Concept of the `sapupra` project

v0.50.0.98

Péter Felecan

assisted by Claude Opus 5.5 (Anthropic)

November 6, 2026

Contents

1 Introduction **2**

2 Motivation **2**

3 Goal **2**

4 The collection **3**

5 Name of the project **4**

6 Prior art **4**

7 Architecture **5**

8 Glossary **8**

8.1 Project terms 8

8.2 Sources 9

8.3 Tools 10

List of Tables

1 Closest existing free software against the requirements 4

List of Figures

1 Data flow of the first prototype, from the scan to the consultation 6

1 Introduction

`sapupra` is a simple management system for a personal library: the books, records, films and other media of one individual, counted in thousands. Its premise is that cataloguing a large collection stays a chore unless entering an item costs no more than pointing a smartphone at its barcode, and that consulting the catalogue needs nothing more than a web browser on the local network.

This document states the goal and the requirements, the result of the survey of existing free software that preceded any design, and the architecture adopted for a first prototype. Nothing is implemented yet; the design of its programs is in Design.pdf, how to install and use it in Manual.pdf; the open questions are listed in FutureResearch.pdf.

2 Motivation

A collection of thousands of items outgrows memory: the same book is bought twice, a record is searched for in the wrong room, a film lent years ago is forgotten. A catalogue answers these questions, but only if it exists, and it exists only if feeding it is fast.

The software that does this today (see Prior art, below) falls into two families. Desktop collection managers catalogue well and look up metadata from many sources, but they are interactive programs bound to one screen: an item is entered at the computer, not where the shelf is. Library systems and self-hosted web applications are reachable from anywhere, but they are heavy (an application server, a database server, often a container stack) for a need that is one person's. Neither offers the short path that this project is about: scan at the shelf, catalogue happens by itself.

3 Goal

The workflow:

Capture a barcode is read with a smartphone.

Transport what the barcode encodes is sent to an MQTT broker, on a machine of the local network running GNU/Linux or FreeBSD; an optional location can be attached, free form or itself a barcode.

Resolution a worker subscribed to the broker is notified, reads the code and searches the bibliographic information of the medium (book, CD, DVD, etc.) in sources on the Internet.

Storage the bibliographic information is stored in a light database.

Retrieval a simple, web based, information retrieval system consults the database.

Locality everything but the gathering of bibliographic information stays on the local network.

The guiding principle is lightness: a short and simple workflow, few moving parts, no fuss. Before designing anything, existing free software was surveyed for these or approaching features, to be adopted or complemented if adequate; the survey is in ResearchJournal.pdf and summarized below.

Not in the goal, unless decided otherwise later: lending and circulation, several users with rights, e-book files management (Calibre does that well), and any service reachable from outside the local network.

4 The collection

What the catalogue has to cope with, from a sample of one collection (ResearchJournal.pdf). Another collection will differ in its languages and its periods; the design does not depend on these examples.

Languages in the sampled collection, mostly French and English books, and a few in Hungarian, Romanian, Italian, Ancient Greek and Sanskrit; another collection has its own set. The language of a book does not tell where it is catalogued: a Sanskrit text may be published in India, Oxford or Paris.

Periods the ISBN exists since 1970 (the international standard ISO 2108; its British forerunner, the 9-digit SBN, since 1967), and each country joined it at its own date; it has 13 digits since

1. Barcodes on books spread from the second half of the 1980s.

So a book older than about 1970 has neither ISBN nor barcode, and one older than about the late 1980s often has an ISBN but no barcode (ResearchJournal.pdf). The dates of the countries of the sampled collection, as far as they were found (ResearchJournal.pdf):

United Kingdom 1967 (the SBN, until 1974);

United States 1968;

France the agency has assigned ISBNs since 1972;

Hungary the national office was set up in 1973, the books of the established publishers carry the ISBN from 1974 and those of the others from 1975;

Italy the ISBN may be printed as an EAN barcode since 1983, by agreement with the Italian authority (Chaumier and Dejean, 1988); the author's reading gives 1977 for the ISBN itself, which no source found here confirms;

Romania 1989, the year the system was introduced (the Romanian Wikipedia, in the author's quotation: "sistemul de numerotare standardizată a cărților – ISBN a fost introdus în anul 1989"); the national ISBN centre of the Biblioteca Națională coordinates it and its page says it has worked since 1996.

