

Résumé de l'échange

- 1) Besoin initial : trouver des bâtiments en vente (site gouv).
- 2) Question élargie : existe-t-il un site gouv qui liste les « sites abandonnés » ? → Réponse : pas de portail unique officiel ; proposer un recouplement de bases publiques (Mérimée/POP, Cartofriches, GéoRisques BASIAS/BASOL, OSM/Overpass, ANFR/Cartoradio, GPU/Cadastre, IGN).
- 3) Prototype front (côté navigateur) : carte Leaflet démonstrative (non-liée ici, hors périmètre PDF).
- 4) Demande « pro » : une app React/Next.js (RJS) sur GitHub, couvrant toute la France, filtre par région, recouplement multi-sources, note d'abandon en étoiles.
- 5) Ajouts requis : voies de repli (fallbacks), console détaillée, auto-critique, anticipation des blocages, roadmaps prêtes à coder.

Objectifs & périmètre (Definition of Done)

- Robustesse : pas d'écran vide ; toujours un résultat dégradé (snapshots/caches) avec explication.
- Traçabilité : provenance, date de fraîcheur et version pour chaque donnée affichée.
- Observabilité : console intégrée + logs structurés exportables.
- Auto-critique : rapport « santé & fraîcheur » (sources, dépendances, score).
- Éthique & légal : exclusion par défaut des sites privés, militaires, dangereux ; bannière permanente.
- Couverture : France entière ; filtres Région/Département/Commune ; export CSV/GeoJSON.

Sources à agréger (publiques & ouvertes)

- Patrimoine : Base Mérimée (POP – Ministère de la Culture) – repérage des notices « ruines/vestiges ».
- Fiches : Cartofriches (Cerema) – inventaire national.
- Anciens sites industriels / sols : BASIAS (inventaire historique) & BASOL (sites et sols pollués) via GéoRisques.
- Données communautaires : OpenStreetMap (tags historic=ruins, préfixes lifecycle abandoned:*, disused:*) , via Overpass/Geofabrik.
- Limites administratives : ADMIN-EXPRESS (IGN) + COG/INSEE pour région/département/commune.
- Contexte radio : ANFR/Cartoradio (statut des émetteurs) – indicateur d'activité plutôt que preuve d'abandon.
- Imagerie/chronologie : IGN « Remonter le temps » (indice visuel, non bloquant dans le score).
- Urbanisme/foncier (contexte) : Géoportail de l'Urbanisme, Cadastre.

Score « Abandon » explicable (0 à 5 ★)

Principe : score probabiliste et transparent, jamais un droit d'accès.

- +2.0★ si notice Mérimée contient « ruines »/« vestiges » (preuve patrimoniale).
- +2.0★ si OSM comporte un préfixe lifecycle `abandoned:*
- +1.5★ si OSM `historic=ruins` ; +1.0★ si `disused:*
- +1.0★ si le site figure dans Cartofriches (signal de friche).
- +0.5★ si BASIAS indique une cessation (industrielle/activité).
- -2.0★ si OSM signale `access=private|no` ; exclusion si `military=*` (masqué par défaut).
- +0.5★ (indice visuel) si les séries IGN montrent une ruine stable > 10 ans (affiché comme indice, pondération faible).

La fiche affiche toujours « Pourquoi X★ ? » (explication des règles déclenchées).

Fusion, dédoublonnage & justification

- Dé-doublonnage spatial : cluster ≤ 50–100 m (centroïdes/boîtes), choix de la géométrie la plus précise (polygone GPU > point OSM).
- Fusion sémantique : similarité de nom/type, source majoritaire, horodatage.
- Piste d'audit : pour chaque attribut, conservation de la source et de la date.

```

# Architecture technique (pro)
Front (Next.js/React, TypeScript)
• Carte (MapLibre/Leaflet), clustering, filtres Région/Département, curseur de note, mode « visite publique » (ouverts/autorisés).
• Panneau Console (onglets Pipeline/Sources/Données/Scoring/Dépendances/Journal/Licences).
• Feature flags : Mode Démo, Caches uniquement, Strict légal.

Backend « Data Hub » (serverless Node + PostGIS)
• Connecteurs par source ; ETL planifié (GitHub Actions) : ingestion → normalisation → fusion → scoring → publication de snapshots régionaux (GeoJSON/tiles).
• Cache/CDN edge des résultats (grille H3).
• Circuit breaker par source, bascule vers snapshots si erreurs récurrentes.

# Stratégies de repli (fallbacks)
Pour chaque source : PRIMARY → SECONDARY → SNAPSHOT.
• OSM : PRIMARY Geofabrik (pré-calcul) ; SECONDARY Overpass (bbox petites, backoff) ; SNAPSHOT dernier export régional.
• Mérimée : PRIMARY API data.culture ; SECONDARY dump/mirror ; SNAPSHOT export régional.
• Cartofriches : PRIMARY API ; SECONDARY export CSV/GeoJSON ; SNAPSHOT cache versionné.
• BASIAS/BASOL : PRIMARY API GéoRisques ; SECONDARY fichiers de référence (si dispo) ; SNAPSHOT dernier agrégat.
• ANFR : PRIMARY Open Data ; SECONDARY absence tolérée (n'agit que sur le score) ; SNAPSHOT dernier dump.
• AdminExpress/INSEE : millésimés → pas de live, snapshot courant.
Règles : timeout 8 s, 2 retries (jitter), badge « Résultats partiels » si bascule, snapshot régional si tout échoue.

# Console détaillée (in-app) & export diagnostic
Onglets :
1) Pipeline : chronologie des étapes (Fetch → Normalize → Merge → Score → Serve).
2) Sources : latence, rate-limit, dernière synchro, fallback actif.
3) Données : nb points, doublons fusionnés, % champs manquants.
4) Scoring : histogramme des étoiles, top règles, anomalies.
5) Dépendances : versions npm, « npm audit/outdated » synthèse.
6) Journal : logs JSONL (ts, corr_id, source, bbox, durée, outcome).
7) Licences : ODbL/Etalab/IGN/ANFR.
Fonctions : export « bundle diagnostic » (.zip), test de la bbox courante, simulateur de panne (feature flag).

# Auto-critique & rapports « Santé »
• Fraîcheur par région/source : âge du snapshot, % succès ETL.
• Couverture : densité vs moyenne nationale (alerte si trous).
• Drift du score : dérive (z-score) vs n-1.
• Qualité fusion : taux de conflits nom/position, orphelins, exclus privés/militaires.
• Dépendances : CVE critiques, packages « outdated » (priorisation).
• Règles obsolètes : détection de tags OSM dépréciés (proposition de mise à jour).
• Changelog : libs/schémas modifiés depuis la dernière release.
Rapport lisible dans l'app + exportable.

# Moteur de règles (scoring) – gouvernance
• Règles déclaratives (YAML/JSON) versionnées, avec poids/conditions.
• « Shadow mode » : tester une nouvelle version de règles en parallèle, comparer les sorties sans impacter l'utilisateur.
• Traçabilité des explications stockée avec chaque site.

