
















Documento Tecnico FRAUDE – Classificazione e Post-Processi Fingerprint + Layer ISWC→ISRC

1. Modulo Fingerprint – Plagi / AI

Scopo

Primo livello di classificazione dei match individuati dal fingerprint. Produce una matrice di rischio che incrocia corrispondenza Titolo, Artista ed esito del match FP.

Titolo Match	Esecutore Match	FP Match	Rischio
 No	 No	Alto	 Alto
 Sì	 No	Alto	 Medio
 Sì	 No	Basso	 Medio-Basso
 Sì	 Sì	Basso	 Basso
 Sì	 Sì	Alto	 Nullo

Miglioramenti:

- Normalizzazione robusta titoli e artisti (accenti, feat., parentesi, ecc.)
- Similarità con rapidfuzz, soglie parametriche (Titolo ≥ 90 , Artista ≥ 85)
- FP “Alto/Basso” calcolato dinamicamente (percentili corpus)
- Output CSV standardizzato compatibile con ISRC e PD Disambiguation

2. Post-processo ISRC Disambiguation

Scopo

Disambiguare casi in cui lo stesso brano appare con più ISRC.

Operazioni

- Input: output Fingerprint + metadati ISRC
- Calcolo Δt , Δf , Δdur , overlap
- Classificazione: Same master / Remaster / Live / Sample_or_transform / Duplicato ammesso

Output

Mappa equivalenze/anomalie ISRC con indice di affidabilità.

Applicazioni autonome

- Audit cataloghi discografici
- Controllo metadati per distribuzioni digitali

3. Post-processo Public Domain Disambiguation

Scopo

Evitare falsi PD in cluster di brani con titolo/artista uguali ma ISRC multipli.

Operazioni

- Input: cluster titolo/artista/ISRC
- Verifica anno di prima pubblicazione
- Regole su Δ dur e pattern overlap
- Decisione: PD escluso / probabile / da verificare

Output

Tabella con stato finale per ciascun cluster.

Applicazioni autonome

- Gestione diritti in scadenza
- Contenziosi legali e auditing collecting

4. Pipeline integrata FRAUDE

Fingerprint (Plagi/AI) → ISRC Disambiguation → Public Domain Disambiguation → Policy & Report → Output validato

Layer di Policy & Report:

- Unificazione ISRC ammessi
- Sospensione riparto per PD probabile
- Flag per revisione manuale
- Report/API con changelog decisionale

5. Punti di forza

- Ogni modulo ha vita autonoma e applicazioni specifiche
- Integrati, producono un flusso robusto che riduce falsi positivi e anomalie
- Compatibilità piena con collecting, editori e audit legali

Addendum – Layer ISWC → ISRC

0) Posizionamento nella pipeline

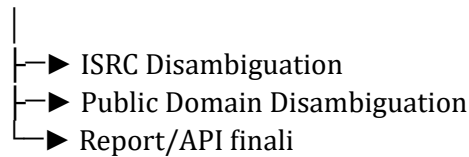
[Ingest] Work (ISWC) —

└─► [Enrichment & Linking ISWC↔ISRC] —► [Fraud/Anomaly]

[Ingest] Recording —

[Ingest] Usage/Play —————►

[Fingerprint → Stage1 Rischio]



1) Modello dati

- work: iswc, titolo, autori (IPI/CAE), publishers, creation_year, work_confidence
- recording: isrc, titolo, performer, label, release_date, duration, fingerprint_id, rec_confidence
- work_recording_link: iswc, isrc, link_method, link_confidence, evidence[], last_verified_at
- estensioni: indici e versionamento storico

2) Linking & disambiguazione

Ordine di precedenza:

1. Mapping diretto ufficiale (conf=1.0)
2. Match editoriale esatto (titolo fuzzy \geq 0.90, writers match, durata \pm 2%) → conf 0.85–0.95
3. Match probabilistico (audio+testo) → conf 0.65–0.85
4. Match manuale assistito

Feature & pesi:

- +0.40 fingerprint robusto
 - +0.25 writer match (IPI/CAE)
 - +0.15 durata compatibile
 - +0.10 titolo fuzzy >0.90
 - +0.05 coerenza label/publisher
 - 0.20 incongruenze (anni/territori)
- Soglie: \geq 0.85 auto-link; 0.65–0.85 review; <0.65 no-link

3) Regole FRAUDE (alert)

