SPECIFICA DI PROGETTO / VERSIONE 1.0

Domuscan
Layer di stampa intelligente

Capitolato operativo per il team di sviluppo. Funzionamento indipendente dalle API dei gestionali condominiali; documenti verificabili, spiegazioni IA e confronto di congruità fondato su dati tracciabili.

Destinatari	Decisione tecnica	Obiettivo commerciale
Programmatore, architetto software, product owner	Core documentale autonomo; i connettori ai gestionali sono opzionali	Accumulare qualità dei dati, metodo, storico e riconoscibilità Domuscan

Mandato. Il prodotto acquisisce il rendiconto e le fatture da file esportati o da stampa; collega ogni voce alle evidenze; produce un PDF arricchito e una pagina web autorizzata per ogni spesa. Il gestionale di origine non deve chiamare né conoscere Domuscan.

Obiettivo difendibile. Nessuna specifica può impedire a Danea o ad altri di sviluppare funzioni simili. La differenziazione perseguita è un sistema trasversale ai gestionali con riconciliazioni validate, tassonomia specialistica, comparabili autorizzati e tracciabili, misure di qualità e domanda per il marchio.

Vincolo di veridicità. Non mostrare numeri di comparabili, intervalli di prezzo, confidenze o badge di verifica come risultati effettivi finché non sono calcolati, misurati e supportati da evidenze. I numeri eventualmente citati negli esempi sono illustrativi.

1. Ambito, priorità e definizione di completamento

Priorità	Consegna
MVP	Pipeline PDF/XML, associazione con revisione, PDF con riferimenti, QR protetto, scheda spesa, spiegazione ancorata ai documenti, log e isolamento tenant.
V1	Stampante virtuale Windows, import batch più ampio, benchmark validati, console feedback, API pubbliche e primi adapter opzionali.
Evoluzione	Anomalie storiche, granularità componenti avanzata, rete di comparabili più ampia, grafi analitici, marchio e distribuzione.

Definizione di completamento MVP. Dato un bilancio PDF digitale e un lotto FatturaPA XML/PDF, un operatore importa, verifica le associazioni dubbie, genera un PDF senza alterare i numeri originali e un condòmino autorizzato apre dal QR soltanto le spese cui ha accesso. Se i dati non sostengono un benchmark, la scheda lo dichiara.

2. Architettura: indipendenza dal gestionale

Flusso: PDF di bilancio / stampa virtuale / XML FatturaPA / allegati → ingestione e parsing → modello canonico → riconciliazione e revisione → tassonomia e componenti → motore benchmark → IA esplicativa → PDF arricchito, QR resolver e scheda spesa.

A01 Core autonomo | MVP

Nessun processo critico può richiedere API, login o formato proprietario di Danea. Supportare almeno upload PDF e XML FatturaPA. Tutti gli adapter dei gestionali sono moduli separati; un errore dell'adapter non blocca i flussi basati sui file.

Accettazione: Eseguire il caso dimostrativo con rete/API dei gestionali assenti; ottenere il PDF arricchito e le schede consultabili.

A02 Canali di acquisizione | MVP

MVP: upload web multi-file e import da cartella controllata. V1: stampante virtuale Windows con anteprima, selezione condominio/esercizio, retry e fallback. La stampante virtuale è un canale d'ingresso, non il motore di estrazione.

Accettazione: Lo stesso documento elaborato via upload e stampa produce oggetti canonici equivalenti; upload ripetuto è idempotente.

A03 Modello canonico e provenienza | MVP

Entità minime: Tenant, Condominio, Esercizio, Bilancio, Voce, Movimento, Documento, Fornitore, Componente, Categoria, Associazione, Analisi, Benchmark, Utente, AccessGrant, AuditEvent. Ogni campo estratto conserva file, pagina, coordinate/riga XML, metodo e versione.

Accettazione: Tracciare da una voce del PDF arricchito al file origine e alle evidenze del dato senza usare ID del gestionale.

A04 Pipeline asincrona e API | MVP

Job ripetibili e versionati: ingest, parse, match, review, analyse, publish. Stati e motivi di errore visibili. API proprie per upload, stato job, lista associazioni, conferma/rettifica, generazione e accesso; webhook firmati per completamento.

