Cahier des clauses techniques particulières (CCTP)

— CCTP LOT 01 —

SOCLE APPLICATIF DU PORTAIL BVH ET SERVICES D'EXPOSITION DES DONNEES (API REST, IIIF, OAI, ARK)

INFRASTRUCTURE NUMÉRIQUE DU PORTAIL « BIBLIOTHÈQUES VIRTUELLES HUMANISTES » (BVH)

EQUIPEX+ BIBLISSIMA+

« Fairisation des données BVH vers les portails et outils Biblissima et partenaires »

Biblissima+ bénéficie d'une aide de l'Etat gérée par l'ANR au titre du Programme d'investissements d'avenir intégré à France 2030, portant la référence ANR-21-ESRE-0005.

1. OBJET DU LOT 01 3
2. CONTEXTE TECHNIQUE ET FONCTIONNEL 4
2.1. Architecture actuelle & architecture cible 5
2.2. Donnees existantes 5
3. OBJECTIFS FONCTIONNELS DU LOT 01 7
3.1. Centraliser et exposer les metadonnees BVH 7
3.2. Faire evoluer le back-end Django existant (et autoriser l'enrichissement des bases et du modele de donnees BVH) 7
3.3. Developper une API REST documentee 7
3.4. Mettre en place un stockage objet compatible S3 8
3.5. Publier les ressources images selon le protocole IIIF 8
3.6. Assurer la continuite des URLs publiques 8
3.7. [PSE obligatoire] Attribuer et resoudre des identifiants ARK 9
3.8. [PSE obligatoire] Mettre en place un moteur d'indexation 9
3.9. [PSE obligatoire] Publier les metadonnees dans l'entrepot OAI-PMH des BVH 9
3.10. Garantir l'interoperabilite et la reutilisabilite des donnees BVH 9
4. PERIMETRE DES PRESTATIONS ATTENDUES 9
4.1. Audit technique initial 9
4.2. Évolution du back-End Django 10
4.3. Base PostgreSQL et modele de donnees 11
4.4. API REST d'exposition des donnees 11
4.5. Stockage objet compatible S3 (MinIO ou solution equivalente) 13
4.6. Serveur IIIF et Manifests 14
4.6.1. Migration et optimisation du serveur IIIF 14
4.6.2. Mise à jour des Manifests IIIF & continuité de service 15
4.6.3. Caractéristiques du serveur IIIF actuel 16
4.7. [PSE obligatoire] Moteur d'indexation Elasticsearch / OpenSearch 16
4.8. Identifiants perennes et gestion des URLs (publiques et redirections) 18
4.8.1. Solution de correspondance et de redirection des URLs 18
4.8.2. [PSE obligatoire] Solution ARK complet 18
4.9. [PSE obligatoire] Entrepot OAI-PMH et exposition Dublin Core 21
5. EXIGENCES TECHNIQUES TRANSVERSALES AU LOT 1 23
6. LIVRABLES ATTENDUS DU LOT 01 24
7. MODALITES DE DEPLOIEMENT SPECIFIQUES AU LOT 01 27
8. TEST & RECETTES SPECIFIQUES AU LOT 01 27
9. SECURITE 28
10. DOCUMENTATION ET TRANSFERT DE COMPETENCES 29

1. OBJET DU LOT 01

Le présent lot a pour objet le développement, le déploiement, le paramétrage et la documentation du socle applicatif back-end de l'infrastructure numérique BVH.

Le lot 01 constitue le socle technique principal sur lequel s'appuieront les interfaces publiques, les fonctions de recherche, la visualisation IIIF, les exports, les flux de numérisation et les services d'exposition des données.

Ce socle permettra :

- l'évolution et l'enrichissement du back-end Django préexistant, liés à l'intégration de nouvelles données issues d'anciennes bases Access et nécessitant l'évolution du modèle de données
- l'exposition des données de la base PostgreSQL au moyen d'une API REST documentée
- le déploiement, le paramétrage ou l'intégration d'un stockage objet compatible S3, de type MinIO ou solution équivalente
- l'optimisation, le paramétrage et la migration du serveur IIIF
- l'exposition des manifests IIIF
- la mise en œuvre d'une solution minimale de correspondance et de redirection entre anciennes URLs BVH et nouvelles URLs publiques, si l'option ARK n'est pas retenue
- l'intégration de ces services dans l'environnement applicatif global BVH
- la documentation, les tests, la recette et le transfert de compétences

Le lot comprend également les options suivantes, qui seront chiffrées séparément :

- [PSE obligatoire] la mise en place d'un moteur d'indexation de type Elasticsearch, OpenSearch ou solution équivalente
- [PSE obligatoire] la génération, l'attribution, la gestion et la résolution d'identifiants pérennes ARK
- [PSE obligatoire] l'audit, l'adaptation, le remplacement ou le redéveloppement de l'entrepôt OAI-PMH existant, afin de le raccorder aux nouvelles sources de données BVH.

Type	Prestations
Offre de base	Back-end Django, PostgreSQL, API REST, stockage objet S3, IIIF, solution de redirection anciennes URLs / nouvelles URLs, documentation, tests
PSE obligatoires	Moteur d'indexation OpenSearch/Elasticsearch, solution ARK complet, solution entrepôt OAI-PMH

2. CONTEXTE TECHNIQUE ET FONCTIONNEL

Le portail BVH vise à publier, interroger, visualiser et pérenniser des corpus numériques patrimoniaux associant images, textes, métadonnées et identifiants persistants.

Le lot 1 s'intégrera dans l'architecture cible principalement sur les modules suivants :

- une base PostgreSQL « Catafonds » constituant la base principale des métadonnées
- un back-end Django
- un stockage objet S3, de type MinIO ou solution assimilée
- un serveur IIIF pour la diffusion des images et des manifests
- un visualiseur IIIF, de type Mirador, intégré dans le lot 2
- un moteur d'indexation de type Elasticsearch ou OpenSearch (si option retenue),
- un service ARK pour l'attribution et la résolution des identifiants pérennes (si option retenue) ou à défaut, un mécanisme minimal de correspondance et de redirection d'URLs
- une plateforme NumaHOP, objet du lot 3, pour la gestion des flux de numérisation
- GitLab pour le versionnement, le suivi des anomalies et la documentation (instance GitLab IR* Huma-Num)
- un entrepôt OAI-PMH, à redéployer ou redévelopper (si option retenue)

Le titulaire veillera à la cohérence d'ensemble entre les métadonnées descriptives, les fichiers numériques, les manifests IIIF, les index de recherche et les identifiants pérennes.

En outre, il garantira la bonne articulation des nouveaux services applicatifs déployés dans le cadre de cette prestation avec, s'ils sont impactés directement ou indirectement par la prestation, les services applicatifs préexistants déjà mis en production par la maitrise d'ouvrage. Le cas échéant, le titulaire en informera en amont la maîtrise d'ouvrage afin de définir une solution optimisée. La poursuite de service concerne potentiellement les services applicatifs suivants :

- un Wikibase data.bvh pour la gestion des référentiels d'autorités communs, en production
- une bibliothèque Zotero pour la gestion des références bibliographiques, en production
- une plateforme Arkindex dédiée au traitement automatisé d'images patrimoniales par l'IA, en production
- une plateforme de publication des textes pour les éditions en XML-TEI xtf.bvh (framework XTF, Berkley), en production
- un CMS Wordpress pour la partie éditorial de présentation des activités du programme BVH, en développement en interne au CESR
- des bases de données (Access, PostgreSQL, Heurist, XML) et sites projets connexes (CMS Drupal, Wordpress) en production, dont les données ont vocation à intégrer les services de publication et d'exposition du portail BVH
- des scripts de transformations ou de migration des données (Python, XSLT)

2.1. Architecture actuelle & architecture cible

Pour soutenir l'infrastructure, quatre machines virtuelles (VM) directement concernées par les prestations ont été déployées par la Direction des Systèmes d'Information (DSI) de l'université de Tours au sein du DataCentre Centre-Val de Loire (DC CVL). Elles sont respectivement dédiées :

- une à la nouvelle version du serveur IIIF (le serveur en production est actuellement hébergé sur une VM de l'IR* Huma-Num et doit donc être migré et optimisé sur la nouvelle VM dans le cadre du lot 01)
- une à l'outil NumaHOP (à mettre à jour et optimiser dans le cadre du lot 3)
- une à l'outil Wikibase data.bvh (non concerné directement par les prestations)
- une à la nouvelle infrastructure de la bibliothèque numérique (PostgreSQL et Django) qui accueillera les nouveaux développements et mises à jour du lot 01 et du lot 02 (API Rest, moteur d'indexation, système ARK ou de redirections d'URLs, OAI-PMH, visualiseur IIIF, etc.)

