



Le projet NUMERIC : des données statistiques pour la numérisation du patrimoine culturel européen

Roswitha Poll
Munster, Allemagne

*Traducteur :
Sophie Palandjian, documentaliste,
Marseille, France*

Meeting: 92. Statistics and Evaluation, Information Technology and Preservation and Conservation

WORLD LIBRARY AND INFORMATION CONGRESS: 75TH IFLA GENERAL CONFERENCE AND COUNCIL
23-27 August 2009, Milan, Italy
<http://www.ifla.org/annual-conference/ifla75/index.htm>

Résumé :

Les archives, les bibliothèques et les musées s'affairent à transformer leurs immenses catalogues « analogiques » en format numérique. Leur principale motivation est de faciliter l'accès aux collections à un public potentiellement très divers.

La plupart des actions de numérisation se focalisent sur le patrimoine culturel mais les données statistiques sur la numérisation n'ont pratiquement jamais été collectées à une échelle nationale.

NUMERIC, un projet de la Commission européenne, a commencé à définir des critères et des méthodologies pour évaluer et décrire l'état d'avancement de la numérisation dans les institutions culturelles européennes. Le but de ce projet est double : mettre en évidence l'impact financier du passage au numérique et décrire les progrès réalisés dans la numérisation du patrimoine.

L'activité principale du projet consiste à élaborer un dispositif de collecte des données statistiques appropriées, et donner ainsi une idée d'ensemble de l'état d'avancement de la numérisation à une échelle nationale. Dans cette optique, différents aspects furent pris en considération lors de l'enquête :

- *La variété des supports numérisés (documents imprimés, matériel audiovisuel, manuscrits, objets dans les musées)*
- *Les formats et les normes de numérisation*

- *Le coût financier*
- *L'accessibilité aux données numériques*
- *Les usages et les usagers de ces services*
- *La différence entre le patrimoine culturel ayant déjà fait l'objet d'une numérisation et celui susceptible de l'être prochainement.*

L'accent a été mis non seulement sur les quantités d'objets ou de documents numérisés mais aussi sur l'utilité des projets mis en œuvre en termes d'éducation, de recherche et d'identité culturelle.

Après avoir testé le questionnaire auprès de plusieurs instituts d'archives, de bibliothèques et de musées, l'équipe du projet a travaillé en relation avec des experts désignés dans chaque pays pour étendre l'enquête dans toute l'Europe. Cet article donne une vue d'ensemble des résultats et des possibles évolutions de la structure du dispositif de collecte.

1. Introduction

Les archives, les bibliothèques et les musées sont les trois principaux organismes à recueillir et préserver l'héritage culturel d'un pays. Depuis déjà plusieurs dizaines d'années, elles se sont attachées à transformer leurs immenses catalogues « analogiques » en format numérique. Leur principal objectif est de faciliter l'accès aux collections à un public potentiellement très divers (chercheurs, enseignants ou public non spécialiste). L'autre objectif est de préserver le document original sans toutefois en restreindre l'accès.

La plupart des entreprises de numérisation se focalisent sur le patrimoine culturel d'un pays. Cela englobe les documents écrits, les images et les sons mais aussi les objets manufacturés et les spécimens naturels. Bien que les actions de numérisation soient nombreuses et variées, souvent financées par des programmes nationaux ou régionaux, il est pratiquement impossible d'obtenir des données fiables sur ce qui a déjà été réalisé. Les statistiques sur la numérisation sont la plupart du temps recensées à l'intérieur même d'une institution ou au sein d'un projet de financement mais n'ont pratiquement jamais été collectées à une plus grande échelle. Les données recensées et les méthodes de collecte diffèrent sensiblement selon les régions, les pays ou les institutions. Par conséquent, même si des informations existent, elles ne peuvent pas être compilées pour offrir une vue d'ensemble nationale et la comparaison entre organismes ou entre pays s'avère impossible.

Cette situation n'étant pas satisfaisante, elle a donné naissance à un projet de la Commission européenne dont le but serait de définir des méthodes de mesures pour la numérisation des collections, qui pourrait s'appliquer à l'échelle de l'Europe et pourrait aussi être utilisé de manière permanente dans toutes les institutions culturelles européennes : NUMERIC¹

¹ <http://www.numeric.ws>

2. Le projet NUMERIC

En tant que projet de la Commission européenne, NUMERIC concerne la numérisation dans les pays européens et plus spécifiquement la numérisation du patrimoine culturel. Son objectif est de définir et tester des méthodes pour évaluer et décrire l'état d'avancement de la numérisation dans les institutions culturelles européennes. L'aspect financier et la valeur ajoutée des objets numérisés sont tous deux pris en compte.

