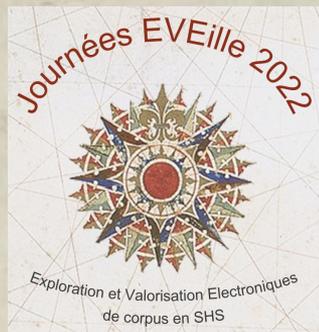


Bibliothèques

Virtuelles

Humanistes



Centre d'études supérieures
de la *Renaissance*



Refonte de l'architecture numérique et du modèle de données : État des lieux, problématiques et perspectives

Sarra Ferjani, Marlène Arruga, Toshinori Uetani
Anne-Laure Allain, Sandrine Breuil

Séminaire EVEille
Session *Chaînes de production*
21/01/2022





Bibliothèques

Virtuelles

Humanistes



Etat des lieux

| Historique, objets, contexte scientifique et numérique

| Présentation de l'infrastructure actuelle

| Refonte : Pourquoi et comment ?



BHH

Bibliothèque numérique de référence pour la Renaissance

- **Fac-similés** - Documents originaux du XV^e au XVII^e siècle sélectionnés principalement dans les fonds patrimoniaux de la Région Centre-Val de Loire
- **Epistemon** - Édition numérique de textes français du XVI^e siècle

Double chaîne de production

Identification des phases de traitement numérique et des compétences correspondantes au fur et à mesure de l'évolution.

The screenshot displays the BVH website interface. At the top, there are navigation tabs for 'Fac-similés', 'Epistemon', 'Manuscrits', 'Iconographie', and 'Projet'. Below this, there is a search bar and a navigation menu. The main content area features a large image of a historical manuscript page with a botanical illustration of a plant. To the right of the image, there is a detailed metadata section in French, including fields for 'Auteur', 'Titre', 'Imprimeur', 'Libraire', 'Lieu de publication', 'Date', 'Format', 'Cotation', and 'Notes'. Below the image, there is a section for 'Adresse typographique' and 'Etat physique'. At the bottom of the page, there is a footer with various logos and contact information.



Bibliothèques

Virtuelles

Humanistes

BVH - Historique

Évolution du paysage numérique

Prospection documentaire
et préparation technique

2003

Epistemon
1^{ère} édition numérique
XML-TEI

2007-2014

Début du projet
*Bibliothèques
françaises*
bibfr.bvh.univ-tours.fr

Engagement *pro-actif*
de la refonte BVH
(site web éditorial,
production, modèle de
données...)

