DOMUSCAN — Piano di implementazione

> Roadmap ad alto livello: fasi, obiettivi, cosa si consegna e quando una fase si considera chiusa. Non è una lista di task. Per il "cosa" e il "come" vedi PROGETTO.md.

La sequenza segue i SAL dell'offerta (Fasi A, B, C ≈ 10–12 settimane), preceduti da una fase 0 di preparazione e seguiti dalle evoluzioni V1 e oltre previste da Capitolato e Global Shadow Layer.

```
Fase 0  Decisioni + corpus pilota          (prima del SAL 1)
Fase 1  Fondamenta + acquisizione          ── SAL 1  (~3 settimane)
Fase 2  Registro, ponte, matching          ┐
Fase 3  Print Layer + Review Console       ┘ SAL 2  (~4 settimane)
Fase 4  Resolver, accesso, collaudo, go-live ── SAL 3  (~3-4 settimane)
Fase 5  V1: stampante virtuale, SdI, IA, scala
Fase 6  Evoluzione: congruità, anomalie, Country Pack
```

Fase 0 — Decisioni e corpus pilota

Obiettivo: iniziare a scrivere codice con decisioni chiuse e dati reali su cui misurare.

- Chiudere i punti aperti di PROGETTO.md §11: Vault o puntatore, regola di auto-associazione, tipo di sigillo, perimetro MVP rispetto al budget, autenticazione del condomino.
- Raccogliere un **corpus pilota autorizzato**: bilanci/consuntivi PDF e lotti di fatture XML/PDF di almeno uno studio con Danea Domustudio e, se possibile, uno con un secondo gestionale. Includere casi difficili: importi uguali, note di credito, fatture ripartite, voci aggregate, PDF scansionati.
- Etichettare un campione come **ground truth** (quale fattura corrisponde a quale riga).
- Verificare sul gestionale reale quali campi (causale, note, protocollo) **compaiono nelle stampe/esportazioni** del consuntivo: da qui dipende la modalità di aggancio del token.
- Definire gli **indicatori di qualità**: accuratezza di estrazione, precision/recall del matching (separando auto-associati e revisionati), tempo umano per lotto.
- Scegliere lo stack definitivo, il provider cloud e l'organizzazione del repository; attivare i domini (`domuscan.it` risultava in stato *inactive/dnsHold* alla registrazione).
- Avviare il percorso **legale/privacy** (ruoli GDPR, informative, accordi con gli studi), che deve chiudersi prima di usare dati reali in produzione.

Fatto quando: decisioni scritte, corpus disponibile e etichettato, metriche concordate, ambienti cloud predisposti.

Fase 1 — Fondamenta e acquisizione (SAL 1)

Obiettivo: ogni documento che entra in DOMUSCAN viene conservato in modo integro, identificato in modo univoco e letto nei suoi dati fiscali.

Made with Markdown Inline Preview

- **Infrastruttura Cloud-Native**: container, ambienti (dev/staging/prod), CI/CD, gestione dei segreti, backup e logging. Dal giorno uno: multi-tenant e audit log append-only.
- **Modello dati canonico** (tenant, condominio, esercizio, documento, fornitore, voce, associazione, utente, audit). Progettarlo già estendibile al modello globale (Country Pack Italia come primo pacchetto, senza logica nazionale nel core).
- **Intake**: upload multi-file e cartella sorvegliata come minimo; email/PEC dedicata se rientra nei tempi.
- **Document Vault**: originale immutabile e cifrato, SHA-256, provenienza (canale, data, utente), deduplicazione idempotente (stesso file da due canali = un documento, due provenienze).
- **Estrazione**: parser FatturaPA XML completo (più aliquote, ritenute, note di credito, righe); parser PDF con testo e coordinate; OCR solo per le pagine che lo richiedono. Ogni campo con la sua provenienza.
- **Identità del documento**: Document ID, chiave fiscale canonica, generazione del **Binding Token** (`DSN-XXXX-XXXX` , Base32 con checksum, senza dati personali).
- **Pipeline asincrona a job** idempotenti e ripetibili, con stati ed errori visibili.