Les interrogations auxquelles furent confrontées l'équipe du projet NUMERIC paraissent simples au premier abord. Les gouvernements, les fondations et d'autres établissements financeurs ont depuis des années dépensé des sommes considérables pour des projets de numérisation. Ce que ces organismes et le public veulent savoir est :

- Qu'est ce qui a déjà été accompli en matière de numérisation ?
- Combien cela a-t-il coûté ?
- Que reste-t-il à faire ?
- Combien cela va-t-il coûter ?

Pour répondre à ces questions, il est nécessaire de disposer des informations suivantes :

- Le volume et le type de collections numérisées dans les institutions
- Le pourcentage des fonds numérisés à une date donnée
- Les ressources (humaines et financières) allouées au projet
- Le pourcentage des collections « analogiques » susceptibles d'être numérisées à l'avenir
- Les ressources nécessaires à la continuité du projet

Pour un seul organisme, il serait facile de compiler de telles données. Mais pour collecter des données qui alimentent un panorama significatif et fiable, il faut absolument définir avec précision ce qui devra être compté et la manière de procéder.

Le projet fut dirigé par Phillip Ramsdale, de l'IPF (Institute of Public Finance, Chartered Institute of Public Finance and accountancy). L'équipe de recherche se constituait de neuf experts dans la numérisation culturelle et les statistiques, dans les archives, bibliothèques et musées.

3. Les différentes phases du projet NUMERIC

Pendant les deux années que dura le projet (de mai 2007 à mai 2009), plusieurs étapes se succédèrent :

- Pendant les six premiers mois, l'équipe examina les rapports et les sites Internet existants sur les projets de numérisation et identifia les différentes méthodes, les statistiques établies et les définitions utilisées.

- Puis le premier jet du questionnaire d'enquête fut rédigé et testé auprès d'un panel d'archives, de bibliothèques et de musées. Les soixante réponses obtenues permirent d'évaluer la qualité du questionnaire et d'importants changements y furent apportés.
- En avril 2008 se tint un groupe de travail au Luxembourg qui réunit une soixantaine de participants issus des 27 pays de l'Union européenne. Des coordinateurs furent désignés pour chaque pays ; charge à eux d'identifier des institutions « pertinentes » et d'en sélectionner au minimum 30 dans chaque pays. Les institutions « pertinentes » pour ce projet furent définies comme celles dont les collections ajouteraient une plus-value substantielle au patrimoine culturel du pays. En plus des archives, bibliothèques et musées, la sélection devrait aussi comporter des instituts audiovisuels et de radiodiffusion. Durant la rencontre au Luxembourg, une séance spéciale fut consacrée à la numérisation des monuments. Leur prise en compte étendrait le champ de l'enquête au patrimoine mobilier : il fut donc décidé de créer une plus petite enquête, spécifiquement pour les organismes gérants des monuments.

Après la sélection des institutions dans chaque pays et la traduction du questionnaire en 14 langues, l'enquête débuta en Juillet 2008.

Parmi les 27 pays, 5752 institutions furent jugées « pertinentes » pour la numérisation du patrimoine culturel. Parmi elles, un échantillon de 1539 avait été sélectionné pour répondre au questionnaire. Le taux de réponse fut de 51 %. Le tableau 1 montre le taux d'échantillonnage et les réponses par institution :

	Institutions « pertinentes »	Echantillon sélectionné	Nombre de réponses	Taux de réponse
Archives	848	262	133	51 %
Instituts audiovisuels ou de radiodiffusion	109	60	41	68 %
Bibliothèques	2754	690	222	32 %
Musées	1932	457	332	73 %
Autres	109	70	60	86 %
Total	5752	1539	788	51 %

Tableau 1 : nombre d'institutions « pertinentes », taille de l'échantillon et taux de réponse²

Le taux de réponse est évidemment supérieur lorsqu'il n'existe qu'un petit nombre d'un certain type d'institutions dans un pays, par exemple les instituts audiovisuels.

4. Le questionnaire

Puisque le programme de l'IFLA traite spécifiquement des statistiques du patrimoine culturel, cet article met l'accent sur le choix des méthodes qui répondraient le mieux à une vue d'ensemble nationale de la numérisation.