Accettazione: Il riavvio di un job non duplica voci, documenti o QR; un fallimento consente retry selettivo.

3. Acquisizione, estrazione e collegamento delle spese

D01 Integrità dei sorgenti | MVP

Conservare l'originale immutabile e impronte SHA-256 di ogni file; creare derivati separati. Il PDF aumentato può aggiungere livelli grafici o pagine complementari, ma non sovrascrivere testi, importi, saldi, millesimi o totali.

Accettazione: Verifica automatica degli hash; confronto di importi e totali estratti prima/dopo; originale scaricabile con autorizzazione.

D02 Parsing PDF e OCR selettivo | MVP

Estrarre testo e coordinate da PDF digitali; rilevare tabelle, intestazioni, colonne e riporti; OCR solo per pagine scansite o illeggibili. Versionare i parser e consentire correzione di campo e di riga. Evitare un solo template rigido per Danea.

Accettazione: Corpus pilota di layout differenti: ogni valore acquisito mostra sorgente; voci non interpretabili vanno in coda di revisione, mai pubblicate come certe.

D03 Import FatturaPA e allegati | MVP

Leggere XML e metadati rilevanti: identificativi, data, cedente, imponibili, IVA, ritenute, totale, linee, quantità, unità, descrizioni e allegati; validare formato e duplicati. Mantenere PDF fattura come fallback.

Accettazione: Import riuscito per esempi con più aliquote, note di credito, ritenute e più righe; nessuna somma desunta tacitamente da campi mancanti.

D04 Riconciliazione probabilistica | MVP

Generare candidati voce-fattura tramite importo, date, numero, fornitore, testo, categoria e periodo; gestire uno-a-molti, molti-a-uno, note di credito, acconti e fatture con importi simili. Lo score dev'essere calibrato su casi verificati, non un numero arbitrario.

Accettazione: Auto-associare solo oltre una soglia tarata sul pilot; mostrare evidenze e alternative; ogni caso ambiguo richiede conferma prima della pubblicazione.

D05 Revisione e apprendimento | MVP

Console di eccezioni con conferma, rifiuto, split, merge e rettifica; registrare attore, data, vecchio/nuovo valore e motivo. Conservare correzioni come ground truth soggetto a controllo qualità.

Accettazione: Una rettifica aggiorna le schede interessate e lascia visibile lo storico; il set di test di regressione include i casi corretti.

4. Documento arricchito, QR e accesso

P01 Riferimenti puntuali nel bilancio | MVP

Collegare una singola voce o gruppo di voci alle schede spesa. Se il margine non permette l'overlay, usare richiamo numerato e pagina indice aggiuntiva; mantenere leggibile il documento. Segnalare voci senza evidenza.

Accettazione: Ogni richiamo punta alla voce giusta; verifica visiva su campioni di impaginazioni; nessun totale o testo originario coperto.

P02 QR opaco e resolver | MVP

Nel QR inserire solo URL con identificatore casuale opaco; nessun importo, nome, dati personali o URL diretto del file. L'identificatore è revocabile e risolto lato server. Distinguere identificatore stampato da sessione/accesso a durata breve.

Accettazione: Scansione anonima non rivela fattura, metadati sensibili né analisi; revoca accesso efficace senza ristampare il bilancio.

P03 Scheda intelligente della spesa | MVP

Mostrare descrizione semplice, importo e periodo, fornitore se autorizzato, documento originale, relazione con riga del bilancio, spiegazione IA, eventuale benchmark e fattori di contesto, incertezza, versione e stato di revisione.

Accettazione: Un campo mancante è etichettato come non disponibile; ogni affermazione fattuale importante rinvia a una fonte o a un calcolo.

P04 Ruoli e separazione tenant | MVP

RBAC/ABAC per amministratore, collaboratore, condòmino, revisore e operatore Domuscan; autorizzazioni per condominio e documento, non solo per URL. Opzioni di autenticazione adeguate alla distribuzione (account/OTP), audit, revoca e scadenza.

Accettazione: Test automatici impediscono accessi tra tenant e tra condomìni non autorizzati, anche alterando ID nelle richieste.