2001-2002

2007

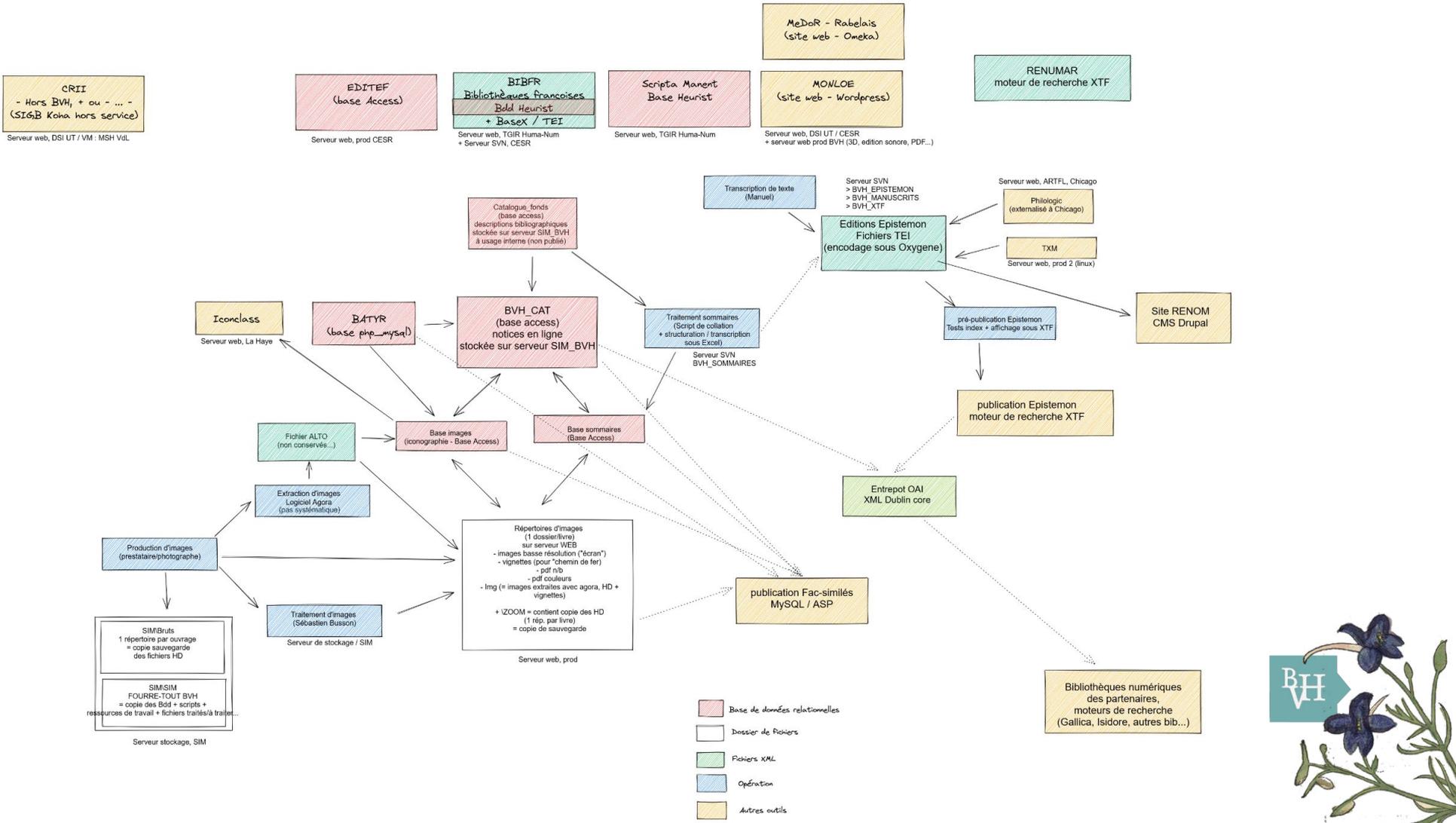
2015

Depuis 2018 - ...

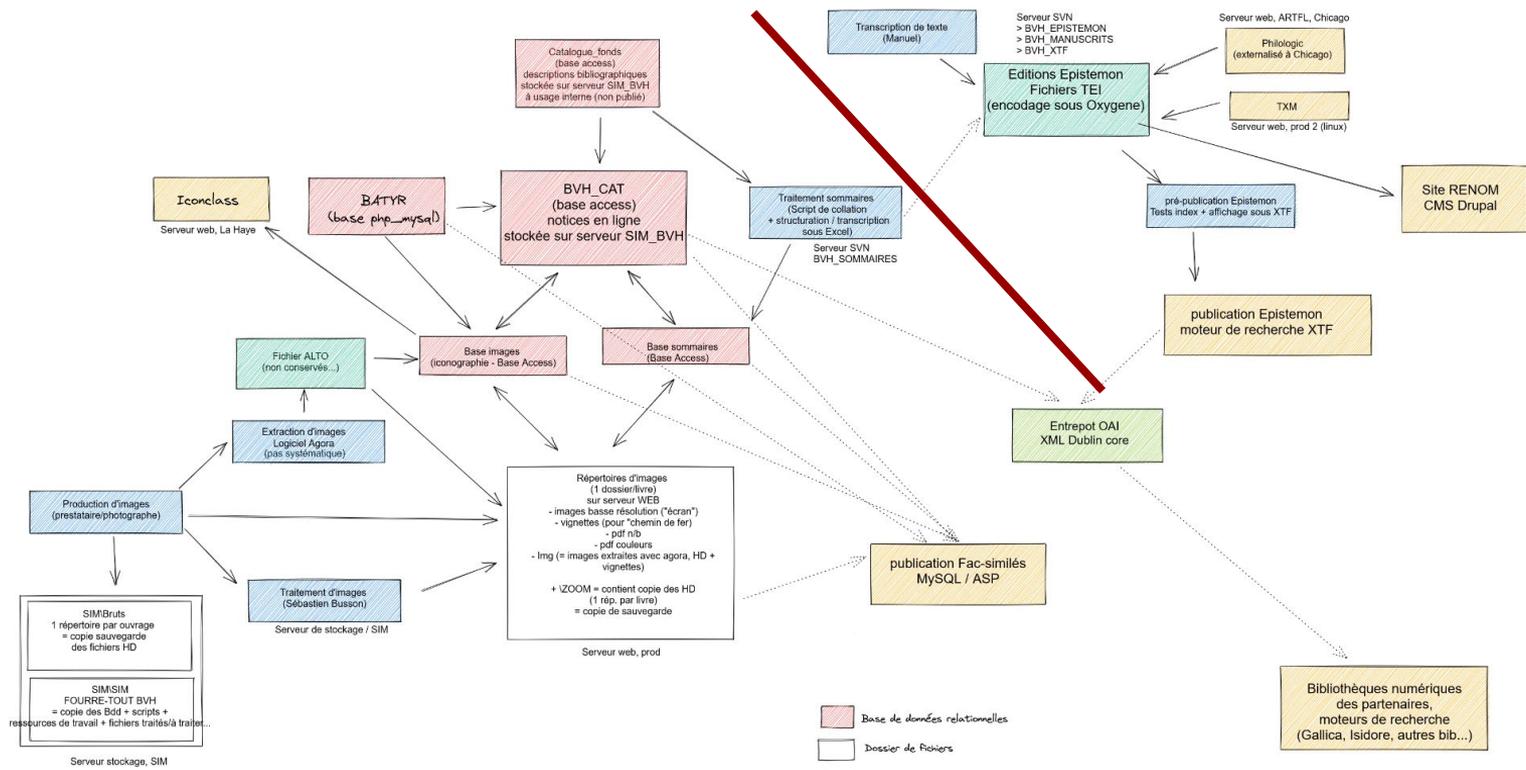
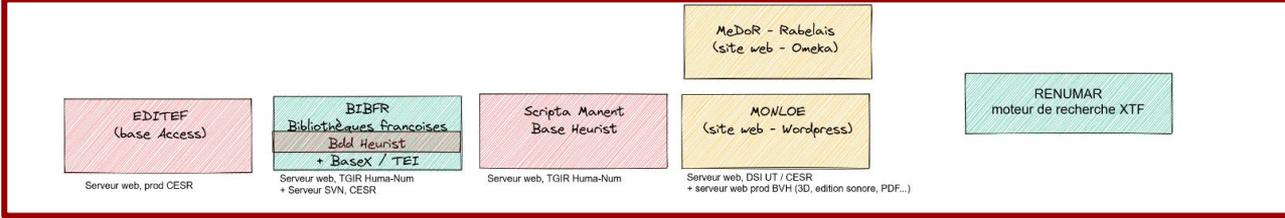
Fac-similés
1^{ères} numérisations
en ligne
www.bvh.univ-tours.fr

