles (catalogues, inventaires...) par une équipe de 7 personnes à Madrid, 5 à Séville, pendant 5 ans ;

- la préparation intellectuelle et matérielle des documents à reproduire : 10 personnes à Séville (5 ans), 10 à Madrid (2 ans) ;

- la scannérisation et le contrôle, à Séville : 30 personnes pendant 5 ans (depuis 1992, l'équipe est réduite à 8 personnes) ;

- enfin, les matériels informatiques (serveurs, écrans haute définition, terminaux⁷) et disques optiques.

Une première estimation donne un coût de 30 pesetas la page numérisée (environ 1,5 F).

7. Notamment un IBM AS-400 comme serveur pour la base de données et le système de gestion des usagers, des IBM PS-2 comme serveurs des disques optiques, ainsi que pour les postes de consultation ; ces derniers comportent des moniteurs IBM 8513 pour le texte, 8508 pour les images en noir et blanc, 6091 pour la couleur. Les stations de numérisation sont équipées de micros IBM AT, de scanners Rank Xerox avec systèmes RF 5010C, et, pour les cartes et plans, le micro est un PS-80, et le scanner un Nikon LS 3500.

Le coût de fonctionnement n'est pas encore complètement analysé, et il n'est pas possible, pour l'instant, de le reporter au coût du projet. Les recettes provenant des reproductions ne peuvent être mises en regard. Elles ne sont de toute façon pas affectables et les tarifs très bas (12 pesetas la page, quel que soit le mode de reproduction) sont fixés pour l'ensemble des archives.

Les responsables du projet reconnaissent que les chercheurs, les universitaires ont été peu impliqués dans la réflexion⁸ ; au départ (en 1985-86), un certain scepticisme entourait l'idée que les petites archives sévillanes puissent faire l'objet d'une telle innovation. Aujourd'hui, la plupart des chercheurs – il y a parfois quelques

personnes âgées réticentes – acceptent volontiers le système, alors qu'il n'y a jamais eu de transition par la lecture de microformes. D'après les intéressés, cet accueil favorable tient au fait que le chercheur dispose de plus d'informations, que l'accès à ces informations et aux documents eux-mêmes est plus rapide, que les consultations de documents électroniques ne sont pas limitées (3 liasses par jour pour la consultation des originaux), et que les reproductions sur papier sont très rapides et peu coûteuses.

Le mode même de consultation – les outils de travail – n'est pas cité parmi les avantages offerts. Pourtant, même si toutes les possibilités n'ont pas été explorées à cette étape, le lecteur peut renverser l'image, augmenter le contraste, effacer les fonds, ainsi que les taches et les transparences para-

6. Les critères de choix suivants ont été croisés : documents les plus consultés en 1984 et 1985, séries complètes de documents les mieux décrits et les moins complexes, couverture géographique intéressante (tous les pays américains), reconstitution de fonds dispersés (autres fonds à Séville, à Madrid, à Simancas). Avec 10 % d'images, on atteint non 40 % mais 25 % des demandes.

8. Pedro GONZALEZ, maintenant directeur des AGI, fut, en tant que directeur du Centre de documentation des Archives, (Centro de Informacion documental de Archivos) au ministère de la Culture, à Madrid, un des principaux responsables de ce projet : il est historien et archiviste de formation.

sites, agrandir tout ou partie de l'image... Plus important encore apparaît le « guidage » de la recherche, *i.e.* la possibilité de croiser en permanence les données textuelles d'information avec la lecture des images elles-mêmes. Un des manques ressentis par les lecteurs est celui d'un accès plus large par mots-clés, qui multiplierait les possibilités offertes par le système. La carence de vocabulaire contrôlé a conduit à la mise en chantier d'un thésaurus de descripteurs géographiques. Les suites du projet (informatisation des archives de l'université de Salammanque⁹, ou des archives du Komintern) devraient permettre une meilleure implication des chercheurs.

Les développements du projet

Le système mis en œuvre aux AGI est, on l'a vu, un système immédiatement fonctionnel, mais susceptible de développements. Un certain nombre d'entre eux est à l'étude dans le cadre de la phase 2 du projet. Cela concerne :

- le remplacement des disques Panasonic (eux-mêmes successeurs des disques IBM d'origine) par des CD-Rom, dont le standard s'impose, mais qui ne sont pas inscriptibles ;
- l'usage d'un seul écran de consultation pour la base de données et la base d'images (au lieu des deux écrans actuels), notamment pour gagner de la place en salle de lecture ;
- l'introduction d'un serveur robotisé pour les disques optiques (ceux-ci, on l'a vu, sont servis manuellement, avec un délai moyen d'une minute) ;
- la recherche en texte libre sur le catalogue (d'où la nécessité d'un vocabulaire contrôlé) ;
- l'exploration des possibilités de reconnaissance optique des écritures manuscrites ;
- et, de façon très immédiate, la communication à distance : un accord a été passé avec la société Telephonica pour une expérience pilote à partir du 1^{er} jan-

