

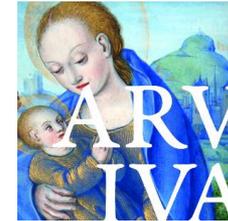
Sobriété numérique :
attention à la course aux
pixels !

Numérique et numérisation : quels modèles pour
un développement durable ?

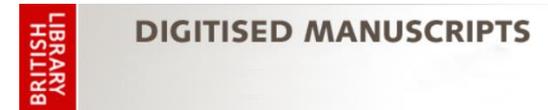
Gilles Kagan

La dématérialisation des documents patrimoniaux

- Des milliers de bibliothèques



Initiale



Interopérabilité des données

- Un cloud de données souvent hétérogènes

Les portails et les bases catalogues



L'IRHT en quelques mots...

Couverture photographique des fonds patrimoniaux dispersés sur tout le territoire français depuis 1976.

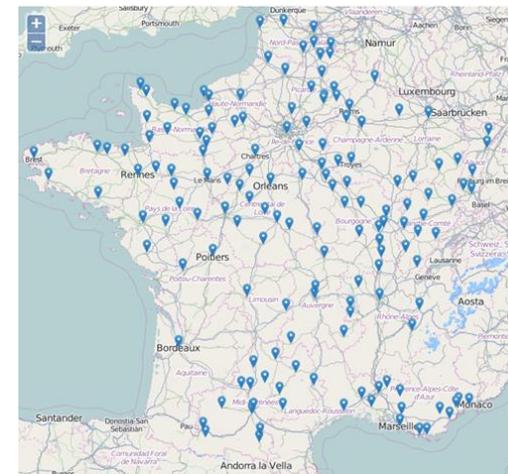
Bibliothèques municipales, universitaires, centres d'archives... :

39 000 reproductions de manuscrits de bibliothèques françaises (**Medium**)

27 494 cotes en ligne dans la **BVMM** (2 870 892 vues)

210 000 reproductions d'enluminures et de reliures (**Initiale**)

23 000 liens (**bases scientifiques** et bases catalogues)



Une mission nationale, soutenue par le SLL



A partir de 2005, la numérisation intégrale des manuscrits en couleur remplace peu à peu l'argentique.

Le Pôle numérique

Le pôle numérique est composé de quatre équipes :

- Systèmes, réseau et archivage
- Développement et interopérabilité
- Images
- Publication

Créé en 2012, ce pôle réunit en une seule entité le personnel compétent dans les technologies complémentaires de la photographie numérique, de l'informatique, de l'édition et de l'infographie.

Des programmes internationaux

La numérisation de manuscrits entre dans pratiquement tous les programmes de recherche de l'IRHT, des projets soutenus par Biblissima, les ANR, l'ERC....



CLAMM

Classification of Latin Medieval Manuscripts

CORPUS / PALÉOGRAPHIE LATINE



Le corpus d'images d'écriture a fourni la matière de deux compétitions (« Competition on the Classification of Medieval Handwritings in Latin Script ») tenues dans les colloques internationaux d'analyse d'image par ordinateur ICFHR (International Conference on Frontiers in Handwriting Recognition) en 2016 et ICDAAR (International Conference on Document Analysis and Recognition) en 2017. À partir des images annotées, les compétiteurs ont dû développer des logiciels capables de classer des images selon le type d'écriture et procéder à leur datation selon des types ou des intervalles chronologiques définis à l'avance (par exemple au quart de siècle pour le xve siècle).

La transmission des textes

Training School 2018

Saint-Bertin, centre culturel du VII^e au XVIII^e siècle

constitution, conservation, diffusion, utilisation du savoir

Une nouvelle perspective sur l'histoire culturelle

Abbaye bénédictine fondée au VII^e s., Saint-Bertin (*Sithiu*), sise à Saint-Omer, et sa bibliothèque de manuscrits sont une ressource de premier ordre pour l'histoire intellectuelle européenne. Centre économique et culturel majeur dès les temps carolingiens, elle est aussi un carrefour entre les espaces français, anglais et rhénans.

Les prélats à la tête de cette abbaye assument un rôle à la fois intellectuel et politique durant le Moyen Âge, et au XVI^e siècle, la bibliothèque sert aux imprimeurs et érudits (François Modius, Allard Tassart, Jacques Malbranque) et donne ainsi accès à des textes importants.

Abbaye de Saint-Bertin, gravure par G. Delonnois, vers 1870

A la recherche des manuscrits de Chartres

Etude et renaissance virtuelle d'un fonds de manuscrits sinistré

une richesse à Saint-Bertin.