Évaluation & Réorientation



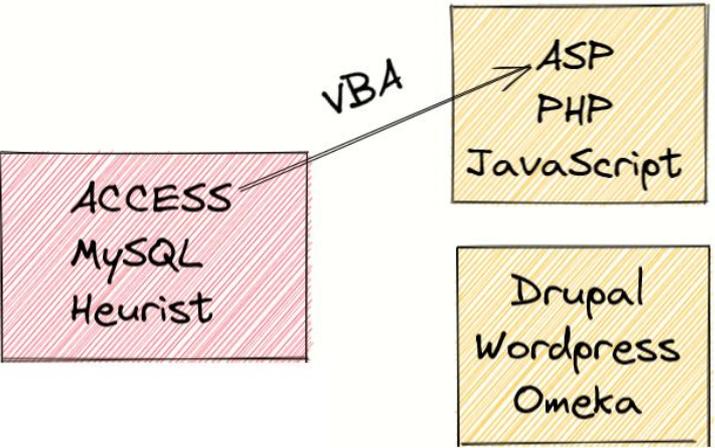


CRII
- Hors BVH, + ou - ... -
(SIGB Koha hors service)
Serveur web, DSI UT / VM : MSH Vdl.



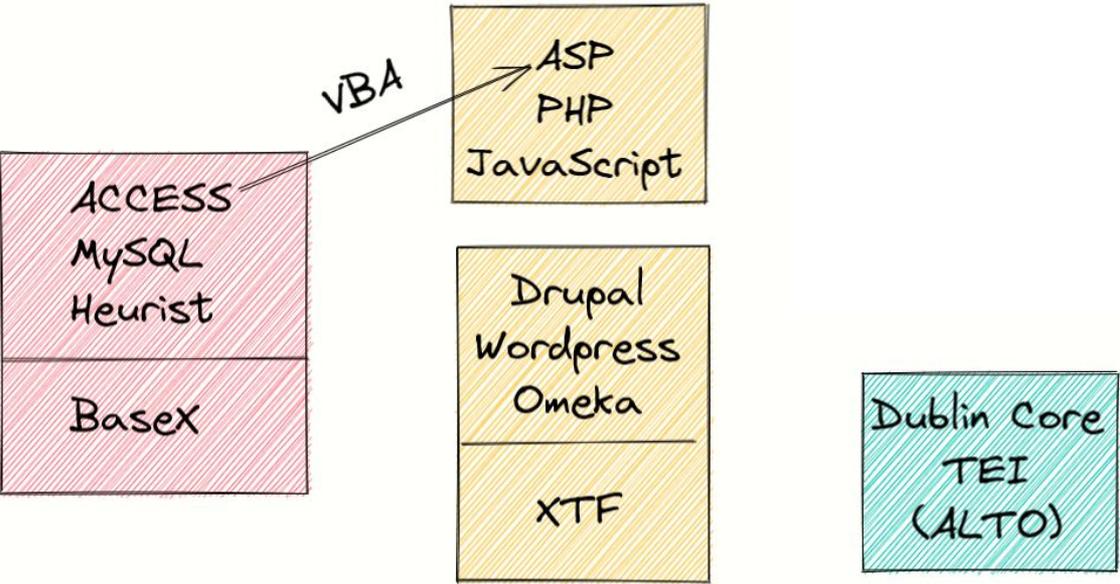
- Base de données relationnelles
- Dossier de Fichiers
- Fichiers XML
- Opération
- Autres outils





