

Media Research v1.0 (Tools · Syntagma Forschung)

A serverless browser tool for researching, editing and evidencing free media

Strand: RESEARCH — Tools · Syntagma Forschung · Vienna **Author:** Thomas Schüller · ORCID 0009-0003-9799-6747 · forschung@tschueller.com **Licence:** CC BY 4.0 (code and documentation) **Date:** August 2026 · **Version:** 1.0

Verification caveat. Statements about interfaces, access rules (CORS), key requirements and licence status of the integrated archives rest on a systematic survey carried out in July 2026. They are snapshots: operators change access rules, key requirements and terms of use without notice. Before productive use, and before any reuse of material, the status must be checked against the primary source. The tool's licence traffic light reflects archive metadata and does not replace legal assessment.

1. Subject

Medienrecherche is a bundle of four HTML files that run in the browser without a server, without installation and without transmitting content. The tool joins three working steps usually spread across separate applications:

1. **Search** across several open archives simultaneously, filtered by media type and licence tier.
2. **Editing** of retrieved audio (cutting, equalisation, pitch, export) plus transcription to MIDI and notation.
3. **Evidence** in the form of an exportable research record with a licence balance.

Two modes of use are addressed equally: journalistic research and media work on one side, artistic production on the other. Both rest on the same foundation — findable material with a clarified usage status and traceable provenance.

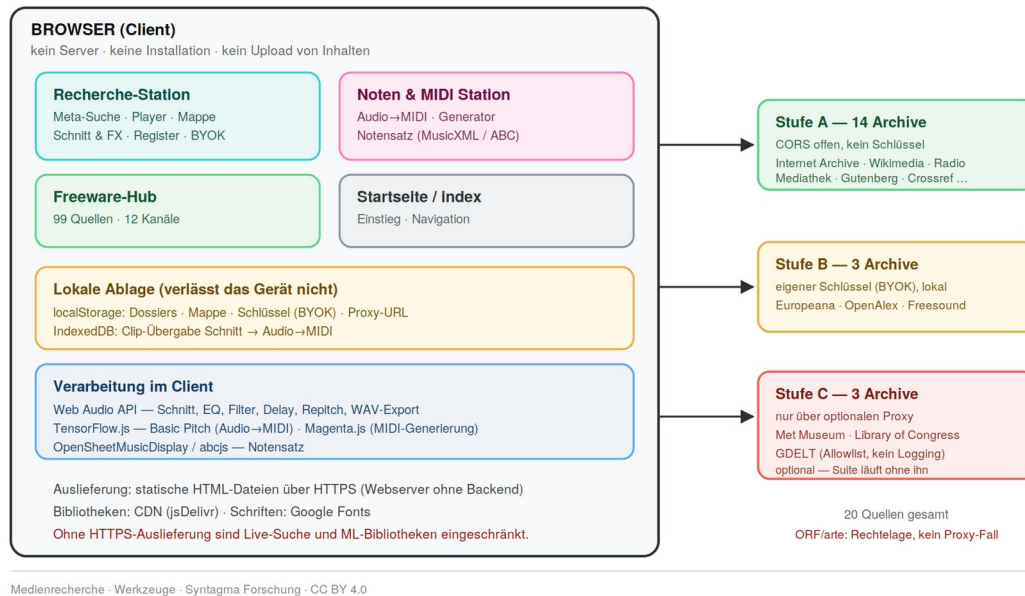
2. Context

Existing tools each cover part of the ground: image search engines with licence filters, sample libraries, notation programmes, verification tools for visual material. The transition between these steps is manual, and licence status is regularly lost along the way because it is not carried forward.

Medienrecherche addresses precisely this: licence status is carried from the result through to the export, and handover between search, editing and transcription happens inside one application. Which archives can be integrated at all follows a technical criterion — whether an interface is reachable from static browser code.

Abb. 1 — Serverlose Architektur / Serverless architecture

Medienrecherche v1.0 · Ausführung vollständig im Browser / executed entirely in the browser



3. Architecture

3.1 Execution model

The application consists of static files. There is no server-side data storage, no user accounts and no transfer of media files. Processing takes place in the browser via the Web Audio API and libraries executed as JavaScript on the client.

Only the user's working data is stored locally:

Store	Content	Scope
localStorage	dossiers, collection folder, own keys, proxy address	this browser
IndexedDB	staging area for clip handover to transcription	this browser

No synchronisation between devices takes place. Dossiers and the collection folder can be exported as files and read in elsewhere.

3.2 Source selection criterion

What governs integration is not the price of an interface but its reachability from the browser. Two properties decide: whether the operator permits cross-origin access, and whether a key is required. Three tiers follow.

Tier A — directly usable (14 archives). Cross-origin permitted, no key. These are available without any setup.