Accueil Corpus Projet Rayonnement Contact

- Manuscrits
- Bibliographie
- Rechercher

Auteurs chartains

- Bernard de Chartres
- Bernard Silvestre
- Claude d'Anas
- Fouquier de Chartres
- Fulbert de Chartres
- Olibert de La Pomerie
- Oultraire de Conches
- Jean de Salisbury
- Jean Le Marchant
- Thierry de Chartres
- Yves de Chartres

Le prestigieux fonds de la Bibliothèque municipale de Chartres était le témoin de l'histoire de toute une région et notamment de sa célèbre école cathédrale, un des plus grands centres intellectuels d'Occident aux XI^e et XII^e siècles. Le 26 mai 1944, un incendie l'anéantit.

Des 519 manuscrits médiévaux à l'origine, 165 subsistent dans des états très variables, du manuscrit presque intact au bloc carbonisé en passant par des fragments recroquevillés. Privés de leurs éléments d'identification habituels, ils étaient depuis près de soixante-dix ans inaccessibles à la recherche.

Le projet piloté par l'Institut de recherche et d'histoire des textes (CNRS) en collaboration avec plusieurs partenaires comprend quatre volets :

- restauration des manuscrits,
- reproduction numérique,
- identification des textes par des spécialistes,
- mise en ligne des fac-similés numériques accompagnés d'une description scientifique des manuscrits.

Origine: Commentaire sur l'Épître aux Romains, Chartres, XII^e siècle (2^e quart). Chartres, B.M., ms. n. 54

patrimoine écrit du Moyen Âge et de la Renaissance

Annel est contenu Lart & Science
e & vraye Proportio des Let
ques, qu'o dit autrement Let
iques, & vulgairement Let
naines proportionnees selon
le Corps & Vilage humain.

De multiples ressources en libre accès sur internet

18 bases scientifiques

BASE DE DONNÉES BASE DEBATE ↗ DEBATE est un outil de dépôt qui vise à collecter et à rassembler des informations historiques, des preuves textuelles, des thèses doctrinales, des traces biographiques d'auteurs qui ont exécuté <i>Principia</i> .	BASE DE DONNÉES BASE MEDIUM ↗ Medium est un répertoire normalisé des cotes de manuscrits, incunables et livres anciens du Moyen Âge et de la Renaissance.	BASE DE DONNÉES BIBALE ↗ Bibale décrit l'histoire de la transmission des livres manuscrits et imprimés et des textes qu'ils contiennent par l'étude des collections anciennes et modernes et de leurs possesseurs.
BASE DE DONNÉES BIBLIOTHÈQUE VIRTUELLE DES MANUSCRITS MÉDIÉVAUX ↗ La Bibliothèque virtuelle des manuscrits médiévaux permet de consulter la reproduction d'une large sélection de manuscrits, du Moyen Âge au XVI ^e siècle.	BASE DE DONNÉES BUDÉ ↗ La base Budé décrit la transmission des œuvres antiques et médiévales par les manuscrits et les imprimés anciens pour une période qui couvre la fin du Moyen Âge et la Renaissance.	BASE DE DONNÉES CALD ↗ La base de données CALD est le principal outil de travail du projet européen ILM (Islamic Law Materialized) qui étudie, dans une perspective comparative, des documents arabes légaux, élaborés en terres d'Islam entre le VIII ^e et le XVI ^e siècle.
BASE DE DONNÉES CALD ↗ CALD s'est fixé pour objectif l'étude comparative des documents arabes légaux, élaborés en terres d'Islam entre le VIII ^e et XVI ^e siècles, et conservés comme actes authentiques.	BASE DE DONNÉES CARTULR ↗ CartulR est consacré aux cartulaires «français» principalement conservés sur l'actuel territoire de la France et, pour un certain nombre de cas, de la Belgique.	BASE DE DONNÉES COLLECTA ↗ Collecta est une base de données qui reconstitue virtuellement la collection de François-Roger de Gaignières (1642-1715) dispersée entre différents lieux de conservation.
BASE DE DONNÉES COMPARATIO ↗ Comparatio facilite l'identification des provenances de manuscrits liturgiques de l'office par l'étude et la comparaison de différentes versions d'antennes et de répons.	BASE DE DONNÉES E-KTOBE ↗ E-ktobe : manuscrits syriaques a pour but de rassembler les informations sur les textes, les aspects matériels, et l'histoire (colophons et notes) des manuscrits syriaques de toute période et de tout lieu.	BASE DE DONNÉES FAMA ↗ FAMA rassemble des renseignements sur le nombre de témoins subsistants des ouvrages les plus lus du Moyen Âge latin et les regroupe par nombre d'exemplaires, par genre, par pays et par date.
BASE DE DONNÉES INITIALE ↗ Initiale est un catalogue de manuscrits enluminés du Moyen Âge lié aux campagnes de reproduction systématique des manuscrits enluminés que	BASE DE DONNÉES ITER LITURGICUM ITALICUM ↗ La base <i>Iter Liturgicum Italicum</i> est un répertoire des manuscrits liturgiques italiens conservés soit en Italie soit dans les bibliothèques et archives du	BASE DE DONNÉES JONAS ↗ Jonas est un répertoire exhaustif de textes et manuscrits en ancien et moyen français et en occitan, hors documents d'archives.