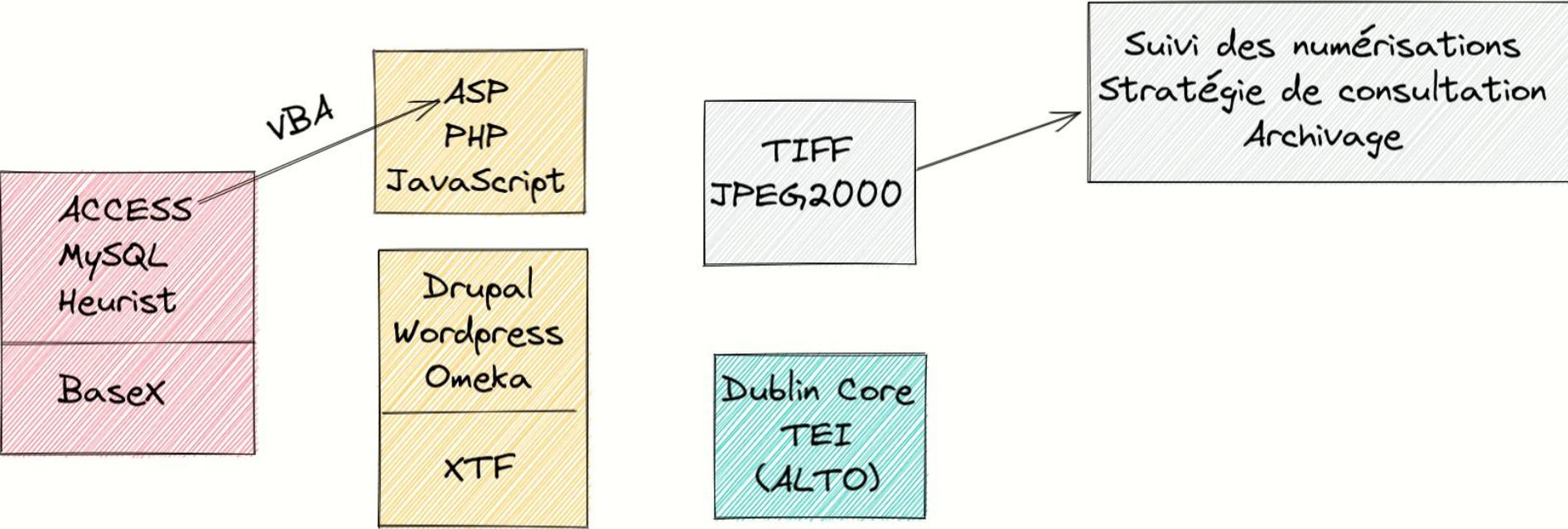
Bases de données relationnelles





Bases de données relationnelles vs. Bases de données XML







État des lieux

Une architecture qui **fonctionne** !

15 067 notices d'ouvrages

1 294 notices en ligne

163 éditions XML-TEI

Bases de données thématiques





Bibliothèques

Virtuelles

Humanistes

État des lieux

15 067 notices d'ouvrages

12 94 notices en ligne

163 éditions XML-TEI

Bases de données thématiques

Une architecture qui **fonctionne...**

... mais qui a atteint ses limites et doit être repensée et optimisée

- en s'appuyant sur l'expérience acquise depuis 20 ans
- en profitant des avancées technologiques





Bibliothèques

Virtuelles

Humanistes

Les “Grands chantiers” de la refonte

Refonte scientifique

Définir les unités documentaires opératoires

Construire les référentiels internes

Reconnecter les bases de données BVH

Interopérabilité des données



Bibliothèques

Virtuelles

Humanistes

Les “Grands chantiers” de la refonte

Refonte scientifique

Définir les unités documentaires opératoires

Construire les référentiels internes

Reconnecter les bases de données BVH

Interopérabilité des données

Refonte technique

Numérisation

Mise en ligne

Stockage /
Archivage

Traitements
des données

Interopérabilité

Serveurs



Bibliothèques

Virtuelles

Humanistes

Les “Grands chantiers” de la refonte

Refonte scientifique

Définir les unités documentaires opératoires

Construire les référentiels internes

Reconnecter les bases de données BVH

Interopérabilité des données

Refonte technique

Numérisation

Mise en ligne

Stockage /
Archivage

Traitements
des données

Interopérabilité

Serveurs

vers une chaîne de production optimisée





Bibliothèques

Virtuelles

Humanistes

Les "Grands chantiers" de la refonte

Refonte scientifique

- Définir les unités documentaires opératoires
- Construire les référentiels internes
- Reconnecter les bases de données BVH
- Interopérabilité des données

Refonte technique



vers une chaîne de production optimisée

La réflexion scientifique **ne doit pas** être guidée/entravée par les choix techniques

Les choix techniques et scientifiques doivent nous permettre de rester connectés à nos partenaires

Les choix techniques doivent s'inscrire dans la durée = choisir des technologies que l'on est en capacités de maintenir



BVH



Bibliothèques

Virtuelles

Humanistes



Vers un nouveau modèle de données

| Méthode

| Définition de concepts opératoires

| Quel modèle et quelles technos pour nos données ?



BHH





Bibliothèques

Virtuelles

Humanistes

Préambule : notre démarche

- ★ Travail en profondeur sur les données que l'on souhaite collecter, interroger & partager
 - Point de vue du chercheur qui décrit
 - Point de vue des utilisateurs de la bibliothèque virtuelle



Du côté des usagers, les objectifs de la refonte

1 TROUVER

Rassembler des informations sur une ressource pertinente en cherchant par tous les critères adéquats.