So about 1970 is the earliest date, and a country has its own: a French book is expected to carry an ISBN from about 1972, a Hungarian one from about 1974, an Italian one from about 1977 and a Romanian one only from about 1989 (the author's newest Romanian book without an ISBN is of 1969). A country may have been much later: the German Democratic Republic made the ISBN mandatory only in October 1986.

Barcodes on books most recent books carry their ISBN-13 as an EAN-13 barcode, sometimes followed by a 5-digit add-on (`90000` means "no price"). Some publishers print instead a product code (GTIN) that is not the ISBN, the ISBN being only in print on the back cover or the colophon.

Barcodes on records and films an EAN-13 or a UPC-A; a film may also carry an ISBN (some publishers of films had one assigned).

Consequence scanning covers most items, not all. Three other ways in are needed: typing an ISBN; finding a book without ISBN in the library catalogues by author and title, and copying the record found; entering the bibliographic information by hand.

5 Name of the project

The name is a romanized abbreviation of the Sanskrit phrase: **सरल पुस्तकालय प्रबन्धन प्रणाली** (saral pustakalaya prabandhan pranali), "Simple library management system"; in French, « Système simple de gestion de bibliothèque ».

6 Prior art

The sources and the details of the survey are in ResearchJournal.pdf. Table 1 sums up how the closest candidates meet the requirements.

Table 1: Closest existing free software against the requirements

Candidate	Kind	Lookup	Web	Light
Tellico 4.3	KDE desktop	many, ISBN and UPC/EAN	no	yes
GCstar 1.8.1	GTK desktop	site plugins	no	yes
BiblioteQ	Qt desktop	SRU, Z39.50, OL	no	yes

Continued on next page

Continued from previous page

Candidate	Kind	Lookup	Web	Light
Calibre	desktop, server	books only	yes	no
Koillection	PHP web app	none (own scrapers)	yes	no
BookLogr	Flask web app	Open Library	yes	no
Homebox	Go web app	none	yes	yes

None covers the workflow end to end. The desktop managers have the best lookups but no unattended entry path (Tellico's D-Bus interface drives a running instance of the program) and no web consultation; Calibre is centred on e-books and heavy; the web applications either have no bibliographic lookup or cover books only, and need a database server or a container stack. None receives scans from a smartphone, and none speaks MQTT.

The survey did find the parts that the workflow needs, each free and mature:

Binary Eye a free (MIT) Android barcode scanner, on F-Droid, with a continuous scanning mode and a built-in forwarding of each scan to a URL by HTTP (as a JSON object carrying the content, the symbology, the EAN add-on, a timestamp and a configurable device identifier); the body of the reply is shown on the phone. It does not speak MQTT.

Mosquitto the MQTT broker, in Debian and in FreeBSD's ports.

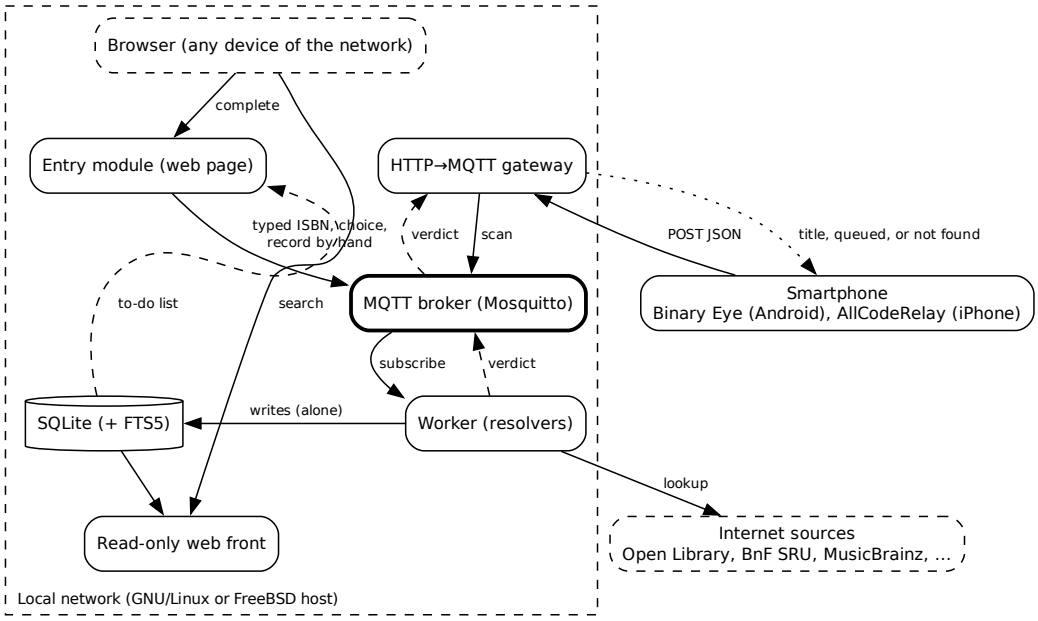
Bibliographic sources Open Library and the national libraries (the BnF and the German DNB by SRU; the Hungarian and Italian national catalogues and the Library of Congress by Z39.50) for books; MusicBrainz and Discogs (barcode search) for records; for films, library catalogues (which record the UPC of a film) and generic UPC/EAN databases, the weakest link.