25 sites (corpus, outils en lignes)

CORPUS BIBLIA GLOSSATA ↗ La Glose ordinaire est un corpus de sentences ou explications brèves, disposées dans les marges et l'interligne du texte de la Bible (dite glosée).	OUTILS CALENDSOCPE ↗ Calendoscope est destiné à aider les spécialistes dans l'analyse et l'identification des calendriers liturgiques médiévaux.	CORPUS CLAMM ↗ Corpus de 8 000 images d'écriture réalisé à partir de la documentation photographique des Catalogues des manuscrits datés pour l'analyse d'image par ordinateur.
OUTILS CMD-F ↗ L'index général interactif des Catalogues des manuscrits datés de France couvre les sept premiers tomes, parus de 1959 à 1984 et recense tous les manuscrits mentionnés.	OUTILS CODICOLOGIA ↗ Codicologia est un vocabulaire multilingue pour la description des manuscrits, instrument de travail à tous les catalogueurs de manuscrits et éditeurs de textes anciens.	SITE DEBATE ↗ <i>Principia</i> est un genre largement inexploré qui correspond à un exercice complexe obligatoire imposé par les statuts des universités médiévales de toute l'Europe au cours des XIV ^e et XV ^e siècles.
OUTILS DIKTYON ↗ Diktyon est un réseau scientifique de ressources électroniques et de base de données sur les manuscrits grecs.	OUTILS GLOSSAIRE DU LATIN PHILOSOPHIQUE ↗ Ce glossaire recueille les définitions de plus de 35 000 concepts relatifs au vocabulaire philosophique du Moyen Âge.	OUTILS GRAPHOSKOP ↗ Graphoskop est un outil informatique destiné à aider le paléographe dans l'analyse des écritures anciennes.
SITE I-STAMBOUL ↗ Ce site est consacré à l'histoire de la bibliothèque du monastère de la Sainte-Trinité, sur l'île de Chalki (au large de Constantinople).	SITE IDEAL ↗ lDeAL est une plate-forme de publication permettant la mise en ligne de documents liés aux travaux des chercheurs.	SITE ISLAMIC LAW MATERIALIZED ↗ Le projet étudie, dans une nouvelle perspective comparative, des documents arabes légaux, témoins directs de la pratique légale, qu'ils soient publiés ou inédits.
CORPUS LUXURY BOUND ↗ Cette publication présente un corpus de plus de 3 800 manuscrits médiévaux illustrés dans les Pays-Bas.	SITE MANUSCRITS DE CHARTRES ↗ Ce site vise à rendre à nouveau consultables les manuscrits de la Bibliothèque municipale fragilisés par la flamme et l'eau en 1944.	OUTILS MILLESIMO ↗ Millesimo est un instrument de chronologie événementielle et technique (selon le calendrier julien pour les années 1 à 1600 de l'ère chrétienne) à l'usage des spécialistes des époques patristique et médiévale.

Prise de conscience dans les SHS ?

Journées Eveille 2022 consacrées à la notion :
Écosophie du numérique.

Promouvoir de nouvelles pratiques pour:

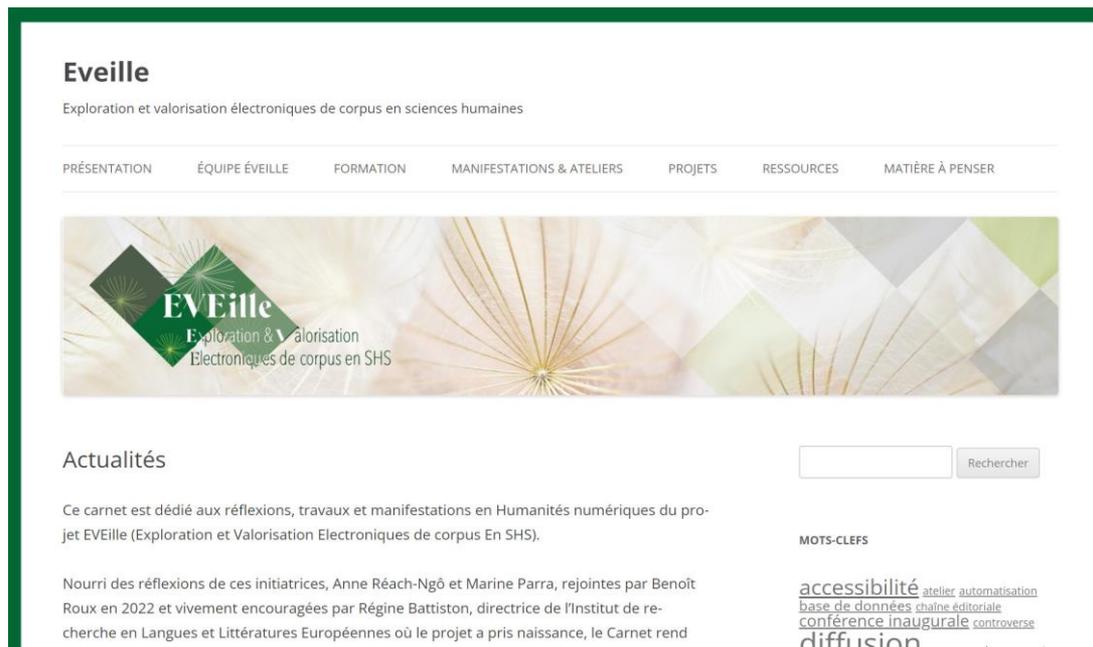
Ralentir la production des données inutiles ou superflues