2 IDENTIFIER

Comprendre clairement la nature des objets trouvés et faire la distinction entre des ressources similaires.

3 SÉLECTIONNER

Déterminer si les ressources trouvées conviennent.

4 OBTENIR

Accéder au contenu de la ressource.

5 EXPLORER

Découvrir des ressources par le biais de relations qui existent entre elles.

Identification des 5 “tâches utilisateur” dans une bibliothèque numérique



Bibliothèques

Virtuelles

Humanistes

Préambule : notre démarche

- ★ Travail en profondeur sur les données que l'on souhaite collecter, interroger & partager

- ★ Redéfinition des concepts clés :
 - Œuvre, Édition
 - Ouvrage, recueil, collection
 - etc.





Bibliothèques

Virtuelles

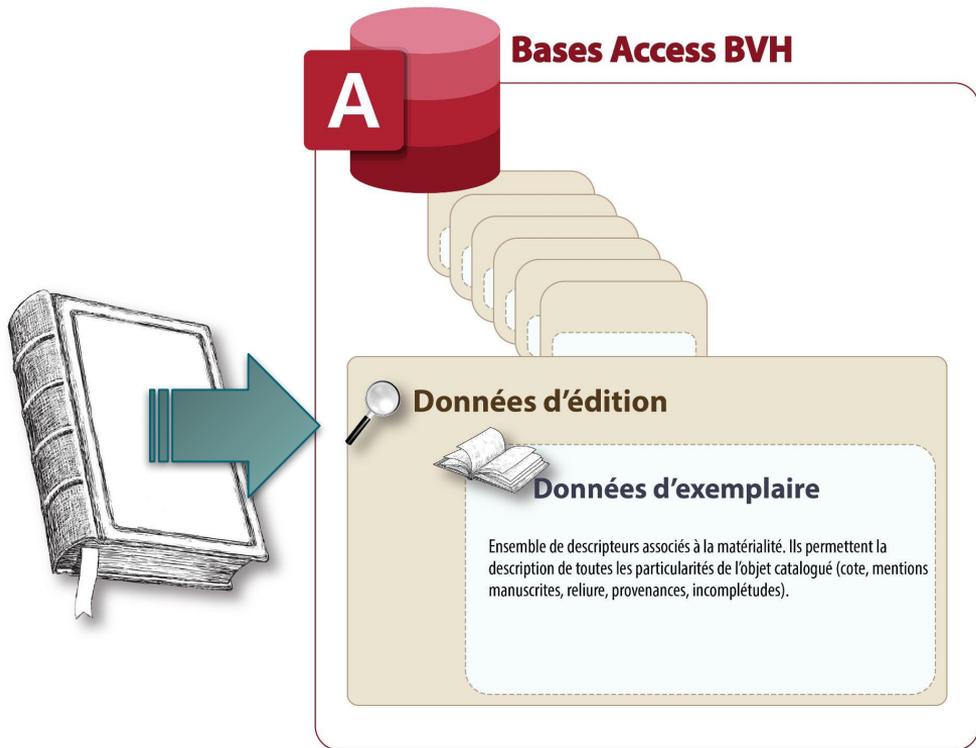
Humanistes

Préambule : notre démarche

- ★ Travail en profondeur sur les données que l'on souhaite collecter, interroger & partager
- ★ Redéfinition des concepts clés
- ★ Définir les unités documentaires opératoires
 - Approche itérative
 - Approche par l'exemple



Du schéma actuel de catalogage dans Access



Limites :

- 👎 Gestion des recueils factices
- 👎 Redondances d'information
- 👎 Normalisation inégale des données
- 👎 Interopérabilité limitée
- 👎 Manque de granularité
- 👎 etc.



Vers le schéma de données envisagé



Œuvre 1

Concept abstrait et opératoire essentiel à l'interopérabilité.
Il sert de point d'accès à celui qui cherche. C'est un élément essentiel à l'interopérabilité qui s'appuie sur des référentiels externes.

Objet physique

Entité concrète reposant sur des descripteurs associés à la matérialité. Elle permet la description de toutes les particularités de l'objet catalogué (cote, mentions manuscrites, reliure, provenances, incomplétude, etc.).

Manifestation 1

Données normées nécessaires à l'identification de l'édition (titre, adresse, date, responsabilités principales et secondaires).

Unité bibliographique 1

Découpage basé sur le contenu de l'objet décrit. Une unité bibliographique se caractérise par la présence d'une page de titre, d'une pagination ou foliotation continue.

Unité bibliographique 2

Manifestation 2

Unité bibliographique 3



Œuvre 2

Publique
Interne

Note(s) sur l'objet

Crédits techniques

Crédits

Droits

Identifiant unique

Type MIME

Date de mise en ligne

Profil colorimétrique

Financier(s)

B
V
H

Objet numérique

B
V
H





Bibliothèques

Virtuelles

Humanistes

Quel modèle pour nos données ?

- ★ Pratique historique : bases de données relationnelles
- ★ Réalité du corpus : données textuelles et documents XML





Bibliothèques

Virtuelles

Humanistes

Quel modèle pour nos données ?