Archive	Region	Media type
Internet Archive	global	audio, video, text, image
Wikimedia Commons	global	image, audio, video
Art Institute Chicago	US	image (CC0)
Cleveland Museum of Art	US	image (CC0)
MediathekViewWeb (ARD, ZDF, ORF, arte et al.)	DACH	video
radio-browser	global	audio (streams)
Project Gutenberg	global	text
Wikiquote	global	text (quotation evidence)
LibriVox	global	audio (audiobook, stage text)
Chronicling America	US	text (historical press)
Open Library	global	text
Openverse	global	image, audio
Crossref	global	publications
Europe PMC	global	publications

Tier B — with your own key (3 archives). Europeana, OpenAlex and Freesound require registration. The key is stored locally and sent only to the respective source. OpenAlex has required a key since February 2026; access was previously open.

Tier C — only via optional relay (3 archives). Met Museum, Library of Congress and GDELT send no cross-origin clearance, or throttle anonymous access. An enclosed relay (Cloudflare Worker, around 100 lines) closes this gap. It works from a fixed target list, does not log and does not store. Without it the suite runs unchanged; the three archives then do not appear.

ORF-TVthek and arte remain outside. The obstacle there is not technical but legal — geoblocking and absent usage rights. A relay changes nothing about that, which is why both are listed in the register as references.

4. Licence handling

Every result receives one of three tiers, derived from the metadata of the supplying archive.

Abb. 2 — Lizenz-Ampel und Nachweiskette / License traffic light and evidence chain

Zuordnung aus Archiv-Metadaten · Prüfung am Original bleibt erforderlich



Nachweis je Treffer (Export: Markdown · HTML · JSON)

Titel	Urheber	Datum	Archiv	Lizenz	Link	Abrufdatum	Suchbegriff
Die Mappe sammelt Nachweise, keine Inhalte — Metadaten und Verweis, kein kopiertes Material. Export gruppiert nach Ampelstufe und weist Bilanz aus (frei / Namensnennung / nur ansehen).							

Die Ampel ist Orientierung aus Metadaten, keine Rechtsauskunft. Prüfung am Original vor jeder Verwendung.

Tier	Assignment	Use
Green	CC0, public domain	free, including commercial, no attribution required
Yellow	CC BY, CC BY-SA	with attribution; BY-SA requires share-alike
Red	broadcasting rights, press, unclear	citable as evidence, not reusable

Assignment is derived automatically from fields such as `edm:rights` (Europeana), `share_license_status` (Cleveland), `is_public_domain` (Art Institute) or `LicenseShortName` (Wikimedia). Where a field is missing or ambiguous, nothing is guessed; the entry is carried as unclear.

One point deserves emphasis because it is routinely underestimated: open access and free reuse diverge. Song MIDI files, commercial acapellas, song lyrics and broadcast archives are often retrievable without being usable. With MIDI files two rights may coexist — in the composition and in the arrangement. The tool carries such finds as red, hence as evidence rather than material.

5. Functional scope

5.1 Research station

The search queries all active archives simultaneously and shows the hit count per archive, so failures remain visible. Filters exist for media type and licence tier.

Dossiers store a question together with the selection of archives and can be passed on as a file, making a search repeatable.

The archive register lists 57 archives, including those the tool does not query. Its purpose is discoverability: an archive whose existence is unknown will not be searched. Each entry carries its tier (A, B, C), region, media type and licence profile.

Perspective comparison works through references to GDELT, Ground News, AllSides and Wikipedia; with the relay configured, GDELT appears directly in the result list. This allows coverage of a topic to be contrasted across language and country boundaries.

Verification per result via reverse image search (Google Lens, Yandex, Bing, TinEye) and a reference to the InVID/WeVerify toolkit.

Collection folder. Results are collected with title, creator, date, archive, licence, link, retrieval date and the search term used. Export is available as Markdown, HTML or JSON, grouped by licence tier and with a balance. What is collected is evidence, not content — metadata and reference, no copied material.

5.2 Editing

Waveform display with region selection, gain, fade in and out, three-band equalisation, cut-off and delay filters, and pitch change with tempo matching. Export as WAV.

Pitch and tempo are coupled, corresponding to tape machine or turntable behaviour. Decoupled stretching is not included; Audacity or Ardour serve that purpose.

5.3 Notation and MIDI

Audio to MIDI via Basic Pitch (TensorFlow.js, Apache 2.0), polyphonic, with adjustable onset threshold. **MIDI generation** via Magenta.js (MusicVAE, MelodyRNN) for melodic and percussion patterns. **Notation** via OpenSheetMusicDisplay for MusicXML and abcjs for ABC notation.

Converting MIDI to notation remains an approximation. MIDI knows neither time signature nor enharmonic spelling; mapping to note values is an interpretation. The result serves as a sketch. For notation of publication quality, MuseScore is preferable.

Abb. 3 — Arbeitskette / Workflow chain

Von der Fragestellung zur belegten Dokumentation · von Rohmaterial zur Notation



Übergabe zwischen den Strängen erfolgt lokal (IndexedDB); Inhalte verlassen das Gerät nicht.