P05 Marchio e comunicazione corretta | V1

Apporre "Bilancio intelligente Domuscan" e un collegamento esplicativo ove concordato. Usare "verificato" soltanto con definizione precisa di controlli effettuati: non implica certificazione legale o attestazione dell'assenza di errori.

Accettazione: Anteprima approvata prima della pubblicazione; spiegazione del badge accessibile e coerente con verifiche realmente eseguite.

5. Tassonomia, benchmark e IA

I01 Tassonomia specialistica versionata | MVP

Categorie e sottocategorie condominiali, componenti elementari e attributi: area, urgenza, quantità, unità, caratteristiche tecniche, dimensione edificio, accessibilità e tempo. Mapping dai conti del gestionale al modello Domuscan senza perdere il conto originale.

Accettazione: Ogni classificazione espone categoria, versione e fonte; le correzioni non alterano le analisi storiche senza nuova versione.

I02 Normalizzazione componenti | V1

Quando la fattura lo permette, separare manodopera, materiali, diritto di chiamata, noleggio, smaltimento, lavori, imposte e oneri. Salvare quantità e unità con qualità dell'estrazione e dati mancanti espliciti.

Accettazione: Non usare il totale fattura come prezzo unitario quando mancano quantità o perimetro tecnico.

I03 Motore comparabili tracciabile | V1

Il confronto è calcolo separato dall'LLM. Selezionare casi pertinenti per categoria, area, data, quantità e variabili tecniche; normalizzare unità, imposte e perimetro. Definire prima del pilot regole per outlier, campione minimo, freschezza, ponderazione e intervallo. Conservare origine lecita, licenza/consenso, qualità e versione del dataset.

Accettazione: Ogni risultato riporta numero di casi eleggibili, periodo, copertura, intervallo, principali differenze e motivi di esclusione; se qualità insufficiente risponde "dati insufficienti".

I04 IA esplicativa vincolata | MVP

Il modello linguistico spiega fattura e risultato deterministico usando solo campi/evidenze forniti; distingue fatti, ipotesi e incertezze; non inventa comparabili, prezzi o giudizi categorici. Versionare prompt, modello e input; filtri contro prompt injection nei documenti.

Accettazione: Suite di domande con dati assenti, fatture discordanti e testo ostile: l'output non crea cifre non presenti e segnala i limiti.

I05 Benchmark riproducibile | V1

Salvare snapshot o riferimento immutabile dei comparabili, regole, tassonomia, data e versione del motore; distinguere risultato storico da eventuale ricalcolo successivo. Un audit interno deve ricostruire il responso.

Accettazione: Ricalcolo sulla stessa versione/snapshot produce lo stesso intervallo entro tolleranze dichiarate.

I06 Feedback validato | V1

Raccogliere correzioni di matching/categoria, giustificazioni degli scostamenti e richieste degli utenti con controlli di qualità; feedback di scansione utile per prodotto, non come prova di congruità. Nessun apprendimento diretto da input non verificati.

Accettazione: Solo correzioni approvate entrano nel dataset dei comparabili o nel test set.

I07 Anomalie storiche | Evoluzione

In fase successiva: duplicati, variazioni improvvise, ripetizione di interventi e concentrazione fornitori, con spiegazioni e revisione; etichette "da approfondire", mai accuse automatiche.

Accettazione: Ogni segnalazione include evidenze e tasso di falsi positivi misurato sul pilot.

6. Sicurezza, privacy e governance dei dati

S01 Protezione documentale | MVP

Cifratura in transito e a riposo, gestione separata delle chiavi, backup verificati, scansione malware, controllo accessi minimo, logging senza fatture o segreti, retention configurabile e cancellazione documentata. I token QR non autorizzano da soli la visione.

Accettazione: Test di autorizzazione, revoca e restore; riesame periodico dei privilegi e delle esposizioni.

S02 Dati per benchmark | MVP

Separare repository operativo e dataset analitico; minimizzare, de-identificare dove possibile e registrare base giuridica/permessi d'uso e limiti contrattuali per ogni fonte. La pseudonimizzazione non va presentata come anonimizzazione garantita.