- ★ Pratique historique : bases de données relationnelles
- ★ Réalité du corpus : données textuelles et documents XML
- ★ Volonté d'interopérabilité avec les bibliothèques partenaires et la Transition bibliographique





Bibliothèques

Virtuelles

Humanistes

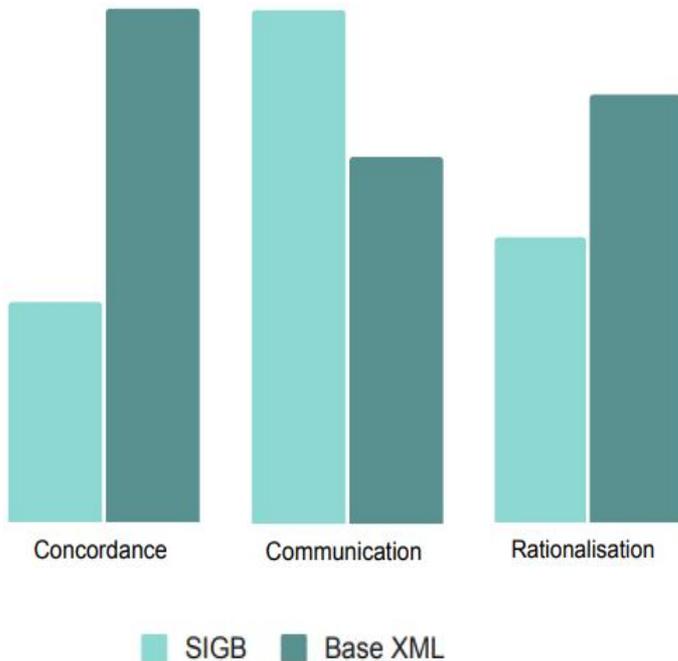
Quel modèle pour nos données ?

- ★ Pratique historique : bases de données relationnelles
- ★ Réalité du corpus : données textuelles et documents XML
- ★ Volonté d'interopérabilité avec les bibliothèques partenaires et la Transition bibliographique
- ★ Réflexion en cours qui s'appuie sur un travail de veille





De l'importance du travail de veille



Notre veille technique et fonctionnelle s'est focalisée sur deux possibilités :

- l'utilisation d'un **S**ystème **I**ntégré de **G**estion de **B**ibliothèque *open source* (PMB, Koha, etc.)
- le recours à une base XML native (BaseX, eXist-db) ou relationnelle compatible





Bibliothèques

Virtuelles

Humanistes

De l'importance du travail de veille

- ★ SIGB peu adaptés à nos besoins et à un modèle de données complexe





Bibliothèques

Virtuelles

Humanistes

De l'importance du travail de veille

- ★ SIGB peu adaptés à nos besoins et à un modèle de données complexe
- ★ BDD Relationnelles compatibles peu performantes





Bibliothèques

Virtuelles

Humanistes

De l'importance du travail de veille

- ★ SIGB peu adaptés à nos besoins et à un modèle de données complexe
- ★ BDD Relationnelles compatibles peu performantes
- ★ Vers un système unique (BBDR ou XML) ou un système hybride ?



Bibliothèques

Virtuelles

Humanistes

Importance des formats XML

Rapidement, un consensus s'est formé autour du format d'archivage des données produites : XML-TEI

Complémentaire des fichiers ALTO (coordonnées issues de la segmentation et de l'OCR)

```
<fileDesc>
  <titleStmt></titleStmt>
  <editionStmt></editionStmt>
  <extent></extent>
  <publicationStmt></publicationStmt>
  <notesStmt></notesStmt>
  <sourceDesc>
    <msDesc>
      <msIdentifier><!--Identifiant de l'objet physique --></msIdentifier>
      <msContents>
        <msItem n="ED1">
          <!--Lien : ŒUVRE/MANIFESTATION -->
          <!-- MANIFESTATION : description -->
          <msItem n="ED1UB1" xml:id="BVH_Cat_0000">
            <!-- UNITE BIBLIOGRAPHIQUE : description -->
          </msItem>
        </msItem>
      </msContents>
      <physDesc><!-- OBJET PHYSIQUE : description --></physDesc>
    </msDesc>
  </sourceDesc>
</fileDesc>
```

OBJET NUMERIQUE



Importance des formats XML

Rapidement, un consensus s'est formé autour du format d'archivage des données produites : XML-TEI

Complémentaire des fichiers ALTO (coordonnées issues de la segmentation et de l'OCR)

Fichier METS en pivot ?

```
<fileDesc>
  <titleStmt></titleStmt>
  <editionStmt></editionStmt>
  <extent></extent>
  <publicationStmt></publicationStmt>
  <notesStmt></notesStmt>
  <sourceDesc>
    <msDesc>
      <msIdentifier><!--Identifiant de l'objet physique --></msIdentifier>
      <msContents>
        <msItem n="ED1">
          <!--Lien : ŒUVRE/MANIFESTATION -->
          <!-- MANIFESTATION : description -->
          <msItem n="ED1UB1" xml:id="BVH_Cat_0000">
            <!-- UNITE BIBLIOGRAPHIQUE : description -->
          </msItem>
        </msItem>
      </msContents>
      <physDesc><!-- OBJET PHYSIQUE : description --></physDesc>
    </msDesc>
  </sourceDesc>
</fileDesc>
```