Favoriser les circuits courts et mutualiser des outils

Permettre la mise en commun des savoirs et de la créativité

Revaloriser les liens qui nous
attachent les uns aux autres
ainsi qu'à notre environnement

eveille.hypotheses.org



The screenshot shows the homepage of the website 'Eveille'. The header includes the title 'Eveille' and the subtitle 'Exploration et valorisation électronique de corpus en sciences humaines'. Below the header is a navigation menu with links for 'PRÉSENTATION', 'ÉQUIPE ÉVEILLE', 'FORMATION', 'MANIFESTATIONS & ATELIERS', 'PROJETS', 'RESSOURCES', and 'MATIÈRE À PENSER'. A large banner image features the 'Eveille' logo and the text 'Exploration & Valorisation Electroniques de corpus en SHS'. Below the banner, there is a section titled 'Actualités' with a search bar and a 'Rechercher' button. The main content area contains a paragraph about the project's focus on digital humanities and a list of keywords under the heading 'MOTS-CLEFS', including 'accessibilité', 'atelier', 'automatisation', 'base de données', 'chaîne éditoriale', 'conférence inaugurale', 'controverse', and 'diffusion'.

Comment mettre en place un modèle soutenable ?

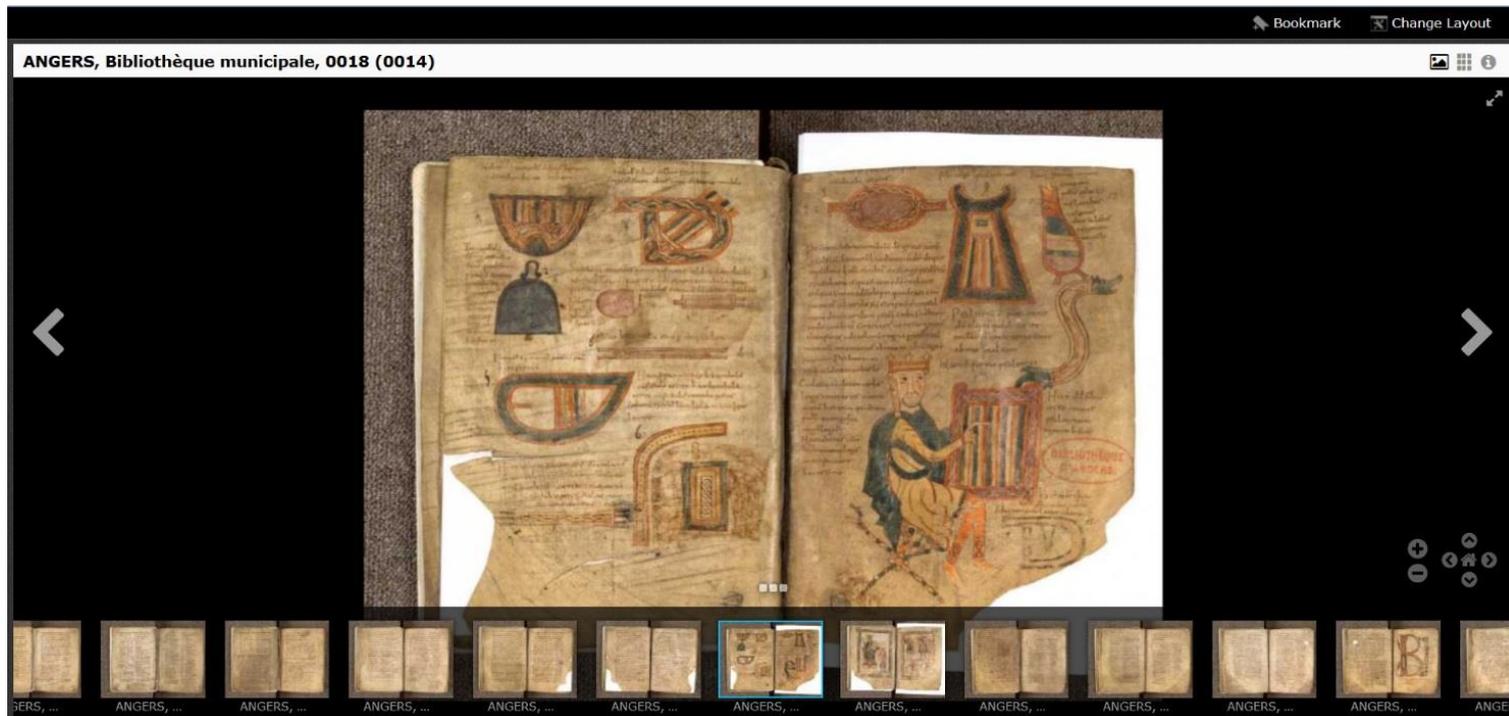
Dès 2016, une réflexion est menée par les composantes du pôle numérique de l'IRHT pour réduire notre empreinte carbone et les coûts :