9. Une convention lie, comme à Séville, la FRA, le ministère de la Culture et les Archives de l'université.

vier 1994. La solution envisagée est le transport d'images en 100 dpi (*dots per inch*), avec seulement 2 niveaux de gris, afin de réduire le coût de la transmission – l'image obtenue étant jugée de qualité suffisante pour la lecture, et les reproductions traitées directement à la source. L'accord prévoit un bilan dès l'été 1994, pour poursuivre ou abandonner l'expérience.

En revanche, d'autres développements ne sont pas envisagés de façon immédiate : introduction d'outils plus performants sur le poste de lecture comme traitement de texte, feuille de calcul, rapprochement d'images, annotations¹⁰... ; gestion plus développée du magasinage (le système gère la communication, pas les adresses), ou de la conservation (pour chaque référence, une zone permet la description de l'état physique, mais les données ne peuvent pas être traitées actuellement).

L'utilisation d'images en couleur peut aussi être considérée comme une expérience prototype : elle concernerait d'abord les 8 000 cartes et plans en couleur. Le format des documents imposera ici le passage par le microfilm couleur, alors que les autres sont numérisés à partir de l'original¹¹.

Le projet d'informatisation des AGI ne s'est accompagné pour le moment que d'une ébauche de politique éditoriale : publication d'un CD-Rom *Tesoros del Archivo general de Indias* présentant le projet et une

10. En plus des possibilités déjà citées, il est possible de marquer les images (pour les rappeler, demander une reproduction) ; le système permet aussi de feuilleter les premières pages de chaque référence (jusqu'au « bloc »).

11. À la question : pourquoi ne pas toujours passer par un microfilm de sécurité (et déjà par ceux qui existaient) ?, Pedro Gonzalez répond que l'objectif de sauvegarde passe d'abord par une réduction des consultations, les originaux étant de grande qualité ; que le lien entre données textuelles et images étant primordial, il est difficile et coûteux de créer ce lien en temps différé ; que les microfilms produits à partir d'images numériques pour la communication à distance sont de bonne qualité ; qu'il sera sans doute possible prochainement de produire simultanément une image photo et une image numérique, lorsque cela sera nécessaire ; qu'enfin la production et la sauvegarde des images numérisées relèvent d'un processus continu, qui dépasse la seule question de la durée du support numérique.

sélection de documents exceptionnels, pour le centenaire ; la publication d'un volume supplémentaire, sur papier, de l'*Inventaire des « passagers »*¹², à partir de la base de données informatisées, est en projet.

Les AGI visent désormais la publication, sur CD-Rom, de certaines séries, accompagnées des données textuelles nécessaires à leur exploitation. Le premier projet est la publication exhaustive de la série citée des « passagers », en reprenant la partie déjà publiée sous forme papier. L'ensemble de cet inventaire rassemblera les 14 volumes prévus (dont les 7 déjà publiés sous forme papier). Ce projet résulte également de la coopération FRA, IBM, ministère de la Culture. Les AGI souhaitent ensuite s'orienter vers la publication, toujours sur CD-Rom, d'inventaires associés à des images (textes originaux, etc.). Pour cela la recherche de nouveaux partenaires sera nécessaire ; mais pour avancer vers une association Université / Archives / éditeur commercial, il faudrait, estime leur directeur, que les AGI disposent d'une plus grande autonomie de gestion.

Un bilan

L'expérience des AGI est arrivée à un stade suffisant de développement pour qu'un bilan puisse en être tiré. Le groupe d'experts de la commission Preservation and Access¹³, qui avait publié un premier rapport en 1989, devait faire le point avec les responsables du projet à Séville en janvier 1994. D'autre part, Pedro Gonzalez, en tant que membre du Comité de technologie des images pour les archives, ou lors du prochain congrès de l'American Society of Archivists, devrait publier plusieurs communications sur ce thème.

12. Il s'agit d'une des séries les plus consultées des AGI ; la liste de tous les passagers partant d'Espagne pour le Nouveau Monde, sur trois siècles.

13. Association américaine dirigée par Hans RUTIMAN, qui s'est fixé comme objectif d'assurer le mécénat des actions de conservation, notamment de microfilmage, dans le monde entier.