Uniquement accessibles aux personnels et prestataires autorisés pendant la phase de développement, ces VM sont ensuite mises à disposition de l'ensemble de la communauté lors de leur passage en production.

Toutes les VM fonctionnent sous Debian 12 ou 13, selon leur date de création. Leur volumétrie est ajustable en fonction des besoins qui seront identifiés au fil des prestations. Afin d'assurer une isolation optimale des environnements et une gestion simplifiée des ressources, de nouvelles VM pourront être provisionnées si jugées nécessaires.

L'infrastructure de recherche BVH ne dispose actuellement pas de stockage objet compatible S3 (un simple test a été réalisé) afin de stocker les ressources numériques nécessaires au portail BVH. Les données stockées, archivées et publiées sont respectivement hébergées sur un Filer hébergé à l'Université de Tours (DSI, DC CVL) et partiellement sur une Huma-Num Box (IR* huma-Num) pour les données partagées et exposées via IIIF.

Les schémas des interactions entre les outils, actuel et cible, seront fournis en annexe.

2.2. Données existantes

Les BVH disposent d'un ensemble de données, de bases de données et de ressources numériques constituées progressivement dans le cadre de leurs activités de description, de numérisation, d'édition et de publication de corpus patrimoniaux.

Ces données comprennent notamment :

- des images issues de la numérisation de documents patrimoniaux
- des textes, notamment des éditions ou transcriptions en XML-TEI
- des métadonnées descriptives, techniques, administratives ou éditoriales
- des fichiers associés aux ressources numériques, notamment PDF de consultation, exports ou fichiers dérivés
- des manifests IIIF et manifests de collections, actuellement générés à partir de plusieurs sources de données

- des données d'autorité centralisées dans un référentiel
- des identifiants internes, des anciennes URLs BVH (ancien portail BVH encore en ligne en date du 26/06/2026), des nouvelles URLs publiques (première version du nouveau portail en cours ; publication fin 2026)

Les principales sources de données sont notamment :

- **la base « Catafonds » sous PostgreSQL**, qui constitue la base principale des métadonnées descriptives BVH
- **le référentiel d'autorité data.bvh sous Wikibase**, utilisé notamment pour les référentiels d'agents, de lieux, et d'institutions de conservation, déjà interfacé avec l'environnement PostgreSQL / Django existant ;
- **la base « Fonds » sous Microsoft Access** (migration prévue dans Catafonds)
- **la base « Sommaires » sous Microsoft Access** (migration prévue, modalités à déterminer)
- **la base « Images » sous Microsoft Access** (migration partielle effective dans BaTyR ; modalités à déterminer pour les données restantes)
- **la base Epistemon**, actuellement publiées via le framework XTF pour l'indexation et la publication des éditions en XML-TEI
- **la base BaTyR sous PostgreSQL**
- les ressources exposées ou destinées à être exposées via l'**entrepôt OAI-PMH**.
- les métadonnées des ressources exposées ou destinées à être exposées via le **serveur IIIF** et les manifests associés.
- à venir dans le cadre de ce marché : les données et fichiers produits ou gérés dans le cadre des **flux de numérisation, notamment via NumaHOP** (lot 03 du présent marché).

Le titulaire prendra en compte cet existant, en assurera la compatibilité avec les développements réalisés dans le cadre du présent marché, et veillera à la non-régression des fonctionnalités d'interfaçage existantes.

La présentation de ces bases de données est fournie en annexe.

La base « Catafonds » repose sur un modèle structuré en 6 niveaux (œuvres, expressions, manifestations, unités éditoriales, objets physiques et objets numériques). Les détails du modèle et des données sont fournis en annexe.

Les opérations de nettoyage, dédoublonnage et arbitrage scientifique des données relèvent du maître d'ouvrage, ainsi que leur intégration ou migration dans « Catafonds » ou dans les bases tierces. À partir des données préparées fournies par l'équipe BVH, le titulaire assurera le développement, l'adaptation et la mise en œuvre des scripts, procédures ou connecteurs nécessaires, ainsi que les vérifications techniques permettant, le cas échéant, leur intégration au back-end Django.

La production de métadonnées et l'alimentation par le maître d'ouvrage de la base « Catafonds » (modification, ajout, suppressions de données) se poursuivront pendant la prestation. À noter que lors de la première phase de mise en production de la base, certains éléments tels que la vie de l'objet physique ou les descriptions des illustrations, ont été simplifiés. Dans une seconde phase,

ces éléments seront approfondis et formalisés, ce qui entraînera une adaptation du modèle pour les intégrer de manière optimale. Si ces modifications structurelles ou fonctionnelles de la base et de ses services dépendants intervenaient ou devaient être mise en œuvre durant la prestation, les membres de l'équipe BVH en informeront le prestataire, et inversement le prestataire devra impérativement solliciter la maitrise d'ouvrage au préalable.

Aucune donnée sensible ni donnée personnelle n'est traitée ou stockée dans ces bases, à l'exception des comptes d'accès des utilisateurs (identifiant, mot de passe, adresse email, nom et prénom) requis pour l'accès aux différents services.

3. OBJECTIFS FONCTIONNELS DU LOT 01

Le lot 01 a pour objectif de constituer le **socle applicatif back-end du futur portail BVH**, en assurant l'articulation entre les métadonnées, les ressources numériques, les identifiants, les services d'exposition, et les outils de diffusion publique.

Le lot 01 répondra aux objectifs fonctionnels suivants :

3.1. Centraliser et exposer les métadonnées BVH

Les données descriptives stockées dans la base PostgreSQL « Catafonds » seront consultées, filtrées, recherchées *via* des interfaces publiques personnalisées grâce au framework Django (lot 02) et exposées via une API REST documentée (lot 01).

Le prestataire prendra en compte le modèle de données BVH, ses niveaux de description, ses relations internes, ainsi que les besoins de publication, de consultation publique ou restreinte, et d'exploitation par les autres composants de l'infrastructure.

3.2. Faire évoluer le back-end Django existant (et autoriser l'enrichissement des bases et du modèle de données BVH)

Durant la prestation, le titulaire prendra en compte le besoin du maître d'ouvrage de reprendre ou d'intégrer des nouvelles données à « Catafonds », dont certaines sont actuellement conservées dans d'anciennes bases Microsoft Access, notamment les bases « Fonds », « Sommaires » et « Images » ou sont en cours de collation.

Le titulaire fera évoluer le back-end Django existant afin de permettre l'administration, la consultation, l'exposition et la synchronisation des données utilisées par le portail BVH. Le back-end sera maintenable, documenté, versionné et compatible avec les autres composants de l'infrastructure.

3.3. Développer une API REST documentée

Le titulaire déploiera et paramétrera une API REST permettant l'exposition des données issues de PostgreSQL et des ressources associées.

Cette API publique (accès en lecture seule) permettra notamment la consultation des métadonnées et des notices, l'accès aux ressources numériques, l'exposition des liens vers images, manifests IIIF, les éditions XML-TEI, les identifiants et URLs publiques.

L'API sera documentée pour permettre des requêtes (filtres, de tris, etc.).

3.4. Mettre en place un stockage objet compatible S3

Le titulaire assurera le déploiement, le paramétrage ou l'intégration d'un stockage objet compatible S3, de type MinIO ou solution équivalente.

Ce stockage devra accueillir les ressources numériques nécessaires en cours de production (flux de numérisation par exemple, dans le cadre du lot 03), pour la publication sur le portail BVH, et pour l'exposition aux partenaires, incluant notamment les images, fichiers sources, fichiers XML-TEI, fichiers METS/ALTO, manifests IIIF, exports dérivés et fichiers de consultation.

L'organisation du stockage permettra une gestion contrôlée des accès, une structuration par usages ou collections, et une articulation avec Django, IIIF, NumaHOP et les autres services concernés.

3.5. Publier les ressources images selon le protocole IIIF

Les images stockées dans le stockage objet seront servies *via* un serveur IIIF et décrites dans des manifests conformes aux standards attendus.

Le titulaire assurera la migration, l'optimisation et le paramétrage du serveur IIIF, la continuité d'exposition des images et manifests existants, sans rupture de service, ainsi que la compatibilité avec les visualiseurs et outils retenus, notamment Mirador et NumaHOP.