- Diminuer le coût du **stockage** pour la diffusion et l'archivage des images
- Appliquer les principes du **FAIR** pour que les données de l'IRHT soient « Faciles à trouver, Accessibles, Interopérables, Réutilisables »
- **OPIDOR** (plan de gestion de vos données)
- Adapter la résolution des documents en ligne aux besoins (recherche et patrimoine)
- Créer des interfaces d'interrogation intuitives, accessibles et multiplateformes (responsive)
- Développer les outils et les applications à partir d'API (Application Program Interface)
- Utiliser des technologies normalisées, ouvertes et pérennes (logiciels libres)

BVMM Bibliothèque virtuelle des manuscrits médiévaux

- Une application full web totalement ouverte

Biblissima



Un système ouvert et soutenable

Technologie API Image IIF

International Image Interoperability Framework

- Une technologie Open Source permettant la visualisation d'un objet numérique à partir d'une API (service Web).
- Objectif : mettre en ligne des ressources de manière standardisé, images, notices....
- Exposer des contenus interopérables (Format JSON-LD).
- Permettre un cadre technique commun et standardisé pour communiquer les images
- Un consortium qui réunit plus de 80 membres

La BVMM un bon exemple où se côtoient actuellement deux technologies....

Evolution du système de visualisation

Architecture informatique :

La base a été développée par l'IRHT à partir de **2008** et s'appuie sur une API développée en php et Symfony .

L'interface Frontend a été développée en php et Symfony. Elle utilise Bootstrap , html 5 et JQuery.

Restructuration de nos données, refonte des bases documentaires de l'IRHT :

ARCA = fusion des deux bases documentaires de l'IRHT : **BVMM + Medium**

- Meilleure gestion des reproductions liées à l'ensemble des bases scientifiques de l'IRHT, Bibale, Jonas, Initiale etc.
- Nouvelle interface plus intuitive (responsive)
- Amélioration au niveau des recherches et des réponses (filtres)
- Nouvelle interface de visualisation des images

Une nouvelle interface : ARCA (BVMM + MEDIUM)



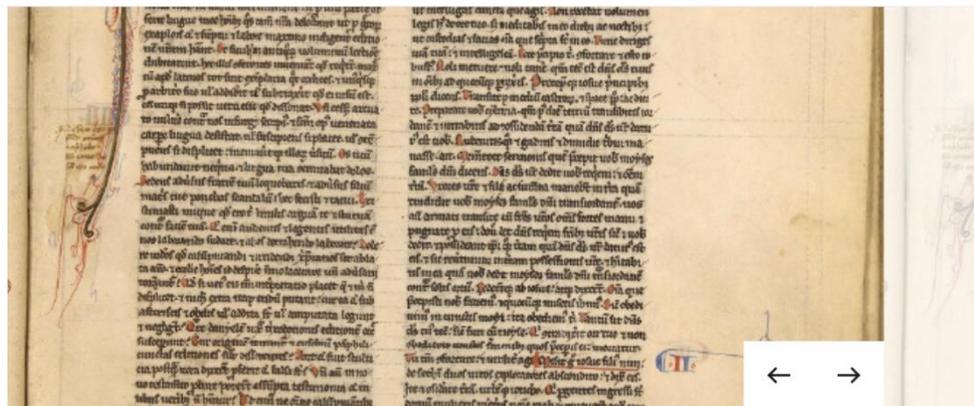
Recherche guidée

Recherche avancée

ARCA

Bibliothèque numérique de l'IRHT, répertoire des cotes de manuscrits, incunables et livres anciens conservés dans les bibliothèques du monde entier.

En savoir plus



Recherche libre

Recherche par cote

EVARDUS de VALLE SCOLARIUM

Manuscrits



Dispose d'une reproduction Numérisation en ligne

1 résultats



France, Tours Bibliothèque municipale, 0483

14e s. (début) ; 16e s.