Il souhaiterait concentrer sa réflexion sur deux points : d'une part l'évaluation économique d'ensemble (ce qui demandera de disposer de tous les éléments chiffrés), d'autre part l'étude des « stratégies » de recherche mises en œuvre par les lecteurs à qui l'on offre des possibilités nouvelles d'accès, de travail, de vérification, etc. Sur ce sujet, les observations, les entretiens, montrent que ces « stratégies » ne sont pas encore systématiques. En revanche, la formation des usagers est très rapide : 2 à 3 jours au plus pour maîtriser toutes les possibilités du système.

L'expérience a apporté encore bien des leçons, ainsi :

– la qualité de la production est meilleure que prévue ; le pourcentage des images mauvaises, donc à refaire, est très faible. Les erreurs relevées au

par an, soit un peu plus de 400 images par jour – soit encore un temps moyen de 1 minute par image (sur 6 h 30 de travail effectif journalier). L'aspect pénible du travail tient à cette cadence lente, qui ajoute à la monotonie. Des scanners plus rapides (un matériel récent travaille à une vitesse cinq à dix fois supérieure) remplaceront un jour les appareils actuels ; mais les AGI souhaitent disposer de personnels permanents (et non de vacataires comme aujourd'hui), afin de les intégrer aux effectifs et d'organiser des rotations¹⁴.

– L'ensemble du personnel des Archives a vécu l'expérience de façon très positive.

– L'impact sur les mouvements des documents originaux est déjà significatif : la communication des documents numérisés atteint le tiers des

– parce qu'on a su replacer un projet « nouvelle technologie » dans le cadre d'une *approche d'ensemble* des besoins d'un établissement, notamment en matière de gestion et de services aux chercheurs ;

– parce qu'on n'a pas cherché à constituer, de toutes pièces, une « archive idéale » (et introuvable), mais qu'on est parti d'une *collection réelle*, dont l'homogénéité, la complétude¹⁵ et l'organisation interne ont été améliorées pour les chercheurs ;

– parce qu'on s'est appuyé sur tout le travail antérieur de documentation des fonds en l'unifiant en une seule base de données ;

– parce qu'on a fait preuve de pragmatisme en recherchant toujours ce qui était réalisable et en écartant les éléments aléatoires (sur le plan technique comme dans la sélection des documents) ;

– parce qu'enfin le projet a été découpé en étapes, et que la réussite de l'une a conditionné la mise en œuvre de la suivante.

Ce projet de numérisation des AGI, tel que nous le découvrons aujourd'hui, appelle une réflexion approfondie chez les professionnels des archives et des bibliothèques : nous pouvons en effet observer, en situation réelle, sur une collection ouverte à la recherche, l'impact des « nouvelles techniques ». Impact, bouleversement même, en terme de préservation, d'accessibilité, d'exploitation, d'enrichissement, de maîtrise générale des collections. A terme ce sont sans doute ces catégories bibliothéconomiques qu'il nous faudra réviser.

Avril 1994

On a su replacer un projet « nouvelle technologie » dans le cadre d'une approche d'ensemble des besoins d'un établissement

contrôle-qualité sont, par ordre décroissant : des images inversées (façon erronée de retourner le feuillet sur le scanner), que le lecteur peut d'ailleurs corriger de lui-même ; des manques d'une partie de l'image (bords roulés par exemple) ; des images répétées ou manquantes ; la non-correspondance entre référence et image.

Le poste de prise de vue avec écran de contrôle permet des vérifications simples (contrôle du nombre de pages de chaque « bloc » par exemple). Le contrôle-qualité est un second niveau, intervenant avec un délai assez important, parfois après la consultation où ces erreurs sont tout de suite repérées. L'insertion des images nouvelles ne soulève aucune difficulté.

– La cadence de production observée est de 100 000 pages par opérateur et

demandes et un lecteur sur trois utilise ce mode de recherche. Il est envisagé dans le court terme de stocker les originaux numérisés dans des rayonnages mobiles, pour dégager deux nouvelles salles de lecture au rez-de-chaussée, dans d'anciens magasins.

Une démarche exemplaire

Ces notes sommaires – qui doivent beaucoup aux entretiens que nous a accordés Pedro Gonzalez – ont essayé de montrer la valeur de cette expérience, au-delà des circonstances particulières. Car ces circonstances, pour importantes qu'elles soient, ne suffisent pas à expliquer cette réussite. C'est la démarche même qui est exemplaire :

14. Les AGI ont un effectif d'environ 50 personnes et 11 vacataires dédiés au projet. L'objectif de Pedro Gonzalez est de disposer de 5 postes permanents pour la vitesse de croisière.

15. La volonté de compléter les fonds des AGI par des séries encore conservées à Madrid, Simancas, ou la Chambre de Commerce de Séville a aussi été un critère déterminant pour la sélection initiale.