- W-MANY-ISRC: troppi ISRC per stesso ISWC
- R-META-MISMATCH: crediti incoerenti
- R-DUP-VAR: duplicazioni quasi identiche
- R-AI-SUSPECT: sospetto AI/voice cloning
- U-ANOM-BURST: spike usage anomalo
- W-NO-MAP: usage senza link ISWC

4) Report & API

Report:

- Work View: ISRC collegati, usage, alert
- Recording View: ISRC+link ISWC+evidenze
- Usage & Compliance: eventi timestamped

API:

- GET /work/{iswc}
- GET /recording/{isrc}

- POST /link/propose
- GET /alerts

5) Pseudocodice

```

for rec in recordings:
    candidates = find_work_candidates(rec.recording_title, rec.performers)
    for w in candidates:
        s = 0.0; ev = []
        if fp_is_robust(rec): s += 0.40; ev.append('fp_match')
        if writers_overlap(w, rec): s += 0.25; ev.append('writer_ipi_match')
        if duration_close(w, rec, 0.02): s += 0.15; ev.append('duration_ok')
        if title_fuzzy(rec, w) > 0.90: s += 0.10; ev.append('title_fuzzy')
        if label_publisher_hint(w, rec): s += 0.05; ev.append('label_publisher_ok')
        if meta_incongruent(w, rec): s -= 0.20; ev.append('meta_incongruent')

        if s >= 0.85: link(w.iswc, rec.isrc, 'probabilistic', s, ev)
        elif s >= 0.65: queue_review(w.iswc, rec.isrc, 'probabilistic', s, ev)

```

6. Orchestrazione & Deduplica

Per evitare duplicazioni di risultati e conflitti tra moduli (Fingerprint, ISRC Disambiguation, Public Domain Disambiguation, ISWC↔ISRC), il sistema segue un grafo decisionale con precedenze chiare.

Entità canoniche:

- Work = ISWC
- Recording = ISRC
- Acoustic Instance = Fingerprint ID
- Usage Event = play rilevato

Precedenze:

1. ISWC↔ISRC diretto ufficiale (CISAC/collecting) ha priorità assoluta.
2. Fingerprint è la verità primaria per associare l'audio a un ISRC candidato.
3. ISRC Disambiguation lavora solo sul set di ISRC coerenti col Work e con i candidati fingerprint.
4. PD Disambiguation si applica solo se:
 - non c'è un link ISWC affidabile, oppure
 - l'opera è in PD e bisogna distinguere le registrazioni valide.

Riuso dei segnali:

- I risultati Fingerprint alimentano direttamente ISRC Disambiguation e PD Disambiguation.
- ISWC↔ISRC fornisce vincoli per limitare i candidati ISRC.

Output consolidato:

Per ogni coppia (Acoustic Instance, ISRC) viene prodotto un unico esito con i seguenti campi:

- fp_conf, fp_overlap, title_match, artist_match, risk_stage1
- work_iswc, work_link_conf
- isrc_class, isrc_class_conf
- pd_status (excluded/probable/review)
- decision_conf_final (0-1)
- alerts[] (chiavi univoche)
- evidence[] (regole/feature applicate)

Alert chiave:

- R-DUP-VAR:{isrcA}:{isrcB}
- W-MANY-ISRC:{iswc}:{window}
- W-NO-MAP:{isrc}
- PD-FALSE-CLAIM:{cluster_id}

Gating rules:

- Se work_link_conf ≥ 0.85 → ISRC Disambiguation si limita agli ISRC collegati a quel Work.
- Se nessun link ISWC affidabile → ISRC Disambiguation valuta l'intero set di candidati fingerprint.
- PD Disambiguation si attiva solo quando l'opera è in PD o in assenza di link ISWC.

In questo modo si evita la duplicazione dei risultati e si ottiene un unico output consolidato, con piena tracciabilità delle decisioni e degli alert generati.

7. Orchestrazione dettagliata (Fingerprint → ISRC → PD → ISWC)

Ottima domanda. La chiave è orchestrare i tre pezzi (Fingerprint → ISRC Disambig. → PD Disambig. → ISWC↔ISRC) in un unico grafo di decisione, con:

- stesso feature store (riuso dei segnali, niente ricalcoli)
- precedenze chiare (cosa "vince" su cosa)
- ID/chiavi di alert per non generare duplicati.

Ecco come lo facciamo in pratica.

