

Insolite, Diadeis, Numérisation pharaonique du patrimoine européen
Insolite

Posté par : JulieM

Publié le : 30/10/2009 0:00:00

Le spécialiste de la numérisation patrimoniale participe à la démocratisation des ouvrages publiés par les Institutions Européennes.

Mise en service à l'occasion de la Foire de Francfort le 16 octobre dernier, la bibliothèque numérique EU Bookshop contient toutes les productions des différentes institutions, agences et autres organismes communautaires depuis 1952, rassemblées au sein de l'OPOCE (Office des Publications Officielles des Communautés Européennes).

Un travail de numérisation pharaonique à traiter en un temps record pour lequel l'OPOCE a choisi Diadeis.

Pour ce projet, Diadeis a ainsi traité environ 110 000 publications, représentant un volume d'environ 12 millions de pages, des plus petits formats publiés jusqu'aux formats supérieurs A0 (841 x 1189) et incluant des plans et des posters. Grâce à un savoir-faire unique, Diadeis se positionne aujourd'hui comme un acteur incontournable de la numérisation patrimoniale en Europe.



Une chaîne de numérisation unique en Europe

L'objectif de l'OPOCE était ambitieux : créer un fonds documentaire online, rassemblant toutes les publications des institutions européennes depuis 1952, afin de les rendre accessibles à tous. Cette bibliothèque numérique complètera également le portail en ligne multilingue Europeana.

Pour mener à bien ce projet, Diadeis, spécialiste de la numérisation patrimoniale, a mis en place une chaîne de numérisation spécifique dans son centre Eurodema : « Basée en Alsace, ce site de production unique en Europe pour son parc machines a été entièrement dédié au projet. », explique **Jean-Charles Morisseau**, Président Directeur Général de Diadeis.

« L'utilisation des 10 scanners spécialisés dans la numérisation de livres ouverts (dont 4 scanners permettant de tourner les pages des ouvrages de manière automatique) a permis le traitement à haute valeur ajoutée de centaines de milliers d'ouvrages en un temps record. »

Les 12 millions de pages ont ainsi été traités sur la durée très courte de 10 mois, ce qui représente 75 000 pages de livres reliés quotidiennement, en travaillant 24H/24 et 6 jours sur 7, et plus de 150 To de données (150 TO) de données.

Pour répondre au défi, Diadeis a conçu et mis en place un processus de numérisation sur-mesure, géré par un Workflow spécifique, permettant :

• La collecte et le stockage de 110 000 publications en vue d'un inventaire minutieux puis d'une préparation spécifique organisée de façon rationnelle en lignes de production.

• La numérisation des documents et des ouvrages reliés auxquels il a fallu apporter un soin particulier, grâce au déploiement de différents matériels et chaînes de traitement :

- 4 scanners automatiques « tourne-page » de type ASSY DL 3000 pour les ouvrages de plus de 100 pages en bon état
- 3 scanners semi-automatiques de type Copybook I2S pour les livres en mauvais état ou constitués de moins de 100 pages
- 1 scanner de grande largeur de type WideTEK 36 IMAGE ACCESS pour les documents de grande taille type plans ou posters
- 1 scanner plat de type WideTEK FB-A2 pour des photos ou des brochures allant jusqu'à un format A2

Tous ces types de scanners permettent la prise d'image TIFF 300 Dpi natif, en niveaux de gris ou en couleurs.

• L'amélioration de la qualité de chaque image afin de répondre totalement aux exigences de qualité requises, grâce à la mise au point d'un outil spécifique.

• Le traitement semi-automatique de l'extraction textuelle (OCR) sur les langues officielles de l'Union en vigueur.

• Le catalogage des publications selon 10 critères.

• Un traitement d'amélioration qualitative de l'OCR sur les tables des matières garantie à 99.975% avec mise du signet électronique associé.

• Un traitement automatique de masse d'encapsulation des données (images et texte OCR) vers le format de diffusion PDF archive [PDF/A-1b] image + texte version «Web-optimised» par publication.

Un outil de suivi, conçu et mis en place par Diadeis pour l'Office des Publications et accessible sur le web, permet de visualiser en temps réel l'avancement de la production.

Outre ce projet, Diadeis participe aussi au programme de digitalisation pour la Bibliothèque Nationale de France représentant plus de 300 000 ouvrages et 45 millions de pages, pour la Bibliothèque Royale de Belgique et la Bibliothèque Nationale de Luxembourg. Ces textes sont accessibles librement sur Europeana, le portail en ligne multilingue qui rassemble les ressources numériques des bibliothèques, archives et musées des pays membres de

l'Union. Une fois les publications de la Librairie Européenne intégrées, ce seront près de 44% des textes présents dans Europeana qui auront été traités par Diadeis.

« Nous sommes particulièrement heureux de pouvoir participer à cette grande aventure qui met le patrimoine européen à la portée de tous et l'inscrit dans l'évolution des pratiques culturelles et des technologies », conclut **Jean-Charles Morisseau**.