6. Operation

The four HTML files, and optionally the relay file, belong in a shared directory; `index.html` forms the entry point. Links are relative.

	Served over HTTPS	opened locally (file://)
Archive search	full	restricted
Playback, editing, export	full	full
Collection, dossiers, register	full	full
Audio—MIDI, notation	full	unavailable

Full scope requires delivery via a web server. Browsers treat locally opened files more restrictively, and the libraries for transcription and notation are loaded from a delivery network.

7. Data protection

Content is not transmitted. Audio data is decoded, processed and exported locally. No analytics tools or tracking pixels are built in.

Outbound traffic is limited to: search queries to the respective archives, fonts from Google Fonts and programme libraries from jsDelivr. When the relay is used, requests to the three Tier C archives pass through it; it sees the search terms and does not store them.

Keys, dossiers and the collection folder remain in the browser.

8. Testing

Before publication the components were tested as far as possible without a browser and without access to the archives.

Test	Scope	Result
Interface parsing logic	20 adapters against reconstructed responses in real field structure	passed, 22 results parsed correctly
Failure behaviour	20 adapters against empty and sparse responses	one crash found and fixed
Central safeguard	malformed results filtered before display	passed
MIDI export	byte-level read-back: header, track length, variable-length encoding, timeline	structurally valid
WAV export	byte-level read-back: header, samples, clipping, channel interleaving	structurally valid

Test	Scope	Result
Notation conversion	scale, chords, rests, empty case	passed
File set	syntax, identifiers, navigation, style variables, cross-references	passed

The crash found concerned the radio query when a response did not contain a field of the expected form. In operation it would have interrupted the running search. Both the finding and the fix are noted here, since the testing would otherwise not be traceable.

What could not be tested is behaviour against the real archives. The test environment had no access to those services. Whether an archive responds on the day of use, and whether its access rules are unchanged, becomes apparent only in operation.

9. Limitations

1. **Source availability.** For Chronicling America, LibriVox and Openverse, browser access was noted as to be verified in the survey. If a source fails, this lies with its server; it can be switched off individually.
2. **Broadcast archives.** MediathekViewWeb supplies metadata and retrieval addresses. The content remains bound by broadcasting rights, may be regionally blocked and is removed after set periods.
3. **Quotations.** Wikiquote results appear as title, brief context and reference. The quotation text itself is not carried over, since quotations may be protected independently of the collection's licence.
4. **No synchronisation.** Without a server component, working data stays bound to the respective browser.
5. **Notation.** See section 5.3 — a sketch, not publication quality.
6. **Pitch and tempo** are coupled.

10. Reuse and extension

Code and documentation are under CC BY 4.0. A source is integrated by adding an object to the SOURCES array:

```
{id:"short", label:"Display name", rg:"Region", media:["audio"],
byok:false, on:true,
run: async q => {
  const d = await jget(`https://api.example.org/search?q=${enc(q)}`);
  return (d.results||[]).map(x => ({
    title:x.name, creator:x.author||"", date:x.year||"",
    type:"audio",          // audio | video | image | text | science
    license:"green",       // green | yellow | red
    licName:"CC0", source:"Example", region:"global",
    page:x.url, thumb:x.thumb||"", file:x.audio||null
  }));
}},
```

The prerequisite is an interface that permits cross-origin access and returns JSON. Check beforehand:

```
curl -I -H "Origin: https://example.com"  
"https://api.example.org/search?q=test"
```

The response header must carry Access-Control-Allow-Origin.

11. Free libraries used

Library	Purpose	Licence
Basic Pitch (Spotify)	audio to MIDI	Apache 2.0
Magenta.js (Google)	MIDI generation	Apache 2.0
OpenSheetMusicDisplay	MusicXML rendering	BSD
abcjs	ABC notation	MIT
TensorFlow.js	model execution	Apache 2.0

Web Audio API, Fetch and IndexedDB are part of the browser and require no library.

12. Competing interests

The author researches independently, without institutional affiliation, privately funded and without grants or third-party funding. Alongside this, the author operates commercial activities in manufacturing, audio and creative software, and event safety.

For this work a connection exists and is disclosed: the same tool is additionally published under the author's own commercial brand as a free tool. Both versions rest on the same source code. No revenue arises from this; an indirect benefit through increased brand visibility is possible and is not disputed. The archives and libraries addressed in the tool have no business relationship with the author.

As there is no institutional affiliation, there is likewise no institutional quality assurance. The methodological statements are designed to make independent scrutiny possible.

Full statement in INTERESSEN OFFENLEGUNG.md; use of AI-assisted tools in AI_DISCLOSURE.md.

13. Citation

Schüller, T. (2026). *Media Research v1.0 (Tools · Syntagma Forschung): A serverless browser tool for researching, editing and evidencing free media*. Zenodo.
[https://doi.org/\[insert DOI after upload\]](https://doi.org/[insert DOI after upload])

Published in the Tools · Syntagma Forschung. Reports of failing sources or missing archives are welcome, since interface reachability changes continuously.