```

```

# Tests, QA & résilience
• Unit : connecteurs, normalisation, scoring.
• Contract : conformité des API aux schémas attendus.
• E2E : parcours carte → liste → fiche → export.
• Charge : carte IDF, cache chaud < 1,5 s médian.
• Chaos : panne simulée d'une source → fallback + badge.
• Snapshots visuels : cartes & fiches pour PR review.

# Anticipation des blocages (par source)
• Overpass : quotas/429/timeout → bbox petites, backoff, Geofabrik prioritaire.
• data.culture : pagination/pertinence → requêtes par région + mots-clés contrôlés ; validation lat/lon.
• GéoRisques : variations schéma → normalisation stricte, mapping INSEE.
• Cartofriches : refresh irrégulier → tolérer, préférer snapshot si lenteur.
• ANFR : volumes → requêtes buffer autour de candidats, pas de nationwide live.
• IGN/GPU : changements d'URL WFS/WMS → canary quotidien, config centralisée.

# Sécurité, licences, conformité
• Filtre légal strict par défaut (privés/militaires/dangereux masqués).
• Mentions ODbL (OSM), Etalab/MTE/Cerema, IGN/ANFR visibles.
• Respect robots/rate-limits ; aucun scraping intrusif ; pas de données personnelles.

# Schéma de données (simplifié)
Table « sites »
• id, geom, name, kind (patrimonial | friche | indus-historique | ruins-osm)
• region_code, dept_code
• sources (JSON : merimee, cartofriches, basias, basol, osm/tags)
• access_flag (public/unknown/exclu), risk_flags (military, etc.)
• score (0-5), score_reasons (JSON), last_seen, links (URLs).

# UX de la note en étoiles
• 0-1★ : indice faible (donnée communautaire seule, peu probante).
• 2-3★ : indice moyen (OSM + BASIAS, ou Mérimée seule).
• 4-5★ : indice fort (Mérimée + OSM + Cartofriches ± BASIAS).
• Icône cadenas si accès non public ; masqué par défaut pour conformité.

# Checklist « Ce que te dirait Claude »
• Ne jamais requérir Overpass « France entière » en live ; préférer Geofabrik + pré-calcul.
• Toujours disposer d'un snapshot régional prêt (downgrade instantané).
• Scoring explicable, versionné ; A/B en shadow avant activation.
• « Why no result? » : expliquer clairement (filtres trop stricts, source down).
• Dead-man switch : si 3 sources clés down → Mode Démo (snapshots + bannière).
• Health Report public (transparence) : fraîcheur par région/date.

# Roadmap d'implémentation
Sprint 1 – Squelette robuste
• Repo front + backend ; feature flags ; Console (Pipeline/Sources/Journal) ; wrapper fetch + circuit breaker ; snapshots IDF de démo.

Sprint 2 – Data & Fallbacks
• Connecteurs OSM (Geofabrik) & Mérimée ; ETL Actions ; PostGIS ; dédoublonnage initial ; snapshots par région ; UI filtres région.

Sprint 3 – Scoring & Auto-critique

```

- Moteur de règles v1 + explications ; Health Report ; page « État des sources » ; export diagnostic .zip.

Sprint 4 – Sources étendues

- Cartofriches + BASIAS/BASOL + ANFR ; réglage pondérations ; tests de charge ; console (simulateur de panne).

Sprint 5 – Qualité & Prod

- Chaos tests ; « shadow rules » ; mentions licences ; runbook incident ; release v1.0 (Vercel/Pages) + snapshots France.

Annexes (références officielles à intégrer dans la doc du repo)

- POP/Mérimée (Ministère de la Culture) – API data.culture.gouv.fr
- Cartofriches (Cerema) – open data/API foncière
- GéoRisques – BASIAS & BASOL – API & portails
- OpenStreetMap – Overpass API – conventions de tagging lifecycle
- ADMIN-EXPRESS IGN – limites administratives ; COG/INSEE
- ANFR/Cartoradio – Observatoire des ondes/open data
- IGN – « Remonter le temps » ; GPU/Cadastre (contexte)

(NB : les liens officiels seront listés dans le README du repo, avec licences.)