EVARDUS de VALLE SCOLARIUM, Summa sermonum de festis ; Office noté pour la fête des Sept-Dormants de Marmoutier

Filtrer par

Tours

Langue

Type de reproduction

Filter

Une nouvelle interface : ARCA (BVMM + MEDIUM)

arca.irht.cnrs.fr/manuscript/35551



Galerie

The screenshot displays the ARCA manuscript gallery interface. On the left, a vertical sidebar contains navigation icons: a plus sign, a bookmark icon with a '1', a gear, three dots, and a square icon. The main area shows a grid of 21 manuscript images, each with a caption below it. The first image, 'plat supérieur', is highlighted with a red border. The grid is organized as follows:

- Row 1: plat supérieur, plat inférieur, dos, contreplat supérieur - garde recto, garde verso - f. 001r avec réglet, garde verso - f. 001r, f. 001v - 002r
- Row 2: f. 001v - 002r avec réglet, f. 002v - 003r, f. 003v - 004r, f. 004v - 005r, f. 005v - 006r, f. 006v - 007r
- Row 3: f. 007v - 008r, f. 008v - 009r, f. 009v - 010r, f. 010v - 011r, f. 011v - 012r, f. 012v - 013r
- Row 4: f. 013v - 013 bis, f. 013 bis v - 014r, f. 014v - 015r, f. 015v - 016r, f. 016v - 017r, f. 017v - 018r

At the bottom left, there is a logo consisting of three vertical bars of varying heights. At the bottom right, there is a vertical scrollbar.

Liens vers les bases scientifiques

Medium

Répertoire des manuscrits
reproduits ou recensés

MANUSCRITS

1 à 49 de 49

FILTRES ACTIFS

PAYS: France(x)

Ville

- Aix-en-provence [1]
- Alençon [1]
- Amiens [1]
- Arras [1]
- Auch [1]
- Auxerre [1]
- Bordeaux [1]
- Boulogne-sur-mer [1]
- Cambrai [1]
- Carpentras [1]
- Chantilly [1]
- Charleville-mézières [1]
- Chartres [1]
- Clermont-ferrand [1]
- Dijon [1]
- Douai [4]
- Grenoble [2]
- Laon [1]
- Lyon [1]
- Metun [1]
- Montpellier [2]

Interrogation par lieu de dépôt et cote du manuscrit ou par contenu.

Pour une recherche approfondie : [Medium, édition avancée](#)

%Etymologiae %



Recherche floue

IRHT
Institut de recherche
et d'histoire des textes

France, Aix-en-Provence, Bibliothèque municipale, 0025 (0914)

ISIDORUS, *Etymologiae* - 12e s. (milieu)



FAMA

Initiale

[Medium - édition avancée](#)

BVMM (Décor)



France, Alençon, Bibliothèque municipale, 0637

(Autre cote : France, Alençon, Bibliothèque municipale, Saint-Martin de Sées, T. f. 21)

ISIDORUS, *Etymologiae* ; fragm. d'Aspremont en français - 12e s. (début)



Bibliographie

Jonas

[Medium - édition avancée](#)

BVMM (Num.)



France, Amiens, Bibliothèque municipale, 0221

ISIDORUS, *Etymologiae* ; MARBODUS, Carmina - 12e s. (troisième quart)



FAMA

Initiale

[Medium - édition avancée](#)

BVMM (Décor)



BVMM (Num.)



Comment réduire notre empreinte carbone ?

Mise en place progressive d'un modèle numérique plus vertueux pour la diffusion (BMMM) et l'archivage des images (+métadonnées)

- 2016, l'archivage pérenne (OAIS) est mis en place à l'Irht avec l'équipe HumaNum et le Cines de Montpellier // convention CNRS -MCC
- 2017, abandon du format Tif (archivage) et Jpg (diffusion)
 - **Choix du format JPG 2000** : 1 seul fichier, gain environ 25 To en moins
- Structurer les métadonnées selon des standards pour préparer l'interopérabilité

JPG 2000 : meilleur sur le plan qualitatif (algo. en ondelette)



Fichier Jpg, agrandissement 800 %

Fichier Jpg 2000, agrandissement 800 %



Jpg 2000: réduit l'empreinte carbone (énergétique)



Gain de 60 % sans modifier la taille en pixels

Tif = 14 Mo (2000 x 2284 pixels)

Jpg = 1,7 Mo

Jp2 = 755 ko

Le JP2 un format plus **robuste** et plus performant pour la diffusion on line plus **pérenne**, ouvert et utilisé par une communauté importante

Préconisations du format jp2 sur le site de la BNF et sur le site de Biblissima :

<https://www.bnf.fr/fr/les-referentiels-de-numerisation-de-la-bnf>

<https://biblissima.fr/>

Evolution des l'infrastructure et des volumes physiques

L'IRHT Orléans il y a 10 ans

- 15 serveurs physiques
- 6 baies de stockage
- 9 switchs réseaux
- 3 KVM

L'IRHT Orléans aujourd'hui

- 10 serveurs physiques
- 4 baies de stockage
- 5 switchs réseaux
- 1 KVM

L'IRHT Paris il y a 10 ans

- 8 serveurs physiques
- 1 baie de stockage
- 2 switchs réseaux
- 1 KVM

L'IRHT Paris aujourd'hui

- 1 serveur physique

La masse de matériel physique a diminué au profit de solutions plus économes et moins énergivores en utilisant l'évolution des technologies VMware et en homogénéisant l'infrastructure en un seul endroit à Orléans.