Fatto quando: il corpus pilota viene importato da più canali senza duplicati, ogni fattura ha hash, ID, token e dati estratti tracciabili; il riavvio di un job non crea doppioni. → **Collaudo SAL 1.**

Fase 2 — Registro delle associazioni, ponte e matching

Obiettivo: DOMUSCAN conosce con certezza quale documento corrisponde a quale registrazione, e sa riconoscere quando non lo sa.

- **Binding Registry** molti-a-molti con importi allocati, macchina a stati (`PROPOSED → VERIFIED → CONFIRMED → PUBLISHED` , `REJECTED` , `REVOKED`), versioning e revoca: nessuna sovrascrittura, ogni correzione è una nuova entry.
- **Ponte di registrazione** (modalità A): l'operatore seleziona la fattura e ottiene negli appunti la causale con il token, pronta da incollare. Versione minima web; widget desktop se il budget lo consente, altrimenti in Fase 5.
- **Parser del rendiconto**: ricostruzione delle righe del consuntivo PDF (voce, importo, causale, gerarchia, coordinate), con un approccio ad adapter resistente a piccoli cambi di layout.
- **Matching Engine** con l'ordine di priorità token → protocollo → chiave fiscale → combinazione univoca → proposta, e i livelli di decisione A–E. Regola fail-closed applicata nel codice e verificata dai test.
- **Casi complessi**: gruppo di voci sulla stessa fattura (QR condiviso solo se la somma chiude al centesimo), voci aggregate, note di credito, acconti, fatture sostituite.
- **Sigillo** secondo la variante scelta in Fase 0, catena degli storni compresa.
- **Controlli di quadratura** e report delle eccezioni (documenti non associati, righe senza documento).

Fatto quando: sul corpus pilota tutte le righe sono associate correttamente o finiscono in revisione; **zero associazioni automatiche errate**; i test T01–T10 della Specifica DIBP passano.

Fase 3 — Print Layer e Review Console (SAL 2)

Obiettivo: l'amministratore ottiene un consuntivo con i QR e gestisce le eccezioni in pochi click.

- **Print Layer / overlay**: dal PDF del consuntivo genera un PDF arricchito con QR sulle righe (o richiami numerati + pagina indice quando i margini non bastano), senza coprire o alterare testi e importi. Eventuale rimozione del token dalla versione per i condomini. Controllo automatico prima/dopo su importi e totali.

Made with Markdown Inline Preview

- **Ingresso del consuntivo**: upload e cartella `DOMUSCAN_OUT`. La stampante virtuale Windows è un canale d'ingresso, pianificata per la V1 salvo diverso accordo.
- **Preflight di pubblicazione**: tutti i controlli prima di emettere un QR (stato del binding, condominio, importo, hash, privacy).
- **Review Console**: cruscotto (verificate / eccezioni), lista delle eccezioni con motivo, azioni di conferma, rifiuto, associazione manuale, split, merge, rettifica e recupero token; ogni azione è tracciata. Le correzioni approvate entrano nel set di regressione.
- **Migrazione dello storico**: import massivo di gestioni già contabilizzate senza token, matching fiscale, revisione dei casi ambigui, pubblicazione solo dopo la quadratura.

Fatto quando: dato un bilancio reale e il suo lotto di fatture, un operatore importa, risolve le eccezioni e ottiene un PDF con QR corretti e numeri originali invariati, anche su impaginazioni dense. → **Collaudo SAL 2.**

Fase 4 — Resolver, accesso e go-live (SAL 3)

Obiettivo: il condomino verifica una spesa dal QR in sicurezza e il sistema va in produzione.

