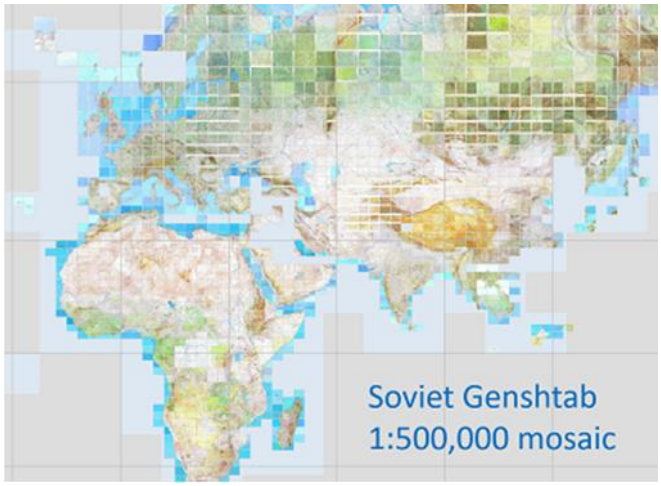
Historical map series online

Matthew Shaw [1]



OpenTiles.org is an online project to stitch historical map sheets into continuous layers, and to present them online as a zoomable tileset. Others are doing similar work, some commercially, but OpenTiles is designed to drive down the cost of preparing the sheets and hosting the huge data volumes involved. The intention is to make the layers free at the point of use.

Opentiles.org is already offering the Soviet Genshtab sheets at scales of 1:25,000 to 1:1 million – some 130,000 sheets stitched together. Other interesting series include Austria-Hungary in 1910, pre-war Latvia and Allied maps of Germany.

Much of the potential value of historical maps cannot be realised because of the difficulty in comparing them with the present-day or with each other. Projections and scales differ, scans are distorted and their organisation into discrete sheets impedes the user experience of collar-free zooming and panning of a multi-sheet map series. Yet high-quality scans of historical maps continue to pour online. Funded by public foundations and private donations, and motivated by educational objectives, there is now a wealth of free historical resources available. Some sources are well catalogued; others rely on users to do their own discovery. But they are usually hard work for the end-user.

Stitching sheets together into a seamless map is not new. As a child, I used paper-clips and adhesive tape. Online out-of-copyright OS maps have long been available from several sites and the National Library of Scotland is a more recent – and more functional – provider. They show the way, but are only scratching the surface of what needs doing internationally.

The tools to stitch maps exist and are maturing. The most generic offering is commercial, from Klokan Technologies, and is deployed by the NLS, British Library, David Rumsey Maps and others. Its approach is to rely on human calibration to precede a process of re-projecting and de-collaring the sheet, with quality achieved by either using skilled technicians or crowd-sourcing multiple attempts. It works well, no matter the cartographic standards of the image, but is inevitably slow.

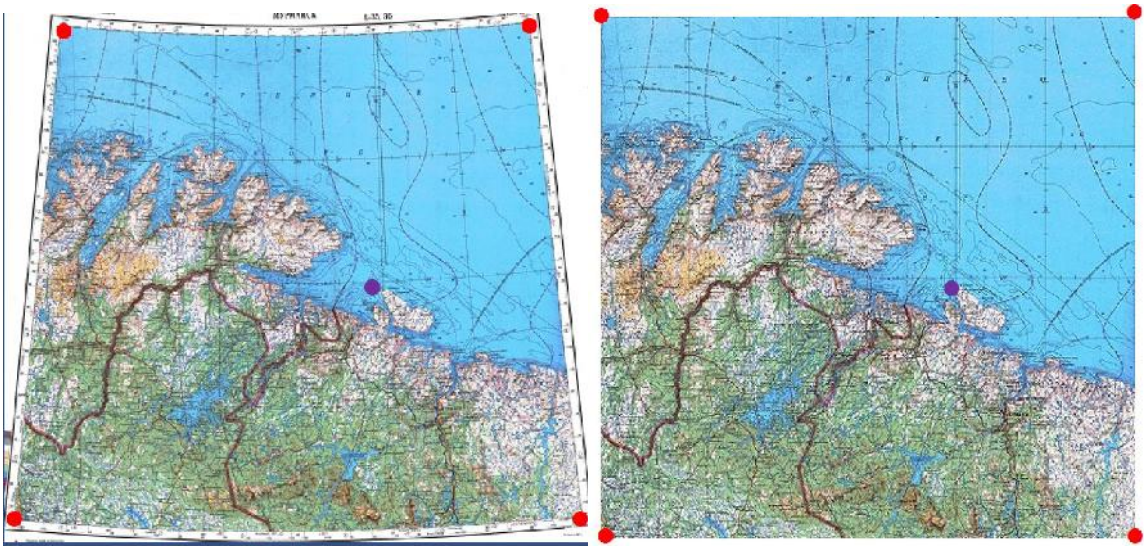
Work began on OpenTiles.org in late-2016. Firstly, it aims to use automated techniques, similar to number-plate or facial recognition systems, to identify the collar on a map sheet. Such computer vision is imperfect but helps to make processing of large sets, such as the 40,000 Soviet Genshtab 1:50,000 series, possible. Secondly, it has built the hardware stack from the bare metal upwards, in order to drive down the cost of a multi-terabyte dataset to an acceptable level.