SQLite with its FTS5 full-text index; Datasette is a ready-made read-only web front for an SQLite database, with search and facets.

7 Architecture

The architecture below is the one adopted for a first prototype. Its centre is the MQTT broker: every change to the catalogue goes through it, and every other part is replaceable. The worker is the only program that writes to the database; the entry module and the web front only read it. Figure 1 shows the data flow.

Figure 1: Data flow of the first prototype, from the scan to the consultation



Capture on Android, Binary Eye in continuous mode forwards every scan as a POST of a JSON object to a URL of the local network. On an iPhone or an iPad, AllCodeRelay does the same with a smaller JSON object; it is still to be tried.

Gateway since the phone applications speak HTTP and not MQTT, a small gateway receives the POST, recognizes which application sent it, and publishes the scan to the broker in one form (QoS 1, so that a scan survives a worker that is down). It then waits a few seconds for the worker's verdict and replies to the phone: the title if the item was identified; "queued" if the verdict did not come in time; "not found, set aside" with an HTTP error status if it was not identified. Binary Eye answers an error status with its error vibration and sound, so that a failure is felt and heard at the shelf without looking at the screen. The differences between scanning applications stop at the gateway.

Location set at the beginning of a batch, it becomes the default location of the scanning device: every item that device scans afterwards is placed there, until another location is set. It is set either by scanning a code, a label stuck on the shelf (a QR code or Code 128 with a reserved prefix) or a QR code holding a free text, or by entering it in the entry module for that device (the scanning applications have no way to type it). A location, scanned or entered, travels through the broker like a scan, and the worker keeps the current location of each device.

Worker subscribed to the scans, it first records each scan (time, device, location, code, symbology), so that nothing that was scanned can be lost, and only then classifies the code and tries the resolvers of that class in order, keeping the raw answer of each source (source record) beside the normalized fields:

- an EAN-13 starting 978 or 979 is an ISBN (979-0, an ISMN, is printed music); its registration group (2 for French-speaking countries, 0 and 1 for English-speaking ones, 963 and 615 for Hungary, 973 and 606 for Romania, 88 for Italy, 81 and 93 for India) says which national library to ask first;
- a typed ISBN-10 is turned into its ISBN-13;
- any other EAN-13, and a UPC-A, is a record, a film or a product that is not a book: the record and film sources are asked.

It then sets the status of the scan and publishes its verdict:

resolved the item is catalogued;

not found every source answered, none knows the code;

sources unavailable a source could not be reached or refused for its rate limit; retried later, automatically;

ambiguous several candidates; a person chooses;

not an ISBN a book whose barcode is a product code; its ISBN has to be typed.

A scan that is not resolved stays in the database with its status, its location, its time and the items scanned just before and just after it, so that the item can be found again on its shelf.

Entry module a page of the local network, opened from any browser, the phone's included. It lists the scans that need a hand (from the database, read-only) and offers, for each: typing its ISBN; searching the library catalogues by author and title, then choosing one of the records found (copy cataloguing); entering the bibliographic information by hand. It catalogues the items that have no code the same ways. Everything it produces (a typed ISBN, a choice, a record entered by hand) is published to the broker, tied to the scan it completes, and written by the worker. Scanning again an item that is waiting adds nothing to the list: the phone is answered that it is already set aside.