² NUMERIC. Study report. Study findings and proposals for sustaining the framework. May 2009, p. 25

Durant la première phase de NUMERIC, la phase de recherche, plus de trente rapports et études sur la numérisation furent analysés pour dénicher des mesures statistiques ayant prouvé leur efficacité. La majorité des enquêtes avait surtout retenu des indicateurs qualitatifs, par exemple la description des données numérisées ; en revanche, les données brutes manquaient. La plupart des études ne contenaient que l'état d'avancement de la numérisation à un moment précis sans prendre en compte l'ensemble des données sur plusieurs années. Enfin, la plupart des projets ne concernaient qu'un seul type d'établissement.

Les institutions qui fournirent le plus d'informations de nature à alimenter le projet NUMERIC furent les suivantes :

- CENL (Conference of European National Librarians) qui avait commencé à collecter des données dès 2007³
- EGMUS (European Group on Museum Statistics)⁴
- IMLS (Institute of Museum and Library Services)⁵

Afin de définir quelles données devraient être collectées, les normes ISO appropriées furent la principale source utilisée⁶. Elles fournissent en effet de nombreuses définitions et modes de calcul pour :

- Les types de documents dans les catalogues des bibliothèques
- Les différentes utilisations possibles
- Les coûts financiers (avec ou sans le coût en personnel et l'amortissement)

En revanche, les définitions et les modalités de calcul pour les archives et les objets conservés dans les musées n'ont pas encore fait l'objet d'une standardisation.

Le groupe d'expert décida que les aspects ci-dessous seraient analysés lors de l'enquête et devraient par conséquent donner lieu à une définition précise :

- Les supports analogiques numérisés (documents imprimés, matériel audiovisuel, manuscrits, objets de musées)
- La quantité de documents numérisés (difficulté du recensement)
- Les formats de la numérisation (par exemple : TIFF, OCR...)
- Les coûts engendrés par la numérisation (aussi bien pour les projets déjà réalisés que pour ceux en cours)
- Les moyens de financement de la numérisation
- L'accessibilité des documents numérisés
- L'utilisation des documents numérisés
- Les tâches restant à accomplir (relation des objets patrimoniaux ayant déjà fait l'objet d'une numérisation, par rapport à ceux en voie de l'être)

³ <http://www.cenl.org>

⁴ <http://www.egmus.eu/index.php?id=139>

⁵ IMLS. Technology and digitization survey. Available at :

http://www.imls.gov/publications/TechDig05/Archives_Survey.pdf

⁶ ISO 2789 (2006), Information and documentation – International library statistics – ISO 11620 (2008) Information and documentation – Library performance indicators.- ISO 5127 (2001) Information and documentation - Vocabulary

De plus, les différents types d'institutions destinataires du questionnaire devaient être clairement définis. Les groupes suivants furent ainsi établis :

- Archives/bureaux d'enregistrement
- Instituts audiovisuels
- Instituts de radiodiffusion
- Musées d'art, d'archéologie ou d'histoire
- Musées de science ou de technologie (ou d'ethnologie)
- Autres types de musées
- Bibliothèque nationale
- Bibliothèque de facultés
- Bibliothèques publiques
- Bibliothèque spécialisée ou autre type de bibliothèque
- Autre type d'institutions

5. Les types de supports analogiques et numériques

La première question fut : quels types de supports doivent être distingués? Concernant les bibliothèques, il était facile d'y répondre, car elles disposent généralement de statistiques très détaillées sur le nombre et le type de supports de leur catalogue. D'autres catégories comme les photos, les posters, les plans et même les tableaux peuvent être présents aussi bien dans des musées que dans des bibliothèques, voire dans des archives. Les enregistrements audio ne sont pas différenciés dans le questionnaire. Les objets de musées, s'ils ne sont pas considérés comme œuvre d'art, furent divisés en manufacture humaine et spécimen naturels du monde entier.

La principale difficulté pour évaluer le nombre de documents numérisés fut de définir des unités de mesure. Les documents imprimés pouvaient être dénombrés par volume, par numéro, par pages ou par feuilles. Les supports audiovisuels quant à eux pouvaient être comptés par boîtier ou par durée. Le tableau 2 montre les unités de mesure choisies.

Type de support	Compté en
Archives	Mètres, volumes ou nombre
Livres, publications en série	Volumes
Journaux	Numéro
Manuscrits	Nombre
Partitions musicale	Nombre
Microfiches et micro films	Nombre
Plans, photographies, gravures, estampes, dessins, cartes postales, posters et tableaux	Nombre
Tout autre objet à deux dimensions	Nombre
Œuvre d'art en trois dimensions	Objets
Artisanat fait main	Objets
Spécimen naturels du monde entier	Objets
Autres objets	Objets
Films, enregistrements vidéos.	Heures
Documents audio (musique et autres sons enregistrés)	Heures

Tableau 2 : Types de support et leur unité de mesure

Les résultats de l'enquête montrèrent que la plupart des institutions n'étaient pas en mesure de fournir des données sur certains documents de leur fonds. « Il apparaît clairement que pour beaucoup d'institutions, le recensement des collections analogiques s'avère aussi difficile que celui de leur fonds numérisés⁷ ». Pourtant, ces données sont essentielles pour calculer les coûts des projets de numérisation, qui sont très différents d'un support à l'autre.