Le maitre d'ouvrage établira les règles de *mapping* entre les données de la base et les champs des manifests et manifests de collection, ainsi que les éventuels *templates* ou exemples de référence pour les tests. Il assurera également la mise à jour des scripts Python permettant la génération automatisée des nouveaux manifests IIIF, conformément aux nouvelles spécifications établies. Ces scripts devront s'appuyer sur la nouvelle architecture applicative et sur la base de données. Le titulaire fournira au maître d'ouvrage toutes les prescriptions fonctionnelles et techniques requises pour cette mise à jour, afin qu'il puisse générer les nouveaux manifests, les valider et assurer la mise à jour de sa documentation interne (manifests, scripts, procédures).

Le titulaire assurera l'accès des scripts aux données et les paramétrages nécessaires pour l'exécution des scripts. Il assurera les tests et signalera toute anomalie relevant des manifests ou des scripts au maitre d'ouvrage qui rendrait impossible, incomplète ou fautive l'exposition sur le serveur IIIF et via ses API.

3.6. Assurer la continuité des URLs publiques

Le titulaire garantira la continuité d'accès aux ressources BVH lors de la migration vers le nouveau portail. Dans l'hypothèse où l'option ARK ne serait pas retenue, le prestataire mettra en œuvre une solution minimale de correspondance et de redirection entre les anciennes URLs BVH, les nouvelles URLs publiques, les identifiants internes des ressources et, le cas échéant, les identifiants OAI-PMH.

3.7. [PSE obligatoire] Attribuer et résoudre des identifiants ARK

Si cette option est retenue, les ressources BVH validées par la maîtrise d'ouvrage recevront des identifiants ARK pérennes, stockés, exposés, résolus et maintenus de manière stable.

Ces identifiants seront intégrés à l'API REST, aux manifests IIIF, aux éventuels index de recherche et aux autres services d'exposition retenus.

3.8. [PSE obligatoire] Mettre en place un moteur d'indexation

Si cette option est retenue, les métadonnées et contenus exploitables seront indexés dans un moteur de type Elasticsearch, OpenSearch ou solution équivalente, afin d'améliorer les performances de recherche, de produire des facettes et tris, et d'alimenter les interfaces du lot 02.

3.9. [PSE obligatoire] Publier les métadonnées dans l'entrepôt OAI-PMH des BVH

La prestation relative à l'entrepôt OAI-PMH constitue une option du présent lot.

Si cette option est retenue, les métadonnées issues de PostgreSQL et les URLs publiques des ressources numériques associées seront exposées *via* un entrepôt OAI-PMH conforme aux standards attendus, notamment Dublin Core simple.

Cette option inclus l'audit de l'entrepôt existant, son adaptation ou son remplacement soit par une solution existante open-source, soit par le développement d'un nouveau service. La solution retenue sera raccordée aux sources de données de l'architecture cible.

3.10. Garantir l'interopérabilité et la réutilisabilité des données BVH

Dans la limite des prestations retenues, le titulaire contribuera par les solutions et développements mis en œuvre, à l'interopérabilité et à la FAIRisation des données BVH, c'est-à-dire à leur capacité à être trouvées, consultées, reliées, échangées et réutilisées par les services du portail BVH, par ses partenaires scientifiques ou par des services tiers.

Cette exigence ne constitue pas une prestation autonome distincte, mais un principe transversal applicable à l'ensemble des développements du socle applicatif du lot 01, notamment à l'API REST, au stockage S3, au serveur IIIF, aux mécanismes d'identifiants ou de redirection, aux exports et aux services d'exposition retenus.

Cette interopérabilité s'appuiera et sera vérifiable au travers du respect des standards, des formats et de la structuration des données définis ou validés par le maitre d'ouvrage, ainsi qu'au travers de la qualité des identifiants, des exports, des dépendances techniques et de la documentation produits par le titulaire.

4. PÉRIMÈTRE DES PRESTATIONS ATTENDUES

4.1. Audit technique initial

Le titulaire réalisera un audit court de l'existant technique afin d'identifier :

- l'état du back-end Django
- l'état de la base PostgreSQL
- l'état du serveur IIIF
- les dépendances applicatives
- les services déjà installés (WikiBase, Zotero…)
- les données disponibles
- les besoins de migration ou de réorganisation du code et de l'architecture applicative (refactorisation)
- les risques techniques
- les prérequis de déploiement (sécurité, performance, etc.)

Le titulaire disposera d'un accès aux dépôts Git, dump des bases de données (PostgreSQL, Access), serveur de test et à toute documentation existante et utile à cet audit initial.

L'audit donnera lieu à une note de cadrage technique et un planning prévisionnel validés en réunion par l'équipe BVH avant le démarrage des développements. Une méthodologie agile sera privilégiée.

Le candidat indiquera dans son offre une durée indicative de l'audit en jours/homme avec justification.

4.2. Évolution du back-End Django

Le titulaire fera évoluer ou développera les modules Django nécessaires au fonctionnement du portail BVH. Le déploiement s'effectuera systématiquement avec la dernière version stable connue du framework Django à la date de mise en production (actuellement 6.1) et le titulaire s'engage à assurer la compatibilité des développements avec les mises à jour ultérieures du framework, en suivant le rythme des versions maintenues par la communauté Django.

Sur la base des résultats de l'audit initial ainsi que des solutions retenues pour l'ensemble des prestations du lot 01 et des dépendances aux lots 02 et 03, les prestations pourront comprendre :

- la définition de nouveaux modèles ou l'évolution de modèles existants
- l'évolution de Django Admin (personnalisation des interfaces de saisie, créations de nouveaux filtres, amélioration de l'interactivité, etc.)
- l'évolution de la gestion des droits et des profils
- le développement de services métiers c'est-à-dire le développement de fonctionnalités applicatives répondant aux besoins propres des BVH (par exemple : génération de manifests IIIF, attribution d'ARK, synchronisation avec les flux de numérisation, exports de métadonnées, automatisation de traitements sur les notices ou les fichiers)
- l'évolution de la gestion des relations entre notices, documents, images, fichiers, exports, ARK et manifests
- la mise à jour ou la rédaction de la documentation des modèles et services impactés. L'intégralité du code créé devra être parfaitement commentée au format compatible avec

Sphinx, afin de permettre une génération automatique et exhaustive de la documentation technique.

La documentation des évolutions et développements sera rédigée en Markdown et le code versionné sur l'instance GitLab Huma-Num dédié de l'équipe BVH (https://gitlab.huma-num.fr/team-bvh - Git privé).

Le titulaire précisera en concertation avec l'équipe BVH la stratégie de versionnement des développements, les modalités d'évolution des versions, la gestion de la rétrocompatibilité, etc.

4.3. Base PostgreSQL et modèle de données

PostgreSQL constitue la base principale des métadonnées BVH. Le titulaire assurera la bonne articulation entre la base, les modèles Django, l'API REST, les index et les services de publication du nouveau portail BVH (bibliothèque numérique).

Le déploiement s'effectuera systématiquement avec la dernière version stable connue à la date de mise en production (actuellement 18.4) et le prestataire s'engage à assurer la compatibilité avec les versions ultérieures du SGBD, en suivant le rythme des versions maintenues par la communauté Postgres.

Sur la base des résultats de l'audit initial ainsi que des solutions retenues pour l'ensemble des prestations du lot 01 et des dépendances aux lots 02 et 03, les prestations pourront comprendre :

- des propositions d'ajustements du schéma relationnel existant
- des scripts de migration
- l'optimisation des requêtes
- l'indexation PostgreSQL lorsque nécessaire

L'équipe BVH fournira au titulaire :

- les correspondances entre tables, entités fonctionnelles et API, en fonction des besoins
- la documentation du modèle

La base de données, composée exclusivement de données textuelles, ne devrait pas excéder les 50 Mo.

4.4. API REST d'exposition des données

Le titulaire développera une API REST permettant l'exposition et le partage des données issues de PostgreSQL.

Le choix de la technologie sous-jacente à l'API (ex. : API native PostgreSQL, Django REST Framework, ou autre solution) fera l'objet d'une concertation préalable avec la maîtrise d'ouvrage,

afin de valider son adéquation avec les besoins fonctionnels, les contraintes techniques et les enjeux de maintenabilité du projet. Cette discussion permettra notamment d'arbitrer entre les approches disponibles (ex. : exposition directe des données via PostgreSQL, développement d'une couche applicative dédiée, etc.) et de définir les modalités de déploiement, de paramétrage, d'authentification et d'accessibilité (lecture/écriture) associées.