Un volume de stockage plus important et moins énergivore

L'IRHT Orléans il y a 10 ans

- Une volumétrie de 80 To

L'IRHT Orléans aujourd'hui

- Une volumétrie de 250 to dont 100 To pour les images

VMWare a permis de s'affranchir des serveurs physiques en permettant la virtualisation des serveurs tout en augmentant par 5 le nombre de serveurs en mode virtuel.

Attention à la course aux pixels

1992 Numérisation HD de la photothèque (18 Mo)

1998 Premières numérisations directes des éléments du décor. (35 Mo)

2000 Acquisition d'un Power Phase FX monté sur une chambre 4 X 5 Linhof dans les BM de Valenciennes, Douai, Cambrai (50 à 200 Mo)



Scanner Kodak 2035 plus (2000 x 3000 pixels)



Dos phase One studio Kit (3500 x 4200 pixels)



Technologie Tri-CCD Linéaire (10 200 photosites 380 Mo)

Attention à la course aux pixels

2002 Premier dos matriciel Phase One (20 Mo)

2004 Mise en service du Phase One P 25 (65 Mo)

2006 Mise en service d'un Hasselblad H2D 39 Mo (96 Mo)

2017 Installation d'un numériseur en berceau BC 100 (155 Mo)



Light Phase 2000 x 3000 pixels



Phase one P 35+ (5500 x 4200 pixels)



Phase One XF + IQ 150 (6208 x 8280 pixels et 8172 x 14199 pixels)

(ouverture à 100 °).

Résolution de sortie et taille du document

Numérisation d'un grand manuscrit in folio : hauteur de 596 mm



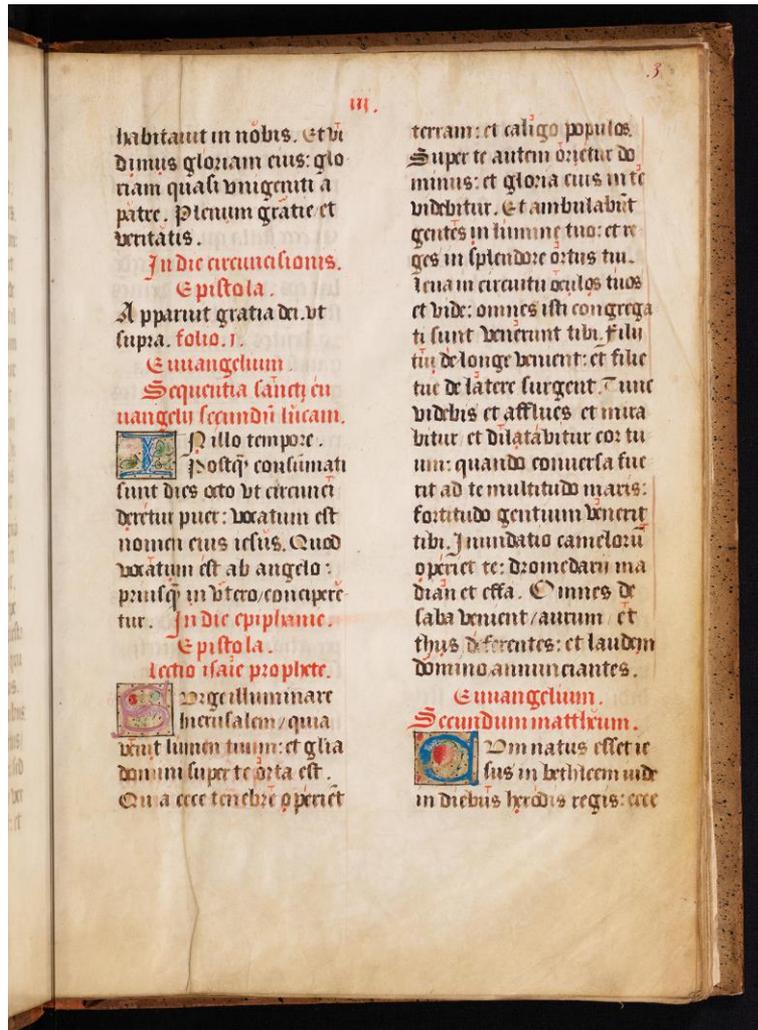
96,8 Mo (4849 x 6975 pixels) : affichage à 100 %