1) Entità canoniche (niente doppioni)

Work = ISWC

Recording = ISRC

Acoustic Instance = Fingerprint ID (hash/ID riconoscimento)

Usage Event = play rilevato (con fingerprint_confidence, clip_offset/duration)

Tutti i moduli leggono/scrivono su queste 4 entità. Così il risultato fingerprint è unico (Acoustic Instance), e viene solo riusato dai moduli dopo, mai ricalcolato.

2) Precedenze (regole d'oro)

1. ISWC↔ISRC diretto ufficiale (collecting/editori/CISAC) → verità primaria per link opera↔registrazione.

2. Fingerprint → verità primaria per associare audio a ISRC.
3. ISRC Disambig. usa (1) e (2) come vincoli.
4. PD Disambig. entra solo se manca un link ISWC affidabile oppure se l'opera è in PD.

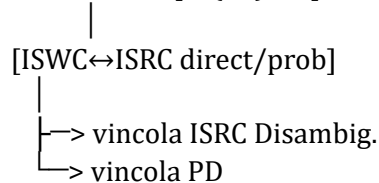
Effetto: ISWC riduce lo spazio decisionale, la fingerprint ancora l'audio a ISRC candidati, PD si applica dopo aver capito l'ISRC corretto e solo dove serve.

3) Riutilizzo dei segnali (feature store condiviso)

- Fingerprint → produce conf, overlap, Δt , Δf , duration_sec, title_norm, artist_norm, risk.
- ISRC Disambig. → riusa questi segnali e aggiunge la logica di classe + link_confidence ISWC ↔ ISRC.
- PD Disambig. → riusa release_date/recording_year, durata, cluster titolo/artista normalizzati, link ISWC.

4) Decision graph (semplice e idempotente)

[Acoustic Instance] --(FP)--> [ISRC candidates]



Gating rules:

- Se work_link_conf ≥ 0.85 → ISRC Disambig. limita ai soli ISRC collegati.
- Se nessun link ISWC → ISRC Disambig. valuta tutti i candidati fingerprint.
- PD Disambig. non si attiva se opera non è in PD; se opera è in PD valuta registrazioni conformi.

5) Unificazione degli score (niente doppio verdetto)

Un solo oggetto esito per ogni (Acoustic Instance, ISRC):

- fp_score
- isrc_class + isrc_class_conf
- work_link_conf
- pd_status

Calcolo decision_conf_final con pesi e precedenze:

- work_link_conf ≥ 0.85 → vincola, altri decadono.
- fp_score robusto + isrc_class same_master/remaster → alta confidenza.
- pd_status = excluded con opera non-PD → blocca claim PD.
- pd_status = probable con opera PD → sospendi riparto.

6) Niente duplicati di alert (chiavi univoche)

Chiavi deterministiche:

- R-DUP-VAR:{isrcA}:{isrcB}
- W-MANY-ISRC:{iswc}:{window}
- W-NO-MAP:{isrc}

- PD-FALSE-CLAIM:{cluster_id}

Il rule engine non ricrea alert già aperti.

7) Quando evitiamo un modulo

- ISRC Disambig. si salta se un solo ISRC candidato ha fp_score robusto e link ISWC affidabile.
- PD Disambig. si salta se opera non è in PD o se claim PD già escluso.

8) Schema operativo (checklist dev)

1. Scrivere i risultati FP in acoustic_instance_link.
2. Applicare resolver ISWC↔ISRC.
3. Eseguire ISRC Disambig. solo sul set vincolato.
4. Eseguire PD Disambig. se necessario.
5. Calcolare decision_conf_final + deduplicare alert.
6. Pubblicare output unico per (Acoustic Instance, ISRC).

9) Regole anti-conflitto

- FP altissima su due ISRC → scegli quello con link ISWC; l'altro diventa R-DUP-VAR.
- ISWC_A linka ISRC_x ma FP indica ISRC_y → cross-work collision = R-META-MISMATCH + review.
- Claim PD su registrazione recente di opera in PD → PD-RESTRICK (nuovo diritto fonografico).

10) Output consolidato (campi chiave)

Per ogni (acoustic_instance_id, isrc):

- fp_conf, fp_overlap, title_match, artist_match, risk_stage1
- work_iswc, work_link_conf
- isrc_class, isrc_class_conf
- pd_status
- decision_conf_final (0-1)
- alerts[]
- evidence[]