L'API permettra, selon les entités concernées :

- la consultation des données de la base retenues pour l'exposition publique par la maîtrise d'ouvrage (les champs à exclure seront précisés le cas échéant).
- la consultation des ressources numériques
- la consultation des métadonnées descriptives
- la consultation des liens vers les images, manifests, fichiers et identifiants
- la mise à disposition de filtres, tris et mécanismes de paginations adaptés aux besoins de consultation et d'exploitation des données, notamment types de ressources, corpus, œuvre, auteur, titre, date, lieu, langue, identifiant, état de publication ou présence de fichiers associés (liste non exhaustive)
- la production de réponses structurées en JSON et/ou JSON-LD
- la documentation automatique ou semi-automatique de l'API

L'API sera conçue pour être consommée par les interfaces du lot 2, par le moteur d'indexation et, le cas échéant, l'API de NumaHop (lot 3) ainsi que les partenaires ou services d'export.

La liste des endpoints sera établie en concertation avec l'équipe BVH : ressources à exposer et opérations attendues. Les endpoints serviront une API publique librement accessible (accès en lecture seule), et éventuellement une API restreinte nécessitant une authentification (accès en lecture seule).

Le titulaire devra distinguer dans son offre les endpoints proposés en tranche ferme et ceux qui relèveraient d'options ou de développements ultérieurs.

La documentation de l'API sera rédigée en Markdown et le code versionné sur l'instance GitLab Huma-Num dédié de l'équipe BVH.

Le titulaire précisera la stratégie de versionnement de l'API (préfixe d'URL de type /api/v1/), les modalités d'évolution des versions, la gestion de la rétrocompatibilité et les règles de dépréciation des endpoints).

Dans l'hypothèse où la mise en œuvre des identifiants ARK ne serait pas retenue ou serait reportée, l'API REST devra pouvoir exposer les informations nécessaires au mécanisme minimal de correspondance et de redirection prévu à la section 4.8, notamment les identifiants internes des ressources, les anciennes URLs BVH le cas échéant, les nouvelles URLs publiques, les types de ressources et les statuts de publication.

4.5. Stockage objet compatible S3 (MinIO ou solution équivalente)

Le titulaire installera, déploiera et paramétrera un stockage objet compatible S3 afin de stocker les ressources numériques nécessaires au portail BVH qui ne dispose pas actuellement d'installation préalable. Un test de déploiement de MinIO a néanmoins été mené en 2025. MinIO n'est pas imposé si le titulaire démontre et justifie le choix d'une autre solution compatible S3 (avantages et contraintes : coût, durabilité, évolution, maintenance, etc.).

La liste initiale des volumes (buckets) à créer ainsi que le plan de nommage à respecter seront fournis par l'équipe BVH qui devra être autonome à l'issue de la prestation pour en créer, modifier ceux existants, ou en supprimer à sa convenance. Les buckets pourront être publics ou privés.

Ces volumes devront être organisés de manière à permettre leur utilisation à différentes étapes de la chaîne de production, en lecture, écriture et modification, par les différents services et applications du système, en tenant compte des contraintes propres à chaque usage (stockage intermédiaire, production, exposition des données, etc.).

Le stockage permettra un accès contrôlé et différencié selon les services, notamment (non exhaustif) :

- le serveur IIIF devra pouvoir accéder à plusieurs collections d'images stockées dans différents volumes afin d'assurer la diffusion des images
- la plateforme NumaHOP devra pouvoir accéder en lecture aux images sources à traiter, et en écriture pour déposer les fichiers après traitement directement dans le volume IIIF de la collection concernée, en vue de leur publication et de leur exposition
- les services applicatifs (Django, API, indexation) devront pouvoir accéder aux ressources nécessaires selon leurs besoins fonctionnels

L'architecture de stockage sera conçue de manière à limiter au maximum la duplication des fichiers et la multiplication des versions d'un même objet numérique, afin d'optimiser l'espace de stockage et de maîtriser les coûts d'infrastructure associés.

La volumétrie globale toutes données confondues est estimée à ce jour 7 To.

Le titulaire documentera :

- les règles d'accès
- les URLs internes et publiques
- les procédures de sauvegarde
- les procédures de restauration
- les modalités d'interfaçage avec Django, IIIF, NumaHOP et la faisabilité de l'interfaçage avec d'autres services/applications de l'infrastructure BVH (plateforme Arkindex, Wordpress, Heurist, XTF…)

L'organisation des buckets et les conventions de nommage seront établis en coordination avec l'équipe BVH qui pourra actualiser et contribuer à la documentation en MarkDown sur son GitLab.

Le stockage accueillera entre autre, en distinguant les différents types de ressources et leurs usages (conservation, traitement, diffusion) :

- les images des numérisations (fichiers issus directement du processus de numérisation, généralement en haute définition et sans traitement, servant de référence)
- les fichiers sources (fichiers maîtres éventuellement retraités ou normalisés à partir des images de numérisation, utilisés comme base pour la production de dérivés ou pour l'exploitation technique)
- les fichiers METS et ALTO (produits notamment dans le cadre des flux de numérisation et d'OCR, utilisés pour structurer les objets numériques et décrire le contenu textuel des images ; ils devront être stockés de manière organisée et interopérable avec les autres composants du système, notamment NumaHOP, l'API et les services d'indexation)
- les éditions textuelles du Corpus Epistemon en XML-TEI, les fichiers XML-TEI dérivés (formatés pour les outils d'analyse linguistique tels que TXM ou Philologic)
- les exports dérivés (par exemple : PDF de consultation, images JPEG compressées, fichiers OCR, exports CSV ou JSON pour diffusion ou analyse)
- les éventuels fichiers dérivés de consultation (par exemple : versions allégées des images pour affichage web, PDF feuilletables, extraits de documents, visualisations adaptées aux interfaces utilisateurs)
- les fichiers nécessaires à la génération ou à la diffusion IIIF : manifests et manifests de collection JSON
- les fichiers nécessaires à la génération des interfaces de consultation et à la recherche sur le nouveau portail BVH, le cas échéant

Les modalités de sauvegarde et de restauration (fréquence, localisation, responsabilité) seront dépendantes de l'infrastructure serveur de l'université. Le titulaire tiendra compte des recommandations de la DSI de l'université pour émettre des préconisations et paramétrer les services requis le cas échéant.

4.6. Serveur IIIF et Manifests

4.6.1. Migration et optimisation du serveur IIIF

Le serveur IIIF du portail BVH est actuellement en production sur une machine virtuelle administrée par le CESR et hébergée dans la grille de service de l'IR* Huma-Num. Les manifests IIIF sont aujourd'hui stockés localement sur cette machine virtuelle. Le serveur d'images (Cantaloupe) est configuré pour pointer vers un unique volume de stockage (actuellement un répertoire Huma-Num Box), ce qui limite la structuration et la séparation des données.

Il est prévu de rapatrier ce serveur IIIF sur une machine virtuelle administrée par l'université de Tours. Dans ce cadre, la mise en place d'un stockage objet compatible S3, de type MinIO ou solution équivalente, permettra de dépasser cette contrainte en offrant la possibilité de définir plusieurs volumes logiques (buckets) distincts. Cette organisation permettra une structuration des données par projet ou par corpus, avec des arborescences indépendantes, tout en assurant leur exposition cohérente via le serveur IIIF.

Une instance de test, incluant une mise à jour du serveur IIIF, a été déployée en interne en 2025 mais n'est pas encore pleinement fonctionnelle. Le titulaire prendra en charge la finalisation et la mise en production de cette instance : il proposera des optimisations et des préconisations techniques (performances, montée en charge, sécurité), puis les mettra en œuvre afin de garantir un service IIIF robuste et pérenne.

La continuité de service lors de la migration vers la nouvelle instance devra impérativement être assurée, avec un minimum de temps d'interruption de l'exposition des images et des manifests existants. Les URLS actuelles doivent rester stables.