- **Resolver pubblico** (`verify.domuscan.it`): codici opachi, casuali, non enumerabili e revocabili; verifica in tempo reale di binding, stato, permessi e **integrità dell'hash**; blocco della visualizzazione in caso di alterazione.
- **Secure Viewer e scheda spesa**: documento in sola lettura, relazione con la riga del bilancio, stato di revisione e versione; pagina di riepilogo per le voci aggregate; dati mancanti dichiarati come "non disponibili".
- **Accessi e ruoli**: autenticazione del condomino (modalità scelta in Fase 0), autorizzazioni per condominio e documento, revoca senza ristampa, log degli accessi. Test automatici anti-accesso cross-tenant.
- **Sicurezza operativa**: cifratura, rotazione delle chiavi, backup e restore provati, monitoraggio, procedura per gli incidenti.
- **Collaudo su flussi reali**: batteria T01–T12 della Specifica DIBP e T01–T07 del Capitolato, compreso il caso "API del gestionale assenti". Ristampa idempotente.
- **Consegna**: codice e istruzioni di deploy, diagramma architetturale, schema dati, contratti API, manuale operatore, piano di retention e incidenti; gate legale/privacy chiuso.

Fatto quando: pilota in produzione con almeno uno studio reale, test superati, documentazione consegnata. → **Collaudo SAL 3.**

Fase 5 — V1: estensione e scala

Obiettivo: ridurre il lavoro manuale e aprire il prodotto a più studi e gestionali.

- **Stampante virtuale Windows** con anteprima, scelta di condominio ed esercizio, retry e fallback.
- **App desktop di supporto** (overlay "Incolla in causale", notifiche, modalità C/D di aggancio).
- **Integrazione SdI / cassetto fiscale / provider** per acquisire gli XML alla fonte e generare il PDF di cortesia; stato del pagamento nella scheda.
- **Scheda spesa con spiegazione IA** vincolata alle evidenze, con prompt e modello versionati, minimizzazione dei dati inviati e test su input ostili.
- **Tassonomia condominiale** versionata e normalizzazione dei componenti (manodopera, materiali, chiamata…).
- **Import massivo, API pubbliche versionate, webhook, primi adapter opzionali** per gestionali, dashboard delle metriche di qualità.

Made with Markdown Inline Preview

- **Marchio** "Bilancio intelligente Domuscan" sul documento, con una definizione chiara di cosa significa "verificato".
- Validazione **multi-gestionale**: stessa pipeline su documenti di almeno due gestionali diversi.

Fase 6 — Evoluzione

Obiettivo: passare dalla verifica di autenticità al valore analitico e all'espansione.

- **Congruità / benchmark** su prezzari regionali, listini delle Camere di commercio e storico autorizzato, con snapshot riproducibili e il fallback "dati insufficienti" (rivendicazione 3).
- **Anomalie storiche** con evidenze e tasso di falsi positivi misurato.
- **Report per assemblea o contestazione** (SaaS), portale con SPID, marca temporale eIDAS.
- **Global Shadow Layer**: Country Pack versionati, wave internazionali A–E con gate G0–G5 per paese (legale + corpus → capture → parse → match → publish → benchmark). Un nuovo paese si aggiunge con un Country Pack e un corpus di test, senza toccare il core.
- White label, SSO, integrazioni ufficiali con i produttori di gestionali come servizio esterno.

Trasversali, da mantenere in ogni fase

- **Fail-closed e zero falsi positivi** come criterio di accettazione permanente.
- **Test di regressione** sul corpus con ground truth, che cresce con ogni correzione approvata.
- **Versioning** di parser, regole, tassonomia, modelli e Country Pack in ogni record di analisi; lo storico resta riproducibile.
- **Privacy by design**: minimizzazione dei dati, isolamento dei tenant, nessun segreto o fattura nei log.
- **Proprietà intellettuale**: allineare le scelte tecniche (sigillo, catena degli storni, verifica indipendente) con la bozza di rivendicazioni del consulente brevettuale.
- **Misurare, non stimare**: nessun indicatore o badge mostrato all'utente senza una definizione e una query riproducibile.

Made with Markdown Inline Preview