6. Le financement de la numérisation

Dans le questionnaire, on demandait aux institutions si elles avaient consacré une part spécifique de leur budget à la numérisation. Seules 48 % d'entre elles répondirent par l'affirmative, y consacrant seulement une toute petite part, soit 1.1 %. L'enquête visait aussi à déterminer le nombre de salariés travaillant au projet de numérisation et quel coût, calculé en équivalent temps plein, cela représentait. La plupart des institutions furent incapable de répondre à cette question. Ces informations sont pourtant indispensables pour définir le coût d'une numérisation.

La majorité des répondants fut en mesure de citer la provenance de leur financement. Au total, les différentes sources se répartissent comme suit :

Origine du financement	%
Ressources internes	62.1
Programmes gouvernementaux	29.9
Dons privés	3.6
Autres	4.5

Dans tous les types d'institutions, les ressources internes constituent la principale source de financement, tandis que les programmes gouvernementaux représentent 30 %.

7. Le coût unitaire de la numérisation

L'enquête faisait aussi l'état des lieux sur les projets en cours et le financement prévu. Pour pouvoir comparer les coûts de la numérisation entre des supports différents, des unités de mesures comme « volumes » ou « mètre linéaire » furent convertis en « pages ». Pour ces pages, les coûts furent alors calculés en fonction des ressources allouées au projet.

Unité	Nombre de pages	Coût par page en euro
Volume (livre)	250	0.45
Volume (publication en série)	350	0.30
Journal	14	0.91
Manuscrit	45	8.74
Partition	23	0.68
Archives institutionnelles au mètre linéaire	768	0.74
Archives historiques au mètre linéaire	300	0.80
Autres archives au mètre linéaire	1868	0.80

Tableau 4 : coût par page numérisée

⁷ NUMERIC. Study report. Study findings and proposals for sustaining the framework. May 2009. p. 66

Les coûts des supports audiovisuels étant très variables en fonction des organismes et des projets, les taux suivants furent définis :

- Support audio : 30.00€heure
- Film : 55.20 €heure
- Vidéo : 34.29 €heure

8. Accessibilité des documents numérisés

La plupart des précédentes études sur la numérisation ne prenaient pas en compte les résultats obtenus, c'est-à-dire l'accès des utilisateurs aux données et l'usage qui en était fait. NUMERIC voulut savoir quels accès étaient rendus possibles, soit par des catalogues en ligne, soit par Internet, ainsi que la politique de diffusion choisie (accès gratuit, restreint...). Les questions suivantes furent donc posées :

- Un accès en ligne aux collections est-il possible et les fonds numérisés sont-ils signalés ? Tous les répondants ne disposent pas d'un catalogue en ligne et font encore moins la distinction entre les documents numérisés et analogiques. Comme il fallait s'y attendre, ce sont les bibliothèques qui possèdent les systèmes les plus performants. Dans les bibliothèques nationales par exemple, 95.5 % des répondants possèdent un catalogue en ligne et 88.9 % distinguent les différents types de documents. Au total, 67.4 % des organismes possèdent un catalogue en ligne, parmi lesquelles 62.2 % distinguent les documents numérisés.
- Quelle proportion du fonds numérisé est disponible en accès libre sur Internet ? Le taux médian est de 20 % mais varie considérablement d'une institution à l'autre. Les bibliothèques (70%) et les archives (48.5%) ont déjà mis en ligne sur Internet une grande partie de leurs fonds numérisés.
- Quelle est la politique d'accès aux données ? l'accès est-il gratuit (sans paiement ni restriction) ou payant avec des restrictions, par exemple réservé à la consultation sur place ? Environ 50 % des institutions dont 75 % de bibliothèques répondirent qu'elles offraient un accès gratuit et illimité. Il existe cependant des restrictions pour certaines parties des collections.