Les prestations attendues comprennent :

- le paramétrage, l'optimisation et la sécurisation du serveur IIIF (configuration, cache, gestion des requêtes, montée en charge)
- la connexion au stockage S3 / MinIO (1 bucket cible par collection IIIF)
- la mise en œuvre des mécanismes garantissant la continuité de service pour l'exposition des images
- la mise en œuvre des mécanismes garantissant la continuité de service pour l'exposition des manifests IIIF existants, y compris des manisfests de collection
- l'intégration des identifiants ARK dans les manifests IIIF, si l'option est retenue
- des tests de compatibilité avec le visualiseur IIIF Mirador V4 ou toute solution retenue pour la bibliothèque numérique BVH (lot 02)
- des tests de compatibilité avec le visualiseur IIIF intégré à NumaHOP (lot 03)
- des tests de compatibilité avec les manifests IIIF exportés en sortie NumaHOP (si cette option est retenue dans le cadre du lot 03)
- la validation d'un corpus pilote
- la mise à jour de la documentation du serveur IIIF et des API afférentes

4.6.2. Mise à jour des Manifests IIIF & continuité de service

La mise à jour des scripts existants (Python) de génération des manifests IIIF conforme au modèle BVH / Biblissima+ sera assuré par la maitrise d'ouvrage. Ces scripts dépendent directement des choix techniques retenus dans le cadre de ce lot 01. Ces manifests concatènent notamment :

- les métadonnées issues de la base « Catafonds » (PostgreSQL, API REST)
- les métadonnées techniques requises des images exposées (stockage S3)
- les données de sommaire (solution à déterminer)
- éventuellement des données de la base image (solution à déterminer)

Dans l'hypothèse où la mise en œuvre des identifiants ARK ne serait pas retenue ou serait reportée, les manifests IIIF devront néanmoins s'appuyer sur la solution de correspondance et de redirection des URLs prévue à la section 4.8.1, afin de garantir la cohérence entre les anciennes URLs BVH, les nouvelles URLs publiques du portail, les identifiants internes des ressources, les URLs des images exposées et les liens de consultation associés aux manifests.

4.6.3. Caractéristiques du serveur IIIF actuel

- Serveur image : Cantaloupe
- API déployées : API présentation v3, API Image, API ChangeDiscovery (level 1)
- Format des images exposées : JPEG 2000 actuellement pour les numérisations les plus récentes et celles déjà sous IIIF ; JPEG sans perte ou TIFF pour les numérisations antérieures à 2017
- Manifests : stockés actuellement ; génération dynamique ou stockage à déterminer à l'issue de la prestation (selon option de script retenue ou non)
- Granularité actuelle des manifests :
    - BVH
        - 1 manifest par objet physique
        - 1 manifest de collection dans le cas de recueils factices
    - Renumar
        - 1 manifest par registre d'archives
        - 1 manifest de collection pour l'ensemble des registres du projet TMD-Renumar

4.7. [PSE obligatoire] Moteur d'indexation Elasticsearch / OpenSearch

La prestation relative à cette partie 4.7 constitue une PSE obligatoire du présent lot. Cette PSE ne sera mise en œuvre que si elle est expressément retenue par le maître d'ouvrage. Le titulaire devra donc la chiffrer séparément dans son offre.

Le prestataire mettra en place un moteur d'indexation destiné à améliorer les performances de recherche et à permettre l'interrogation de contenus complexes du portail BVH dans la base postgreSQL.

Le titulaire assurera le déploiement, le paramétrage et l'intégration du moteur d'indexation retenu. À ce titre, les prestations attendues comprennent notamment :

- l'installation et la configuration du moteur d'indexation (Elasticsearch, OpenSearch ou solution équivalente) dans un environnement conteneurisé
- la définition et la mise en œuvre des mappings d'index adaptés aux données BVH et validés avec le maitre d'ouvrage
- la configuration des index, et incluant si nécessaire :
    - la définition des analyzers (tokenizers, filtres, normalizers) adaptés aux langues et aux types de contenus
    - la mise en place des pipelines d'ingestion (ingest pipelines) pour le prétraitement, l'enrichissement et la normalisation des données avant indexation
- le développement des scripts d'indexation initiale et de réindexation
- l'intégration avec l'API REST et la base PostgreSQL
- la mise en place des mécanismes de synchronisation entre les données sources et les index
- la configuration des performances (sharding, réplication, cache) en fonction de la volumétrie

- la sécurisation des accès au moteur d'indexation
- la documentation des procédures d'exploitation, de maintenance et de réindexation
- la réalisation de tests de bon fonctionnement et de performance

Le moteur devra pouvoir indexer, selon les données disponibles :

- les métadonnées de « Catafonds » issues de PostgreSQL
- les données du référentiel data.bvh sous Wikibase
- les données de la base bibliographique Zotero
- les identifiants ARK (si option retenue)
- les liens vers les manifests IIIF
- les champs nécessaires aux facettes par exemple : auteur, titre, date, lieu, imprimeur, possesseur, type de document, corpus, langue, type d'objet numérique, statut de publication
- les champs nécessaires aux tris (par exemple : titre normalisé, date de publication, nom d'auteur, cote, identifiant interne, date de mise à jour)
- les données utiles aux interfaces du lot 2, notamment les champs nécessaires à l'affichage des résultats de recherche, aux filtres dynamiques, aux aperçus de notices, aux liens vers les ressources numériques et aux interactions avec le visualiseur IIIF.

À court terme, le moteur devra pouvoir évoluer pour permettre l'indexation des données textuelles exploitables par une recherche plein texte : les données de sommaires en priorité, et ensuite celles des éditions XML-TEI.

Le maître d'ouvrage, après concertation avec le titulaire du lot 02, fournira la liste détaillée des champs à indexer (métadonnées descriptives, identifiants, textes, liens, champs techniques) et la liste des facettes retenues pour les IHM de recherche et d'affichage des résultats.

Le titulaire fournira les *mappings* d'index, les scripts d'indexation et les procédures de réindexation.

Le titulaire proposera une solution de réindexation programmée de préférence (batch planifié), avec possibilité de déclenchement manuel ou automatique (à l'événement) en tenant compte des contraintes identifiées (volumétrie, durée, impact sur les performances, gestion des mises à jour incrémentales).

Le titulaire proposera une solution d'interface d'administration dédiée au moteur d'indexation (*ex.* intégration à un outil existant comme Django Admin, développement sur mesure avec un framework front-end, etc.).

Cette interface inclura au minimum les fonctionnalités suivantes :

- Consultation des index (état, volume, champs indexés)
- Suivi des processus d'indexation (statut, durée, erreurs)
- Déclenchement manuel de réindexations (totale ou partielle)
- Gestion des erreurs (logs, alertes, relances)
- Visualisation des logs et métriques de performance

4.8. Identifiants pérennes et gestion des URLs (publiques et redirections)

4.8.1. Solution de correspondance et de redirection des URLs

Le titulaire mettra en œuvre un mécanisme de gestion de correspondance et de redirection des URLs permettant d'assurer la continuité d'accès entre les anciennes URLs BVH et les nouvelles URLs publiques du portail et de ses services d'exposition.

Ce mécanisme permettra :

- de maintenir une table de correspondance entre les anciennes URLs BVH, les nouvelles URLs publiques, les identifiants internes des ressources, les identifiants OAI-PMH le cas échéant, le type de ressource et le statut de publication
- de rediriger automatiquement les anciennes URLs vers les nouvelles URLs publiques correspondantes
- de gérer les cas de ressources déplacées, remplacées, dépubliées, supprimées ou temporairement indisponibles
- de documenter les règles de redirection mises en œuvre
- de garantir que les URLs exposées dans l'entrepôt OAI-PMH, dans les exports et dans les éventuels flux partenaires restent cohérentes avec les nouvelles URLs du portail
- de prévoir une procédure de mise à jour de la table de correspondance en cas d'évolution ultérieure des URLs

Le titulaire proposera la solution technique la plus adaptée à l'architecture retenue, par exemple : règles de redirection au niveau du serveur web ou du reverse proxy, table de redirection gérée dans PostgreSQL, *middleware* Django dédié, script de génération de règles de redirection, ou toute autre solution documentée, maintenable et compatible avec l'infrastructure BVH.

Cette solution de base ne se substituera pas à un véritable système d'identifiants pérennes, mais elle devra garantir la continuité d'accès aux ressources existantes et limiter les risques de liens rompus lors de la migration vers le nouveau portail BVH.

Le titulaire fournira le cas échéant :

- le modèle de table de correspondance anciennes URLs / nouvelles URLs / identifiants internes / identifiants OAI-PMH / statut
- la documentation du mécanisme de redirection
- des exemples de redirections fonctionnelles
- la procédure de mise à jour et de contrôle des redirections
- un rapport de test des redirections

4.8.2. [PSE obligatoire] Solution ARK complet

La prestation relative à cette partie 4.8.2 constitue une option du présent lot. Cette prestation supplémentaire ne sera mise en œuvre que si elle est expressément retenue par le maître d'ouvrage. Le candidat devra donc la chiffrer séparément dans son offre. Dans le cas où la proposition du candidat serait retenue par le maître d'ouvrage, cette option complétera ou se substituera à l'offre de base de solution de gestion et de redirection des URLs (cf. partie 4.8.1).