Database one SQLite file, with an FTS5 index on titles, people and subjects; written by the worker alone.

Retrieval a read-only web front on the local network: search, facets by medium and location, the item page. Decided in 0.18.0: a few pages of the project's own, in the entry module; Datasette stays an optional extra outside the build.

Implementation the gateway, the worker and the entry module are written in C++17 with Qt 6 (the non-graphical modules Core, Network and Sql), libmosquitto for MQTT and YAZ for the library catalogues served by Z39.50.

Later an export of the catalogue in Tellico's XML format.

8 Glossary

The one place the terms of the project are defined; the other documents point here. Code names are written between tildes, as `deviceId`. Where the world outside uses a different word for the same thing, the entry says so.

8.1 Project terms

Item one physical object of the collection: a book, a CD, a DVD. Two copies of the same edition are two items. Libraries say *copy* or *holding* (French: *exemplaire*).

Location where an item is: a free-form text or the content of a location label. Optional; the current location of a scanning device is the default of the items it scans.

Scan device a device that reads a barcode and transmits what it read to the broker, directly or through the gateway: today a phone with a scanning application that forwards its scans by HTTP (Binary Eye on Android). Each has an identifier, sent with its scans.

Scan one reading of a barcode by a scan device, with what the scanner reports (content, symbology, device, time); by extension, an ISBN typed in the entry module.

Status where a scan stands: resolved, not found, sources unavailable, ambiguous, or not an ISBN.

Entry module the web page of the local network where the scans that need attention are completed, and the items without code are catalogued.

Gateway the program that receives scans by HTTP from the phone and publishes them to the MQTT broker.

Worker the program, subscribed to the broker, that turns a scan into a catalogue entry.

Resolver the part of the worker that asks one source about one code.

Source record the answer of one source for one code, kept as received.

Edition the bibliographic description of one publication, shared by its copies: each copy is an item of that edition. Libraries say *manifestation*.

Input a message to the worker: a scan, a location, a typed ISBN, a search, a choice or a record entered by hand.

Receipt the worker's answer that an input is recorded in the database.

Verdict the worker's answer that settles an input: its status, and the title found or the reason.

8.2 Sources

EAN-13 the 13-digit product barcode (GTIN-13). On books it carries the ISBN-13 (prefixes 978 and 979), on printed music the ISMN (979-0).

GTIN Global Trade Item Number, the family of product numbers that EAN-13 and UPC-A barcodes carry. A GTIN that does not start with 978 or 979 is not an ISBN, even on a book.

ISBN International Standard Book Number; 10 digits before 2007, 13 since. An ISBN-10 becomes an ISBN-13 by the prefix 978 and a new check digit.

Registration group the part of an ISBN after the prefix that names a country or a language area (2 for French, 963 for Hungary); it tells where the book was published, not its language.

UPC-A the 12-digit North American product barcode; an EAN-13 with a leading zero.

EAN add-on the small 2- or 5-digit code printed right of an EAN-13; on books, often the price (`90000`: no price).

MQTT a publish/subscribe messaging protocol; QoS 1 means a message is delivered at least once.

SRU Search/Retrieve via URL, the HTTP successor of Z39.50 for querying library catalogues.

Z39.50 the older binary protocol for querying library catalogues, still served by many national libraries.

UNIMARC, MARC21 the bibliographic record formats of library catalogues (UNIMARC at the BnF).

Copy cataloguing cataloguing an item by copying a record that a library has already made for the same edition, rather than describing it anew.

8.3 Tools

Mosquitto a free MQTT broker, packaged for Debian and FreeBSD; the centre of `sapupra`.

SQLite a database held in one file and used through a library, with no server to run; the catalogue is one SQLite file.

FTS5 the full-text search module of SQLite: a virtual table that indexes the words of chosen columns (titles, people, subjects), answers searches by words, prefixes or phrases, and orders the results by relevance.

Datasette a free program (Apache 2.0, written in Python) that serves an SQLite database as a web site: tables to browse, facets to filter by a column's values, full-text search where an FTS index exists, and an API; it opens the database read-only. A optional extra for the web front of `sapupra` (the pages of the entry module are the front itself).