268,8 Mo (8080 x 11625 pixels) : affichage à 100 %

Résolution de sortie et taille du document

Numérisation d'un petit manuscript in quarto : hauteur de 262 mm



114,8 Mo (5380 x 7432 pixels) affichage à 100 %



317 Mo (8666 x 12387 pixels) affichage à 100 %

Ne pas mélanger sobriété et efficacité

Simplification des chaînes de traitement

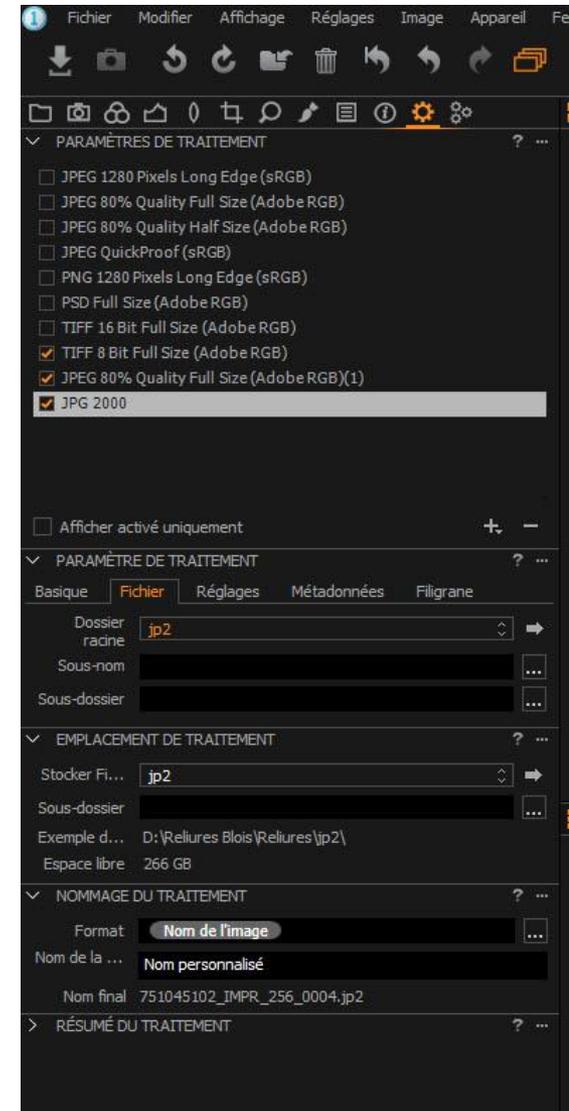
Fourniture des manifests IIF aux bibliothèques

Privilégier les outils et plateformes en libre accès comme :

- hypothèses,
- Open Edition,



Sulpice Sévère expédiant son ouvrage
Saint Martin partageant son manteau, Tours, ms. 1023



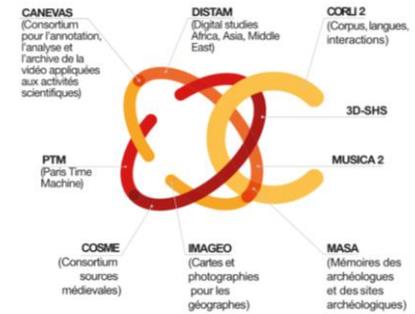
Bien préparer les données (interopérables).

- Structurer les données selon des standards (Dublin Core)
- Fournir pour chaque document des **métadonnées minimales** permettant d'identifier la source (titre, auteur, date, éditeur, licence, etc.)
- Pour le signalement utiliser des référentiels en lignes :

Geonames (localisation)

Periodo (chronologie)

- Utiliser des identifiants pérennes (ARK, Handle, DOI)
- Exploiter les services Web pour récupérer dynamiquement les informations
- Travailler en réseaux en s'appuyant sur un consortium et une ontologie lorsque celle-ci existe



Cycle de vie des outils numériques

Utilisation de technologies plus responsables en s'appuyant sur la règle des 4 R :

- Réduire
- Réparer
- Réemployer
- Recycler et/ou renoncer (résilience)

Force est de constater que l'innovation numérique est plus aujourd'hui au service de la croissance et non de la sobriété

Lutter contre l'obsolescence programmée des outils de production ou de diffusion

Cycle de vie des outils numériques

Conclusion:

Réfléchir collectivement à la mise en place de modèles plus durables au niveau des infrastructures, mieux évaluer les besoins des utilisateurs, être attentif aux changements sociétaux et comportementaux.

Il y a un important travail à faire sur l'usage (utile) que nous avons des nouvelles technologies. Savoir utiliser les outils numériques, c'est aussi savoir prendre en compte l'aspect environnemental.

L'éducation est la meilleure façon de changer les comportements :

<https://m1ehst1920.hypotheses.org/284> Sarah Descamps