Le titulaire mettra en œuvre un service de gestion et de résolution d'identifiants ARK, en s'appuyant de préférence sur un service intégré au back-end Django et à la base PostgreSQL, ou proposé sous forme de service autonome conteneurisé, à condition que son fonctionnement soit documenté, maintenable et compatible avec l'architecture technique BVH.

Le service sera compatible avec les recommandations de l'ARK Alliance et conçu de manière à garantir la stabilité des identifiants dans le temps, indépendamment des évolutions futures de l'architecture applicative, des URLs publiques ou des modalités de stockage des ressources.

Le titulaire prendra en compte le NAAN attribué à l'établissement et accompagnera le maître d'ouvrage dans les choix techniques liés à son utilisation, sans se substituer au maître d'ouvrage pour les démarches institutionnelles d'obtention ou de gouvernance du NAAN.

Le système permettra donc :

- de paramétrer et proposer une syntaxe stable des identifiants.
- l'attribution d'un ARK aux ressources identifiées par la maitrise d'ouvrage : le service devra générer un identifiant unique par ressource et vérifier qu'il n'a jamais été attribué.
- le stockage des ARK dans la base PostgreSQL : le stockage des liaisons dans une table de correspondance entre l'ARK, sa ressource BVH, son URL cible (nouveau portail BVH), son ancienne URL cible (ancien portail BVH), le type de la ressource, son statut (par ex. actif/redirigé/dépublié/supprimé…)
- la résolution des ARK vers les URLs publiques : quand un utilisateur appelle l'ARK, le service doit rediriger vers la ressource publique correspondante
- la gestion des redirections en cas d'évolution des URLs
- l'exposition des ARK via l'API REST
- l'intégration des ARK dans les manifests IIIF
- l'intégration des ARK dans les index de recherche Elasticsearch / OpenSearch (si l'option de développement correspondante est retenue)

Par défaut, la résolution publique d'un ARK devra diriger vers la page de consultation publique de la ressource sur le portail BVH, sauf cas particulier validé par la maîtrise d'ouvrage, par exemple résolution directe vers un manifest IIIF, un fichier XML-TEI publié ou une ressource numérique spécifique.

Le service de résolution des identifiants ARK mis en œuvre par le titulaire, pour chaque ressource identifiée, devra donc distinguer :

- l'accès à la ressource elle-même et à sa page de consultation publique
- l'accès aux métadonnées descriptives associées à l'identifiant
- l'accès à une information minimale sur la politique de persistance appliquée par BVH

Le titulaire devra proposer et documenter le mécanisme technique retenu pour ces différentes modalités de résolution, par exemple au moyen de paramètres ou d'inflexions de type `?info`, `?metadata` ou tout autre mécanisme conforme aux recommandations ARK retenues.

Cette documentation précisera :

- la syntaxe des ARK générés
- les types de ressources concernés
- les règles de redirection vers les ressources publiques
- le format des métadonnées retournées
- le comportement attendu lorsqu'une ressource est déplacée, remplacée, dépubliée, supprimée ou temporairement indisponible
- la liste des statuts de gestion des ARK
- les modalités de mise à jour des URLs cibles
- les responsabilités de maintenance du service de résolution
- les procédures de sauvegarde et de restauration de la table de correspondance entre ARK, ressources internes et URLs publiques

La documentation de la politique d'identifiants ARK BVH sera rédigée et mises à jour par le maître d'ouvrage en tenant compte des contraintes de la solution proposée par le titulaire. Elle précisera notamment les règles d'attribution, la granularité retenue, les conventions de nommage, les procédures de gestion et de maintenance, les modalités de pérennisation (garantie de stabilité des identifiants dans le temps, indépendamment des évolutions techniques) et de gouvernance des identifiants (rôles et responsabilités des acteurs, processus de validation, de modification et de suppression éventuelle).

Les identifiants ARK envisagés concernent prioritairement les ressources documentaires publiées de manière pérenne sur le portail BVH :

- ARK obligatoire pour les Objets physiques, les Objets numériques, les Manifestations
- ARK obligatoire pour les manifests IIIF
- ARK obligatoire pour les fichiers XML-TEI publiés
- ARK souhaité pour les notices Agents (personnes, collectivités et familles) et les Œuvres
- ARK souhaité pour les pages/images (pour les besoins de citation fine)
- Aucun ARK systématique n'est attendu pour les fichiers dérivés temporaires, les fichiers techniques intermédiaires ou les ressources non destinées à une exposition pérenne.

Afin de valider l'intégration dans l'infrastructure BVH, le titulaire fournira un jeu d'exemples d'ARK attribués et résolus, couvrant plusieurs types de ressources, ainsi qu'un protocole de test permettant au maître d'ouvrage de vérifier : l'unicité des identifiants générés, la résolution vers la ressource publique, la mise à jour des redirections, l'exposition des ARK dans l'API REST, l'intégration des ARK dans les manifests IIIF, l'indexation des ARK dans Elasticsearch / OpenSearch (si l'option correspondante est retenue), la restauration de la table de correspondance ARK en cas de sauvegarde ou de reprise.

4.9. [PSE obligatoire] Entrepôt OAI-PMH et exposition Dublin Core

La prestation relative à l'audit, à l'adaptation, au remplacement ou au redéveloppement de l'entrepôt OAI-PMH des BVH constitue une option du présent lot. Cette option ne sera mise en œuvre que si elle est expressément retenue par la maîtrise d'ouvrage.

Les BVH disposent déjà actuellement d'un entrepôt OAI-PMH indépendant, accessible à l'adresse suivante : https://www.bvh.univ-tours.fr/oai2/oai2.php?verb=Identify.

Cet entrepôt s'appuie aujourd'hui sur l'ancienne base Access des BVH ainsi que sur la plateforme XTF utilisée pour les éditions XML-TEI (corpus Epistemon). Dans le cadre de la refonte de l'infrastructure BVH, cet entrepôt devra être audité, maintenu, adapté ou remplacé afin de s'appuyer sur les nouvelles sources de données de l'architecture cible, en priorité la base PostgreSQL et l'API REST, et éventuellement étendu aux métadonnées issues des fichiers XML-TEI publiés, les manifests IIIF et les identifiants ARK.

Le titulaire réalisera un audit technique de l'entrepôt OAI-PMH existant afin d'identifier :

- son architecture actuelle
- les sources de données actuellement utilisées
- les formats de métadonnées exposés
- les sets existants
- les identifiants OAI actuellement produits
- les dépendances techniques
- les limites de maintenabilité
- les risques liés à la migration vers la nouvelle infrastructure BVH

À l'issue de cet audit, le titulaire proposera une solution technique argumentée parmi les scénarios suivants :

- mise à jour de l'entrepôt OAI-PMH existant afin de le raccorder aux nouvelles sources de données BVH
- remplacement de l'entrepôt existant par une solution open source maintenable et documentée
- développement d'un service OAI-PMH spécifique, intégré au back-end Django ou raccordé à l'API REST, à condition que cette solution soit conforme au protocole OAI-PMH, documentée, maintenable et compatible avec l'architecture technique BVH.

La solution proposée devra préserver la continuité du service d'exposition OAI-PMH, notamment pour les moissonneurs existants (Gallica, Isidore, etc.). Le titulaire précisera si la base URL actuelle de l'entrepôt peut être conservée ou si une nouvelle URL doit être mise en place. En cas de changement d'URL, le titulaire proposera les modalités de redirection ou de signalement nécessaires.

L'entrepôt OAI-PMH s'appuiera sur le mécanisme de correspondance et de redirection d'URLs retenu et présenté à la section 4.8 (voir *Supra* : soit l'offre de base *4.8.1 Solution de redirection*, soit l'option *4.8.2 Solution ARK complet*), afin de garantir la cohérence entre les identifiants OAI-PMH,

les anciennes URLs BVH, les nouvelles URLs publiques du portail et les identifiants internes des ressources.

L'entrepôt OAI-PMH devra respecter le protocole OAI-PMH 2.0 et prendre en charge a minima le format Dublin Core simple, non qualifié (`oai_dc`). La solution choisie devra également offrir la possibilité d'exposition de formats complémentaires, par exemple Dublin Core qualifié, EDM, voire METS, TEI simplifié ou tout autre format pertinent pour les usages BVH et Biblissima+.

