What each set makes possible.

Every set here answers one question: what can you build from it? The paired production sets come first, because a song with its project beside it can become almost any dataset you need. The other sets follow, each with what it gives you and what it leaves out.

The most complete set

Paired production sets

One song, everything that made it: the producer's project, every lane printed from it, the MIDI it played, the unmastered mix and the mastered master.

3,874	about 8,464 h	24	16,777,215
songs, each with its project	of audio, lanes and masters added up	lanes in a typical song	lane combinations in that one song

1 Any stem set you want

Every lane is its own file, and the lanes sum to the mix. So you choose the split: four stems, six, the drums against the rest, one lead alone. A song with 24 lanes gives 16,777,215 different combinations of lanes. Each one is a mix you can sum, and the lanes you left out are its exact answer. Eight songs in ten have 12 to 45 lanes. At 20 lanes a song passes one million combinations.

2 New versions, printed from the project *

Our stem printing tool runs as an MCP server. Your agent or your own scripts open a project in its DAW and print the versions you need: with and without effects, with a send, with one thing changed, on one song or on all of them. One set of projects becomes many datasets, each made for one question. The tool drives the DAW on your own machines, so a project prints in full where its plugins are installed.

3 The project, read as data

Each project is parsed into one JSON schema: tracks, plugin chains, volume, pan, sends, buses, sidechain, clips, notes, automation and sections. A model can learn the decisions, not only the sound they made.

4 Notes beside the sound

3,764 of the 3,874 songs carry the MIDI the session played, 36,568 files in all. Notes on one side, the printed lane on the other: the pair a transcription or a synthesis model needs.

5 From lanes to master

The lanes, the unmastered mix they sum to, and a mastered master on 1,655 songs. Three stages of one song, for mixing and mastering models.

Side by side

Set	Project file	Every lane as a file	MIDI	Masters	Any stem set	New prints with the tool *
Paired production sets	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Stems	–	**Yes**	In part	In part	**Yes**	–
Vocal and instrumental pairs	–	In part	–	–	–	–
DAW sessions	**Yes**	–	In part	–	–	–
Tutorial sessions	**Yes**	**Yes**	In part	**Yes**	**Yes**	–
MIDI files	–	–	**Yes**	–	–	–
Instrumentals	–	–	–	**Yes**	–	–
Stock music	–	–	–	**Yes**	–	–

In part: stems carry MIDI where the session had notes and ship the unmastered mix, not a mastered one. Vocal and instrumental pairs hold two parts per song. In DAW sessions and tutorial sessions the MIDI sits inside the project file.

* The stem printing tool comes with a minimum order of paired production sets. Ask us and we send the terms.

The other sets

Set	What it makes possible	Next to the paired sets
Stems 4,649 songs · about 10,323 h	• Any stem set you want, by the same sum: every lane is its own file and the lanes add up to the unmastered mix that ships beside them. • Separation, stem generation and instrument recognition. Each lane carries one of 53 instrument labels. • MIDI where the session had notes.	No project file. You get every lane as it was printed, but no new versions and no mixer or plugin data.
Vocal and instrumental pairs 2,006 pairs · about 222 h	• The vocal against the rest of the song: two aligned files from one session that sum to the mix. • Vocal separation, vocal removal and singing models.	Two parts per song, not every lane. No project and no MIDI.
DAW sessions 9,479 projects · about 692 h	• The set with the most songs: projects made in Ableton, Logic, FL Studio and Cubase. • Tracks, plugin chains, mixer state, clips, notes, automation and sections, parsed into one JSON schema. • Production assistants, mixing models and anything that learns how a track is built.	The project with its own audio and MIDI, but no lanes printed from it and no masters.
Tutorial sessions 393 episodes · about 231 h	• One producer building a track on camera, every word timed, from the first sound to the mixdown. • The finished Logic project, its stems and a flat and a mastered master for every episode. • Models that explain, teach or follow a production step by step.	A paired set with the reasoning on top, on a smaller number of songs, all by one producer.
MIDI files 291,223 files	• Chords, leads, basslines and drums as note data, tagged with source and tempo where known. • Symbolic music models.	Notes only, with no audio beside them.
Synth presets 167,471 files	• Patches for the synths producers use, as parameter data. • Sound matching and synth programming models.	Parameters only, not audio.
Loops and one-shots 460,951 files	• Drum hits, loops, vocal chops and effects, labelled with the producer's own names. • Sample generation, drum models and sound classification.	Short sounds on their own, not songs.
Instrumentals 3,351 tracks · about 184 h	• Full-length instrumental performances with tempo and key. • Full-mix training, and human-made reference audio for detection models.	Finished mixes. There are no parts to take apart.
Stock music 1,929 tracks · about 82 h	• Finished tracks written for licensing, full length, with tempo and key. • Full-mix training, and human-made reference audio for detection models.	Finished mixes. There are no parts to take apart.
Marketplace data 54,536 orders	• What sold, when, to which country, at which tempo and key, over fourteen years. No names, no emails. • Trend, demand and recommendation models.	Records, not audio.

Tell us what you are building. We answer with the set that fits and a quote. A sample follows before you sign.

ximaudio.com/request · Book a 30-minute call · mikas@ximaudio.com

Figures counted 25 Nov 2026 from the live catalogue, as listed per set at ximaudio.com/data. The same comparison with today's figures: ximaudio.com/compare. The number of lane combinations in a song with n lanes is 2 to the power n, less one. XIM Audio Inc., Vancouver, Canada.