Accettazione: Un cliente escluso dall'uso secondario dei dati non alimenta i comparabili; cancellazioni e limitazioni si propagano.

S03 Revisione giuridica e documentale | MVP

Prima della produzione definire ruoli privacy, informative, accordi con studi/fornitori, tempi di conservazione, eventuale valutazione d'impatto e trasferimenti a fornitori IA, con consulenza competente. Non usare le fatture per training di terzi senza autorizzazione appropriata.

Accettazione: Checklist legale e configurazione tenant approvate prima del pilot con dati reali.

S04 Audit e incidenti | MVP

Log append-only dei caricamenti, estrazioni, correzioni, pubblicazioni, accessi e ricalcoli; allarmi su anomalie di accesso; procedura per incidenti e responsabilità di risposta.

Accettazione: È possibile ricostruire chi ha pubblicato una scheda e quale documento/analisi era visibile in una data versione.

7. API, portabilità e misure di differenziazione

E01 API e adapter propri | V1

Endpoint versionati e documentati per upload documenti, job, associazioni, scheda, benchmark e PDF; webhook firmati, idempotenza e limiti. Adapter per gestionali sempre opzionali, testati contro il modello canonico.

Accettazione: Rimuovendo un adapter, import PDF/XML e pubblicazione funzionano senza modificare il core.

E02 Import massivo | V1

Caricare un esercizio con centinaia di file, deduplicare, mostrare avanzamento ed eccezioni; parallelismo limitato per tenant e coda ripristinabile. Dimensionare su misure reali del pilot.

Accettazione: Un lotto interrotto riparte senza doppioni; il tempo di lavoro umano per 100 fatture viene misurato.

E03 Metriche di qualità e rete | V1

Dashboard: studi/condomìni attivi, documenti, percentuale di associazioni auto-confermate e poi smentite, voci con evidenza, analisi con benchmark valido, accessi, errori, tempi e correzioni. Pubblicare solo numeri verificati e aggregati.

Accettazione: Ogni indicatore ha definizione e query riproducibile; niente numeri stimati come dati effettivi.

E04 Asset proprietari governati | V1

Documentare e versionare tassonomia, normalizzatori, metodologia dei benchmark, dataset di test e origine dei comparabili. Motore e dati sensibili restano lato server, con controlli accesso e piano di portabilità interna.

Accettazione: Il client locale non contiene comparabili o segreti; chiavi e credenziali possono essere ruotate.

8. Contratti dati ed esempi per l’implementazione

Schema minimo di una voce: ID interno, tenant_id, building_id, fiscal_year_id, source_document_id, page, bbox, raw_label, raw_amount, currency, normalized_category_id, category_version, extraction_confidence, match_status, linked_document_ids, created_at e version.

Schema minimo dell’analisi: expense_id, source_hashes, matched_evidence, component_items, benchmark_status, benchmark_snapshot_id, benchmark_range, sample_size, comparability_factors, confidence_basis, explanation_text, model_version, rules_version, reviewer_id, publication_status e timestamps.

Stati raccomandati: IMPORTED → PARSED → NEEDS_REVIEW / MATCHED → ANALYSED → APPROVED → PUBLISHED; stati di errore e revoca separati. Nessuna analisi con associazione ambigua deve essere pubblicata come certa.

9. Piano di consegna e gate decisionali

Fase	Consegna concreta	Gate di uscita
0 - Corpus pilota	Raccogliere bilanci e fatture autorizzati, diversificati per layout e gestionale; etichettare un campione indipendente.	Definizioni di categoria, accessi e metriche concordate.
1 - MVP documentale	Import PDF/XML, modello canonico, match/revisione, PDF con QR, resolver e scheda spiegata.	Test integrità, correttezza link e isolamento tenant superati.
2 - Motore prezzi	Normalizzazione, corpus comparabili autorizzato, regole statistiche e motivazioni verificabili.	Qualità del confronto misurata per categoria; fallback “dati insufficienti”.
3 - Scala e difesa	Stampante Windows, batch, API/adapter, feedback, benchmark più ampio, dashboard.	Adozione multi-gestionale e precisione stabile misurate su studi reali.