Le service devra permettre :

- l'exposition des notices publiées sur le portail BVH
- l'exposition des métadonnées descriptives issues de PostgreSQL
- l'intégration des identifiants ARK dans les métadonnées exposées, si l'option correspondante est retenue ; à défaut, les anciens identifiants devront être conservés et des redirections serveurs ou applicatives devront être prévues
- l'intégration des liens vers les ressources publiques du nouveau portail BVH
- l'intégration des liens vers les manifests IIIF, le cas échéant
- l'intégration des liens vers les éditions XML-TEI publiées, le cas échéant
- la gestion de sets OAI-PMH permettant de distinguer les corpus, collections ou types de ressources ; la liste des sets devra correspondre a minima à l'actuel, d'autres sets pourront être envisagés
- la prise en compte des dates de création, de modification et de publication des notices
- la gestion des mises à jour incrémentales
- la gestion des notices supprimées, dépubliées ou devenues indisponibles, selon la politique retenue par la maîtrise d'ouvrage
- la validation des réponses OAI-PMH produites

L'équipe BVH fournira la documentation des *mappings* entre les données sources BVH et les champs Dublin Core exposés en OAI-PMH. Ces *mappings* préciseront, pour chaque champ exposé :

- le champ source dans PostgreSQL (ou dans l'API REST)
- la règle de transformation éventuelle
- le champ Dublin Core cible
- le caractère obligatoire, recommandé ou optionnel du champ
- un exemple de valeur
- les règles applicables en cas de champ vide ou de valeur multiple

La solution proposée par titulaire sera conçue pour ne plus dépendre des anciennes bases de données Access et pour limiter la dépendance à la plateforme XTF. Dans le cas de données encore non migrées et nécessaires à l'exposition OAI-PMH, le titulaire identifiera les dépendances et proposera une solution transitoire ou cible validée par le maître d'ouvrage.

Le prestataire fournira :

- une note d'audit de l'entrepôt OAI-PMH existant

- une recommandation argumentée sur la conservation, l'adaptation ou le remplacement de l'entrepôt
- la solution technique retenue, comprenant son architecture détaillée, ses composants logiciels, ses dépendances, ses modalités d'installation, de déploiement et de configuration, ainsi que sa documentation technique et fonctionnelle complète
- les fichiers de configuration nécessaires
- les scripts ou connecteurs d'alimentation depuis l'API REST, PostgreSQL ou les autres sources retenues
- des exemples de réponses pour les verbes `Identify`, `ListMetadataFormats`, `ListSets`, `ListIdentifiers`, `ListRecords` et `GetRecord` ;
- un protocole de test et un rapport de validation
- les procédures de maintenance, de mise à jour, de sauvegarde et de restauration

Le maître d'ouvrage assurera la mise à jour de la documentation des *mappings* Dublin Core et de la documentation des sets OAI-PMH.

Le titulaire vérifiera également le bon fonctionnement de l'entrepôt à l'appui d'un corpus pilote fourni par le maître d'ouvrage, pour démontrer que les métadonnées exposées sont moissonnables sans erreur ni anomalie par un client OAI-PMH standard tiers.

5. EXIGENCES TECHNIQUES TRANSVERSALES AU LOT 1

Les composants livrés dans le cadre du lot 1 respecteront les exigences techniques communes décrites dans le CCTP commun.

Le titulaire devra en outre respecter les exigences suivantes :

- Recours prioritaire à des composants open source
- Conteneurisation Docker
- Scripts de déploiement reproductibles
- Documentation complète
- Sécurisation des API et services
- Gestion des droits d'accès
- Logs applicatifs
- Tests automatisés lorsque pertinent
- Procédures de sauvegarde et restauration
- Procédures de mises à jour
- Compatibilité avec les contraintes des serveurs de l'établissement
- Livraison du code dans GitLab
- Respect des standards ouverts
- Maintien de la cohérence entre données sources, API, manifests IIIF, entrepôt OAI-PMH, identifiants ARK ou mécanisme minimal de redirection

6. LIVRABLES ATTENDUS DU LOT 01

Les livrables devront être fournis dans le dépôt GitLab de l'équipe BVH, dans des formats éditables, documentés et exploitables par l'équipe projet. Lorsque cela est pertinent, les livrables devront également être accompagnés d'exemples, de fichiers de configuration, de scripts, de procédures d'exécution et de tests.

Les livrables associés aux trois prestations supplémentaires — option système ARK, option moteur d'indexation Elasticsearch / OpenSearch, option entrepôt OAI-PMH— devront être clairement identifiés dans l'offre du candidat et faire l'objet d'un chiffrage distinct tel que cela est demandé dans le règlement de consultation ou dans le cadre de réponse financière.

N° livrable	Prestation concernée	Livrables attendus
L1-01	Audit technique initial	**Note d'audit de l'existant technique**, incluant l'état du back-end Django, de la base PostgreSQL, des anciennes bases Access concernées, du serveur IIIF, de l'entrepôt OAI-PMH et des dépendances applicatives
L1-02	Audit technique initial	**Note de cadrage technique** présentant les recommandations, les risques identifiés, les prérequis, les dépendances entre composants et les choix techniques à valider par la maîtrise d'ouvrage
L1-03	Back-end Django	**Code source des modules Django développés ou modifiés**
L1-04	Back-end Django	**Documentation des modèles**, services métiers, dépendances, paramètres de configuration et procédures d'exploitation
L1-05	Back-end Django	**Documentation des droits**, profils, comptes techniques et modalités d'administration, le cas échéant
L1-06	API REST	**Code source et configuration de l'API REST**
L1-07	API REST	**Documentation** de l'API : endpoints, paramètres, filtres, tris, pagination, formats de réponse, exemples de requêtes et réponses
L1-08	API REST	**Documentation** de la stratégie de versionnement de l'API, des règles de rétrocompatibilité et des modalités de dépréciation des endpoints
L1-09	Migration ou intégration des données complémentaires	**Scripts, procédures ou connecteurs de reprise des données issues des anciennes bases Access**, notamment sommaires et données relatives aux images, si cette prestation est retenue
L1-10	[le cas échéant] Cas de migration ou d'intégration des données complémentaires par le maitre d'ouvrage	**Rapport de reprise** indiquant les répercussions ou modification sur le back-end et les services afférents, ainsi que les contrôles effectués

L1-11	Stockage objet S3	**Configuration du stockage objet compatible S3**, incluant buckets, droits/accès, URLs internes et publiques
L1-12	Stockage objet S3	**Documentation** de l'organisation des buckets, des conventions de nommage, des règles d'accès, des procédures de sauvegarde et de restauration
L1-13	Serveur IIIF	**Configuration, paramétrage et documentation du serveur IIIF**
L1-14	Serveur IIIF	**Procédure de migration** ou de bascule depuis l'instance existante vers la nouvelle instance, incluant les mesures de continuité de service
L1-15	Moteur d'indexation Elasticsearch / OpenSearch, si option retenue	**Configuration du moteur d'indexation**, mappings, analyzers, pipelines d'ingestion et paramètres techniques
L1-16	Moteur d'indexation Elasticsearch / OpenSearch, si option retenue	**Scripts d'indexation initiale, de réindexation et de synchronisation** avec les données sources
L1-17	Moteur d'indexation Elasticsearch / OpenSearch, si option retenue	**Documentation d'exploitation** du moteur d'indexation : installation, réindexation, supervision, sauvegarde, restauration et diagnostic des erreurs
L1-18	Solution de redirection d'URLs	**Mécanisme de redirection** ancienne URL / nouvelle URL, documenté et maintenable
L1-19	Solution de redirection d'URLs	**Modèle de table de correspondance** ancienne URL / nouvelle URL / identifiant interne / identifiant OAI-PMH / type de ressource / statut
L1-20	Solution de redirection d'URLs	**Rapport de test des redirections**, indiquant les redirections valides, les cas ambigus et les URLs sans correspondance
L1-21	Identifiants ARK, si option retenue	**Service de gestion et de résolution des ARK**, intégré au back-end Django ou livré sous forme de service autonome conteneurisé
L1-22	Identifiants ARK, si option retenue	**Modèle de table de correspondance ARK** / ressource interne / URL cible / ancienne URL / type de ressource / statut / historique
L1-23	Identifiants ARK, si option retenue	**Documentation technique du service ARK** : syntaxe, règles d'attribution, résolution, redirections, statuts, sauvegarde et restauration
L1-24	Identifiants ARK, si option retenue	**Jeu d'exemples d'ARK** attribués et résolus, couvrant plusieurs types de ressources
L1-25	Entrepôt OAI-PMH, si option retenue	**Note d'audit de l'entrepôt OAI-PMH** existant