OBJET NUMERIQUE



Bibliothèques

Virtuelles

Humanistes



Vers une nouvelle chaîne de production

| Déploiement du protocole IIF



BHH



Vers l'interopérabilité des images

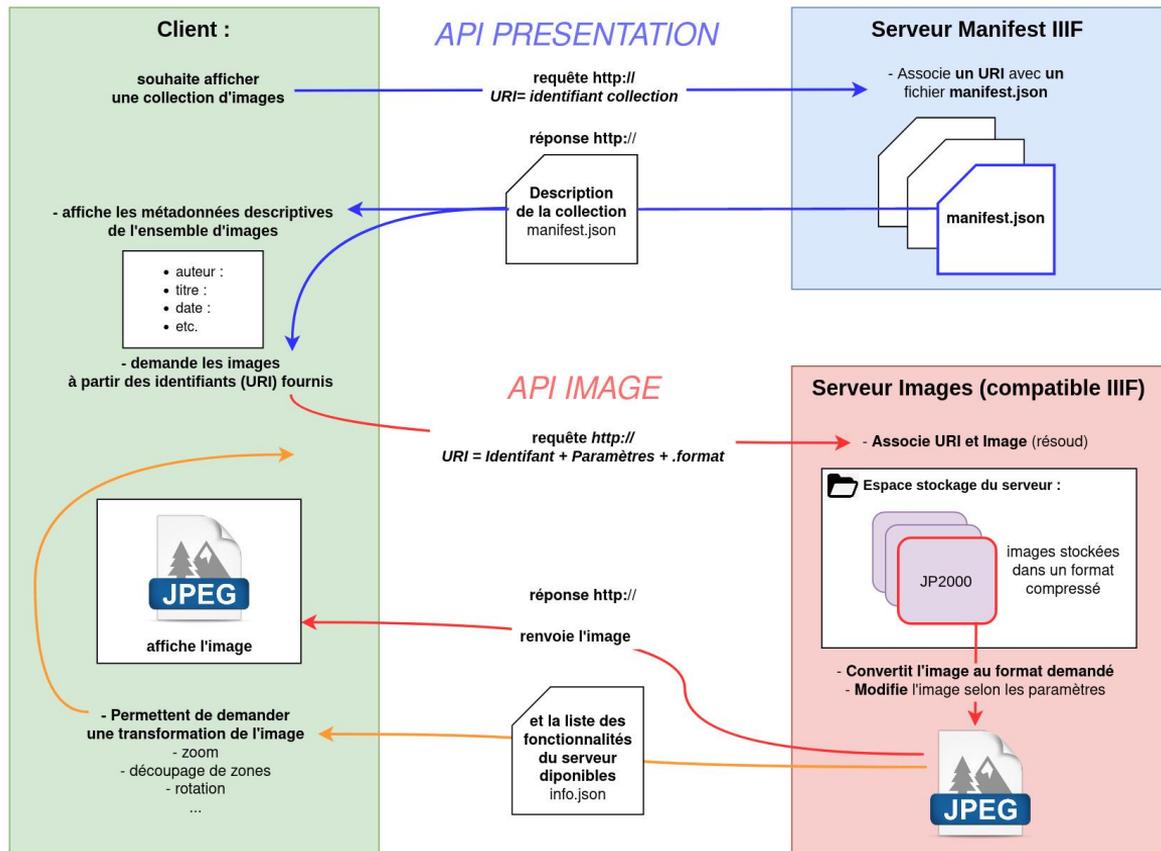


- **Interopérabilité des données** produites avec le Portail Biblissima et les autres bibliothèques partenaires des Bibliothèques Virtuelles Humanistes

- **Le protocole IIF** : repose sur des normes et l'établissement d'un système complexe