Ordine commerciale coerente con la difesa tecnica: validare l’uso su studi con gestionali diversi; misurare adozione e correzioni; costruire comparabili utilizzabili; proporre integrazioni a Danea e ad altri come servizio esterno. Nessun volume di dati o esito negoziale è garantito da questa architettura.

10. Test di accettazione richiesti al fornitore

ID	Scenario	Esito richiesto
T01	Bilancio PDF nativo, XML con più linee, PDF scansionato	Estrazione con provenienza, OCR quando serve e revisione dei campi incerti.
T02	Due fatture uguali per importo, nota di credito, pagamento parziale	Alternative visibili; nessuna associazione incerta pubblicata automaticamente.
T03	Bilancio con margini stretti e molte voci	Riferimenti leggibili o indice aggiuntivo; testo e valori originari invariati.
T04	QR fotografato da utente anonimo o di altro condominio	Nessuna esposizione di documento o dati sensibili; accesso autorizzato registrato.
T05	Benchmark con pochi casi o perimetri incompatibili	Risposta “dati insufficienti” e motivazione; niente cifra o confidenza inventata.
T06	Ripetizione import/job, correzione operatore, revoca	Nessun doppione; storico versionato e accesso revocato.
T07	API di ogni gestionale indisponibile	Flusso PDF/XML e pubblicazione operativi.

Metriche da concordare sul corpus pilota: accuratezza per estrazione campi e per categoria, precisione/recall del matching separando i casi auto-associati da quelli revisionati, tasso di schede con evidenza, tempo per lotto, errori di autorizzazione, copertura e stabilità dei benchmark. Le soglie vanno fissate dopo aver visto mix di documenti e ground truth; “0 accessi tra tenant” è un requisito assoluto.

11. Deliverable richiesti al programmatore

• Codice sorgente e istruzioni riproducibili per installazione, migrazioni, backup, restore e deploy; elenco componenti esterni e relativi costi.

• Diagramma architetturale, schema dati e contratti API con esempi; documentazione del flusso di autorizzazione QR.

• Corpus di test anonimizzato o sintetico con ground truth, risultati misurati e casi di errore noti; report visuale su PDF generati.

• Manuale operatore per revisioni, pubblicazione, revoca e rettifica; audit di versione dei risultati.

• Piano di osservabilità, incidenti, retention e trattamento dei dati da approvare con il responsabile privacy.

Criterio di selezione del fornitore: chiedere una demo su un bilancio reale autorizzato e fatture pertinenti, compreso un caso ambiguo. Valutare la qualità della riconciliazione e delle evidenze prima della resa grafica del QR.

12. Decisioni da fissare nel preventivo

Decisione	Richiesta al fornitore
Stack e proprietà	Indicare linguaggi, servizi cloud, modelli IA e dipendenze con licenze; consegnare repository e credenziali alla società committente.
Campione iniziale	Proporre quantità e varietà di bilanci/fatture per addestrare le regole e misurare parsing e riconciliazione.
Qualità di pubblicazione	Definire soglie e workflow di revisione per associazioni, classificazioni e confronti; consentire configurazione per tenant.
Costo unitario	Stimare costi per fattura, pagina OCR, chiamata IA, storage e scansione; separare costi fissi e variabili.
Tempi e milestone	Preventivare per fasi 0-3, con demo, test e pagamento legati ai gate di uscita del capitolo 9.

Preparazione al “Domuscan Price Graph”. Conservare relazioni normalizzate fra categoria, componente, quantità, zona, periodo e caratteristiche tecniche. È sufficiente uno schema relazionale indicizzato nell’MVP: introdurre un database a grafo solo se i casi d’uso e le misure di prestazione lo giustificano. Non aggregare fornitori e interventi in modo che possano essere reidentificati impropriamente.

Prova della differenziazione. Il fornitore deve mostrare la stessa pipeline su documenti di almeno due origini diverse, con riconciliazione e spiegazione coerenti. La vera verifica competitiva successiva è la crescita di comparabili leciti e di correzioni validate, unita a utilizzo ricorrente degli studi e riconoscimento del marchio da parte degli utenti.

Nota: questo documento è una specifica di prodotto e implementazione, non una garanzia di esclusiva, una certificazione legale o una valutazione di congruità già disponibile.