L1-26	Entrepôt OAI-PMH, si option retenue	**Recommandation argumentée** sur la conservation, l'adaptation ou le remplacement de l'entrepôt existant
L1-27	Entrepôt OAI-PMH, si option retenue	**Solution OAI-PMH retenue, installée, configurée et documentée**
L1-28	Entrepôt OAI-PMH, si option retenue	**Scripts ou connecteurs d'alimentation** depuis PostgreSQL, l'API REST, les fichiers XML-TEI ou les autres sources retenues
L1-29	Entrepôt OAI-PMH, si option retenue	**Exemples de réponses** pour les verbes Identify, ListMetadataFormats, ListSets, ListIdentifiers, ListRecords et GetRecord
L1-30	Déploiement et conteneurisation	**Dockerfiles, fichiers de configuration Docker, docker-compose ou dispositif équivalen**t
L1-31	Déploiement et conteneurisation	**Documentation d'installation, de configuration, de lancement, d'arrêt, de mise à jour et de retour arrière**
L1-32	Versionnement, GitLab et CI/CD	**Organisation du dépôt GitLab**, branches, tags, README, **procédures de livraison et documentation associée**
L1-33	Versionnement, GitLab et CI/CD	**Fichiers de configuration, scripts ou procédures** permettant l'intégration des composants du lot 1 à la chaîne de livraison, de test et de déploiement définie dans le CCTP commun
L1-34	Logs et traçabilité	**Documentation des logs** produits par les composants du lot 1 : emplacement, format, niveaux, rotation, consultation, purge et précautions relatives aux informations sensibles
L1-35	Sécurité	**Documentation des paramètres de sécurité, comptes techniques, variables sensibles, droits d'accès et modalités de protection des services exposés**
L1-36	Tests et recette	**Protocoles de tests couvrant les composants livrés** : API REST, PostgreSQL, S3 / MinIO, IIIF, manifests, ARK ou redirection minimale, OAI-PMH, indexation si option retenue
L1-37	Tests et recette	**Rapports de recettes** indiquant les tests réalisés, les résultats obtenus, les anomalies corrigées, les anomalies restantes et les réserves éventuelles
L1-38	Sauvegarde et restauration	**Procédures de sauvegarde et de restauration pour les composants livrés ou modifiés** : stockage S3, index, tables ARK ou redirection, configuration, OAI-PMH, IIIF
L1-39	Documentation générale	**Documentation technique complète du lot 1**, incluant architecture, installation, exploitation, maintenance, dépendances et points de vigilance
L1-40	Rapport final	**Rapport final de livraison** synthétisant les prestations réalisées, les composants livrés, les choix techniques, les limites, les risques résiduels et les recommandations de maintenance

7. MODALITÉS DE DÉPLOIEMENT SPÉCIFIQUES AU LOT 01

Les modalités générales de déploiement, de mise en recette, de mise en production, de sauvegarde préalable et de retour arrière sont définies dans le CCTP commun.

Pour le lot 1, le titulaire appliquera ces exigences aux composants relevant de son périmètre, notamment :

- le back-end Django
- l'API REST
- la base PostgreSQL et ses éventuelles migrations
- le stockage objet S3 / MinIO
- le serveur IIIF et les manifests
- le moteur d'indexation Elasticsearch / OpenSearch, si l'option est retenue
- la solution de gestion et de redirection des URL, ou le service ARK le cas échéant, si l'option est retenue
- l'entrepôt OAI-PMH, si la PSE est retenue

Le titulaire fournira, pour chacun des composants livrés ou modifiés, les fichiers de configuration, scripts, variables d'environnement documentées et procédures nécessaires à un déploiement reproductible en environnement de développement, de recette et, le cas échéant, de production.

Une attention particulière sera portée aux opérations de migration ou de bascule depuis les services existants, notamment le serveur IIIF actuel, l'entrepôt OAI-PMH existant, les anciennes bases Access et les anciennes URLs BVH.

Le titulaire documentera les prérequis, les dépendances entre composants, les étapes de déploiement, les sauvegardes préalables, les points de contrôle, les procédures de vérification après déploiement et les modalités de retour arrière.

8. TEST & RECETTES SPÉCIFIQUES AU LOT 01

La recette du lot 1 et des options retenues devra vérifier le bon fonctionnement de l'ensemble des composants livrés.

Pour chaque composant livré, la recette s'appuiera sur un corpus pilote fourni par le maître d'ouvrage et sur des critères de validation documentés : déploiement reproductible, absence d'erreur bloquante, conformité des formats produits, accessibilité des services, documentation fournie et validation par l'équipe BVH.

Les tests porteront notamment sur :

- le déploiement Docker
- la connexion à PostgreSQL

- la saisie, consultation, recherche dans Django Admin
- l'exposition de l'API REST
- la validité des réponses API
- l'accès au stockage S3
- l'exposition des manifests IIIF
- l'ouverture des manifests dans un visualiseur IIIF
- le bon fonctionnement du mécanisme de correspondance et de redirection des URL
- l'attribution et la résolution des ARK, si l'option est retenue
- l'indexation et la recherche, si l'option est retenue
- le fonctionnement de l'entrepôt OAI-PMH, si option retenue
- la validité des réponses OAI-PMH, si option retenue
- la documentation fournie
- la sauvegarde et la restauration

9. SÉCURITÉ

Les exigences générales de sécurité sont définies dans le CCTP commun. Le titulaire les appliquera à l'ensemble des composants relevant du lot 1.

Le titulaire devra prévoir et assurera :

- la séparation entre les services publics, les services internes et les interfaces d'administration
- la protection des interfaces d'administration, notamment Django Admin ou toute interface équivalente
- la gestion des droits d'accès aux API publiques, restreintes ou d'administration
- la sécurisation des accès à PostgreSQL, au stockage S3 / MinIO et aux éventuels services d'indexation
- la gestion sécurisée des variables sensibles, mots de passe, clés d'accès, jetons techniques et fichiers de configuration
- la non-exposition publique des services internes qui n'ont pas vocation à être accessibles depuis l'extérieur
- la journalisation des erreurs et opérations critiques, sans inscription de secrets ou données sensibles dans les logs
- la protection contre les robots et attaques automatisées
- les recommandations de mise à jour de sécurité des composants livrés

Pour l'API REST, le titulaire prévoira et distinguera les endpoints publics, les endpoints restreints et les endpoints d'administration requis, et précisera les mécanismes de contrôle d'accès associés.

Pour le stockage objet S3, il documentera les règles d'accès aux buckets, les droits en lecture et écriture, les comptes ou services autorisés, ainsi que les précautions à prendre pour éviter toute exposition non maîtrisée des fichiers.

Pour le serveur IIIF et les manifests, il précisera les modalités d'exposition des images et les éventuelles restrictions appliquées selon les collections ou types de ressources.

Pour l'entrepôt OAI-PMH, il s'assurera que seules les données et ressources destinées à l'exposition publique sont moissonnables.

Pour le service ARK ou la solution de gestion des URL retenue, les règles de redirection seront maîtrisées, documentées et ne permettront pas de redirections ouvertes non contrôlées.

Le titulaire fournira une documentation de sécurité précisant les choix effectués, les services exposés, les services internes, les comptes techniques, les droits associés et les points de vigilance pour l'exploitation et la maintenance.

10. DOCUMENTATION ET TRANSFERT DE COMPÉTENCES

Le titulaire établira une documentation permettant à l'équipe BVH de maintenir, exploiter et faire évoluer les composants livrés.

Une session de transfert de compétences pourra être proposée par les candidats, à destination des administrateurs et développeurs de l'équipe BVH.

Le cas échéant, cette session de transfert de compétence pourra porter sur :

- l'architecture du lot 1
- l'installation
- les commandes principales
- la structure du code
- les modèles Django
- l'API REST
- le serveur IIIF
- la solution de gestion/redirection des URL
- la solution ARK, si l'option est retenue
- l'indexation, si l'option est retenue
- l'entrepôt OAI-PMH, si l'option est retenue
- les procédures de sauvegarde et restauration
- les erreurs fréquentes et leur diagnostic