DOT: CC0 & Public-Domain Audio Resources

A strict-ingestion resource map for instruments, game SFX, music, MIDI, vocals, loops and impulse responses

Policy target: only explicitly verified CC0 1.0 or public-domain assets enter DOT's trusted production corpus.

Tier 1: strongest matches for DOT

1. Meta Open Sound - CC0 instruments

A strong fit for an AI agent because its repository explicitly addresses AI agents. The instrument collections are CC0 1.0/public domain and include an 88-key upright piano, nylon-string guitar, and steel-string acoustic guitar, with multiple dynamic or velocity layers, WAV/OGG audio, SFZ sampler mappings, and instrument design specifications.

Important restriction under DOT's policy: the Oculus Audio Pack in the same repository is CC BY 4.0, not CC0. DOT should ingest the CC0 instruments and reject that pack.

DOT priority: VERY HIGH

2. Open Game SFX Index - CC0 game sound warehouse

This is particularly suited to autonomous DOT operation. The project describes an AI-searchable collection of 1,451 CC0/public-domain game effects sourced from Kenney and CC0-verified OpenGameArt packs.

- Categories include UI/interface, impacts, sci-fi, digital/retro, RPG, casino, jingles, voiceover, footsteps, GUI, zombies, battle sounds, and coins/level-up sounds.
- Machine-readable index.json includes category, tags, use cases, mood, duration, SHA-256, download URLs, license, and source.

DOT priority: EXTREMELY HIGH

3. OpenGameArt CC0 Sound Effects

Verified CC0 collections include creatures, RPG effects, sci-fi, water/splash/slime, metal/wood, hits, interface beeps, footsteps, ambience, spell effects, retro/8-bit effects, and loops. These collections are useful as an upstream expansion source.

DOT priority: HIGH

CC0 music and reference material

4. OpenGameArt CC0 Music

Dedicated CC0 music collections cover game/background styles including battle, space, fantasy, orchestral, chiptune, ambience, dungeon, action, menu, and electronic material. Individual entries may explicitly declare CC0 or public-domain status.

DOT can use this material not only as finished assets but also as reference material for structural, spectral, rhythmic, and production analysis.

5. CC0-1.0 Music corpus

This GitHub corpus describes itself as a collection of public-domain/CC0 music gathered from multiple sources for uses including procedural music analysis and ethical audio-driven ML research. It reports roughly 7,000 songs and about 40 GB of material.

Caution: the repository also acknowledges that tracks may potentially have been incorrectly categorized and has a takedown process. DOT should therefore validate provenance and licensing of individual tracks before promoting them into the trusted corpus.

Loops, vocals, FX and MIDI

6. Producer Space - CC0 sample packs

The site states that its packs are CC0/public domain. Material described includes vocals, spoken word, vocal percussion, computer and sci-fi FX, piano MIDI, and pop-progression MIDI. It also states that commercial use is allowed without attribution.

DOT priority: HIGH

7. CC0-midis

A GitHub repository containing MIDI material under CC0 1.0. MIDI is useful to DOT because it provides direct symbolic information about pitch, velocity, duration, timing, polyphony, chords, melody, and rhythm without needing audio transcription first.

8. Imperfect Piano Performances

A CC0 1.0 repository of repeated piano practice performances and MIDI showing progression from earlier attempts toward more polished performances. The repository explicitly states that its files can be used for commercial projects and AI training.

This is potentially valuable for studying changes associated with performance refinement rather than studying only unrelated finished recordings.

Discovery-only MIDI indexes

A repository such as Awesome MIDI Sources can be useful for discovering third-party collections, but the license of the index itself does not establish the license of every external MIDI file it links to. DOT should treat such indexes as discovery-only until each underlying asset passes license verification.

Impulse responses

9. Verified CC0 impulse responses

A bundled collection was found containing ten impulse responses derived from source sounds identified as CC0 1.0. Listed spaces include a studio room, attic, concert hall, cathedral, parking structure, tunnel, courtyard, and shower. This provides a starting set for DOT's convolution and reverb experiments.

Not acceptable under the strict rule

- Conner's Impulse Response Library is MIT licensed, not CC0/public domain. Reject from the strict corpus.
- Sampled contains instruments, drums, loops, and impulse responses with per-instrument licensing. Treat it as discovery-only until an individual collection is verified CC0/public domain.

DOT trust model

Instead of a simple allowed/denied flag, use four trust classes:

Class	Meaning	DOT action
A	CC0-1.0 explicitly verified	Automatic production use
B	Explicit public-domain dedication	Automatic production use
C	Unrestricted-use statement, not explicitly CC0/public domain	Quarantine

Class	Meaning	DOT action
D	CC-BY / GPL / NC / SA / royalty-free / unknown / ambiguous	Reject

Under the strict policy, only classes A and B enter DOT's working sample library.

Immutable provenance record

Every accepted download should receive a manifest that preserves license evidence, origin, retrieval data, and a cryptographic hash.

```
{
"asset_id": "dot-audio-000042",
"filename": "metal_impact.wav",
"rights": {
"license": "CC0-1.0",
"public_domain": true,
"verified": true
},
"provenance": {
"source": "OpenGameArt",
"source_url": "...",
"license_evidence": "...",
"retrieved": "...",
"sha256": "..."
},
"classification": {
"type": "sfx",
"category": "impact",
"tags": ["metal", "heavy", "short"]
}
}
```

DOT should never overwrite an original accepted asset. Originals and provenance remain immutable; transformed copies belong in a derived-assets area.

```
/original/
original.wav
provenance.json

/derived/
dark.wav
pitched.wav
reversed.wav
impact-layer.wav
```

Recommended DOT ingestion architecture

```
CC0 MUSIC + CC0 INSTRUMENTS + CC0 GAME SFX
+ CC0 FOLEY + CC0 VOCALS + CC0 MIDI + CC0 IRs
|
v
DOT INGESTION
|
license verification
provenance + hashes
DSP analysis
semantic description
embeddings + categories
|
v
```

```
DOT AUDIO MEMORY
|
SEARCH / ANALYZE / COMPARE / COMPOSE / DESIGN / PROCESS / MASTER
```

Hard agent rule

> Never infer licensing from the hosting site, filename, repository license, surrounding collection, or a "free" designation. An audio asset enters the trusted corpus only when DOT has asset-level or clearly applicable collection-level evidence establishing CC0 1.0 or public-domain status. Otherwise quarantine or reject it.

Source notes from the research response

- Meta Open Sound - GitHub (CC0 instruments; Oculus Audio Pack separately CC BY 4.0)
- Open Game SFX Index - GitHub (AI-searchable CC0 game SFX index)
- OpenGameArt - CC0 Sound Effects and CC0 Music collections
- Producer Space - CC0 public-domain sample packs
- CC0-1.0 Music - GitHub corpus
- CC0-midis - GitHub
- Imperfect Piano Performances - GitHub
- CC0 bundled impulse responses in cendance

This document reproduces and formats the preceding Copilot response. License status should still be validated at ingestion time because upstream content and metadata can change.