9. L'utilisation des collections numérisées

En élaborant le questionnaire, l'équipe de NUMERIC était consciente qu'il serait difficile de répondre à cette question et que les réponses risquaient de ne pas être suffisamment fiables pour être exploitées. L'utilisation des données dans les bibliothèques virtuelles est toujours un problème pour les statistiques des bibliothèques nationales, même pour ce qui concerne les publications à usage commercial, pour lesquelles les vendeurs fournissent un compteur. Il fallait donc s'attendre à ce que de nombreuses institutions culturelles n'aient pas encore trouvé la solution miracle pour déterminer quel usage est fait des données numérisées.

Les questions portaient sur le nombre de requêtes d'utilisateur pour des documents numériques, soit en ligne, soit hors ligne, par exemple en consultant un CD-ROM dans la bibliothèque. Les données sur les consultations en ligne étaient trop disparates pour fournir un résultat pertinent. Apparemment, les visiteurs uniques et ceux qui revenaient plusieurs fois ne pouvaient être distingués entre eux. En revanche, les données furent plus fiables pour la consultation hors ligne pour lesquelles 27 222 732 requêtes furent dénombrées.

Bien que ce premier essai pour évaluer l'utilisation des données numérique ne fût pas concluant, ce paramètre est néanmoins indispensable pour démontrer les bénéfices que peut engendrer une numérisation.

10. L'état actuel de la numérisation et ce qui reste à accomplir

Ce sujet, avec celui du coût unitaire, est probablement le plus intéressant pour tous les organismes financeurs. On peut comprendre qu'ils cherchent à connaître l'état d'avancement de la numérisation du patrimoine culturel et ce qui reste à accomplir. Le questionnaire posait donc trois questions :

- Quelle part des collections a déjà été numérisée ?
- Qu'est-il nécessaire de numériser ?
- Qu'est-ce qui n'a pas besoin d'être numérisé ?

La « nécessaire numérisation » peut s'expliquer par des raisons de conservation et /ou parce que l'intérêt des documents justifie d'en améliorer l'accessibilité à un public plus large.

«Ce qui n'a pas besoin d'être numérisé » concerne des documents dont l'ouverture à un plus large public ne présente pas d'intérêt à priori. Ce sont notamment les documents qui ne constituent pas une part importante du patrimoine culturel, qui font doublon avec d'autres collections ou qui ont déjà été numérisés par d'autres organismes. La proportion de documents qui n'ont pas besoin d'être numérisé est importante dans les bibliothèques, car elles possèdent beaucoup de documents en plusieurs exemplaires, et faible dans les musées où la plupart des objets sont uniques. Au total, les pourcentages sont les suivants :

- Déjà numérisés : 19.3 %
- N'ont pas besoin d'être numérisés : 30.2 %
- Restent à numériser : 50.5 %

La proportion varie considérablement entre les différentes institutions, mais il apparaît que partout, il reste beaucoup à faire. L'enquête demandait aussi la principale motivation à l'origine d'une numérisation, soit la conservation, soit l'accès à un plus large public. Mais les réponses montrèrent que les institutions ne pouvaient pas choisir entre ces deux aspects.

11. Le développement à venir

NUMERIC s'est achevé en mai 2009 mais de nombreuses actions existent pour tenter d'améliorer et d'exploiter les méthodes et les données collectées. L'équipe de NUMERIC a

utilisé les réponses pour estimer l'état d'avancement de la numérisation en Europe. Cela n'est possible qu'en tenant compte des spécificités de chaque pays et en compilant les données avec soin. Mais une seule ligne de conduite doit être suivie par les enquêtes à venir, si l'on veut estimer le niveau de développement de chaque pays.

Cette première expérience a montré que certaines définitions et procédés méritent d'être redéfinis. Le comité ISO approprié s'intéressera à cette problématique⁸. L'autre aspect essentiel concerne les institutions « pertinentes » pour le patrimoine culturel, qui gagneraient à être plus clairement définies et identifiées dans chaque pays selon ces critères.

De l'avis général des répondants, le questionnaire devrait être raccourci. Mais s'agissant d'une première enquête, il était nécessaire d'obtenir un panorama d'ensemble. Le prochain questionnaire pourrait se limiter aux questions qui mettent le mieux en évidence les apports et les résultats de la numérisation en Europe, dont des questions sur son influence (usages et utilisateurs) sur l'héritage culturel numérique. L'accent a été émis sur l'évaluation, pas seulement quantitatif mais aussi sur la valeur des projets pour l'apprentissage, la recherche et l'identité culturelle. Cet article donne une vue d'ensemble des résultats et des possibles évolutions de la structure du dispositif.

⁸ ISO TC 46 SC 8 Information and documentation – Quality, statistics and performance evaluation