Le protocole IIIF



Fonctionnement schématique des APIs IIIF





Bibliothèques

Virtuelles

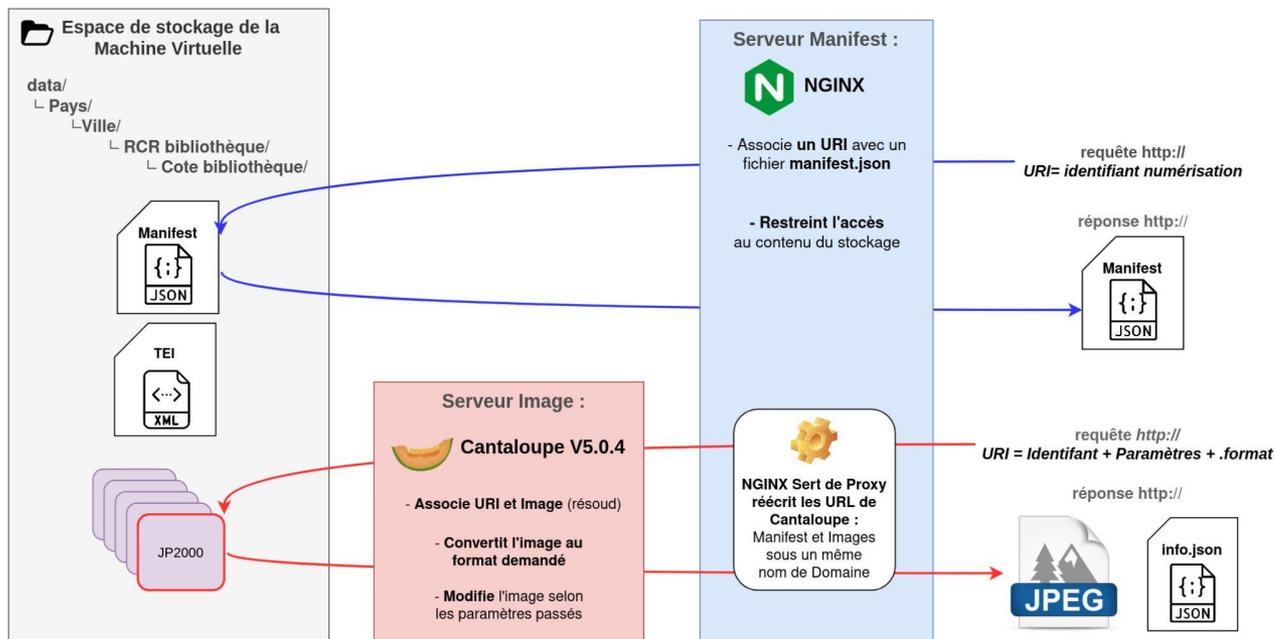
Humanistes

Le déploiement de IIF : Réflexion sur la chaîne de production

- Données et métadonnées à récupérer
- Format des données :
 - Le format des fichiers
 - Le format des images
- Outils et logiciels utilisés
- Stockage et l'organisation des données au sein de l'infrastructure



Le déploiement de IIF : Mise en place



Serveur Manifest :

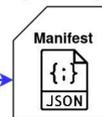


- Associe un **URI** avec un fichier **manifest.json**

- **Restreint l'accès** au contenu du stockage

requête `http://`
URI = identifiant numérisation

réponse `http://`



Serveur Image :



Cantaloupe V5.0.4

- Associe **URI** et **Image** (résoud)

- Convertit l'**image** au format demandé

- Modifie l'**image** selon les paramètres passés

NGINX sert de Proxy réécrit les URL de **Cantaloupe** : Manifest et Images sous un même nom de Domaine

requête `http://`
URI = Identifiant + Paramètres + .format

réponse `http://`



Architecture des serveurs IIF des BVH (Machine Virtuelle Huma-Num)



Le déploiement de IIF : Développements futurs

- Travail en cours
- Objectifs futurs :
 - Personnalisation du Serveur Image via un script délégué
 - Récupération de métadonnées encapsulées dans les images
 - Limitation de l'accès à certaines images, en fonction des droits
 - Paramétrage d'un visualiseur





Bibliothèques

Virtuelles

Humanistes



Conclusion

| Perspectives

| Calendrier



BH



Bibliothèques

Virtuelles

Humanistes

Conclusion

- ★ Bilan qui nous a conduit à repenser nos objectifs :
 - créer un outil unique de visualisation des données
 - stabiliser et pérenniser les données de la recherche



Conclusion

- ★ Bilan qui nous a conduit à repenser nos objectifs :
 - créer un outil unique de visualisation des données
 - stabiliser et pérenniser les données de la recherche

- ★ Travail qui est un challenge
 - Remet en question les grilles de description traditionnelles
 - Sollicite des compétences multiples
 - Avancer dans le présent en pensant au futur



Conclusion

- ★ Bilan qui nous a conduit à repenser nos objectifs :
 - créer un outil unique de visualisation des données
 - stabiliser et pérenniser les données de la recherche

- ★ Travail qui est un challenge
 - Remet en question les grilles de description traditionnelles
 - Sollicite des compétences multiples
 - Avancer dans le présent en pensant au futur

- ★ Créer un outil heuristique adapté à la nature des sources
 - Offrir une nouvelle vision historique de la Renaissance





Bibliothèques

Virtuelles

Humanistes

Conclusion

Projections : calendrier ?



Mise en place des ateliers de réflexion scientifique

Lancement des “chantiers” IIF, Archivage TEI

Veille sur les outils



Mise en place de l'infrastructure serveur

Déploiement IIF et reversement Biblissima

Validation du modèle de données

Conceptualisation de la chaîne de production et choix technos



Validation de la chaîne de production et mise en place des outils

Nettoyage, enrichissement, transfert des données fac-similés



Intégration du Corpus Epistemon et des autres BDDs



BHH



Bibliothèques

Virtuelles

Humanistes



| Questions ?



BH

Bibliothèques

Virtuelles

Humanistes



Centre d'études supérieures
de la *Renaissance*



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Partage dans les Mêmes Conditions 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/).