Open-source software has been used throughout. Free software is still good software, and doesn't come burdened with onerous restrictions on adapting it.

---

[1] *Matthew Shaw is software hobbyist, with an interest in historical cartography. Professionally, he is data architect specialising in business intelligence*

Collar detection and reprojection



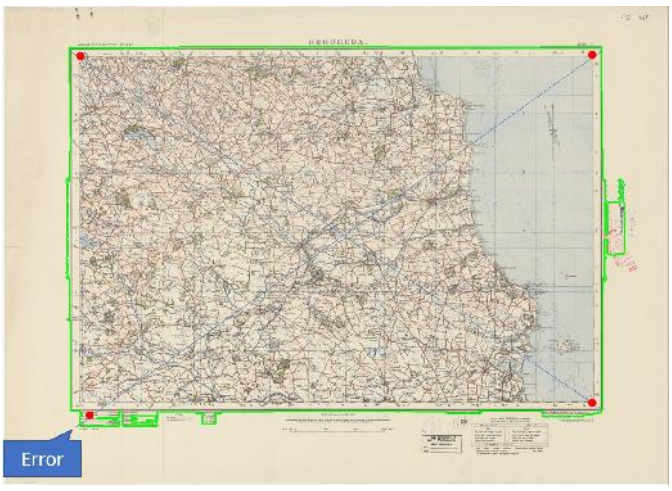
Original map, far left, and target warped to WGS84 and de-collared

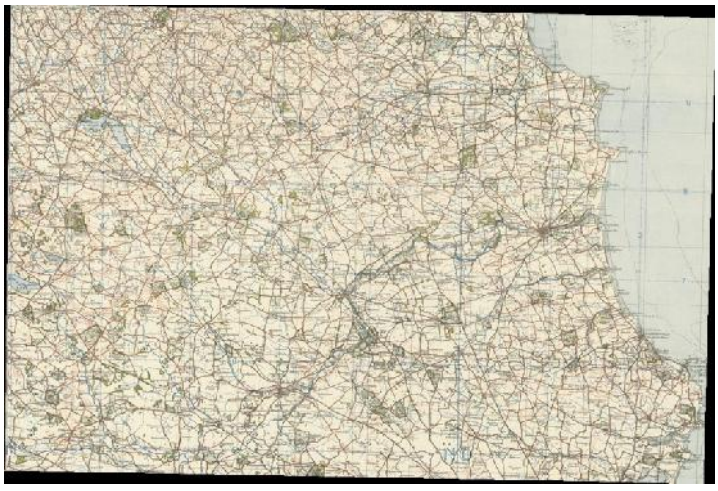
The challenge is to take a map image, detect and remove the collar and re-project it to WGS84. In tests, automated collar detection using Open Computer Vision and the CV2 Python bindings achieved a 98.5% success rate. Typical errors arise where the border is unclear, due to poor scan quality or user markings, and it cannot handle map irregularities such as insets and overflows.

The OS Ireland map of Drogheda (*below*) illustrates automated border detection using Open CV's Contour algorithm, shown in green, and the use of diagonals to find the corners of that contour (shown in red). The algorithm detects hundreds of contiguous areas, but the largest area is usually the one we want – as formed by the sheet border. One error here is caused by some text bleeding into the border lower-left. Nevertheless, the result is 'good enough' for a first pass.

A Python library then helps to re-project the OS Ireland grid to WGS84. The black area is actually transparent, which means that adjacent sheets show through when viewed together, irrespective of the order they are applied.

The map bounds of each sheet are required as input. In this case, the calibration is by reference to the OS Ireland grid, and it may be necessary to know the date of the sheet to apply the correct version when converting to WGS84. Other map series are often easy to calibrate, so long as the sheet naming conventions are understood. Nevertheless, map bounds in latitude-longitude still need to be transformed to the WGS84 datum, and the original datum may not be obvious.





Online presentation

Raster maps are presented online after conversion into a 'tile pyramid'. Each zoom level is a layer of the pyramid, and each level is made up of tiles of 256x256 pixels. Tiles are typically stored as binary records in a database rather than the JPEG or PNG file that the browser receives, because millions of separate files are exceptionally difficult to manage. The web-server must query the database for the required records and render them as files in milliseconds. Mapbox.com has been instrumental in driving the standards for online tile serving. But the compliant open-source tools fall short when scaling up to the 40 million+ tiles already available on OpenTiles.org. Many enhancements to the established tile-making tools have become necessary. The task of building a large pyramid is now subdivided in order for it to be executable, and this has the added benefit of being able to handle partial refreshes as new sheets are found or individual detection errors are corrected. The web-server has been extended to enable new layers to be added without restarting it, and the database indexing redesigned to support multi-million tile pyramids. The database is already several terabytes in size.

Hardware

'Storage is cheap' we are often told. But hosting a multi-terabyte database stresses this assertion – and many hosting companies have upper limits of just 2Tb. OpenTiles.org recognised that building its own infrastructure from off-the-shelf components was the best way to achieve a solution that could scale indefinitely. Cost per Tb is under 10% that of the cheapest hosting company and 17Tb is already installed. Hosting the raw files, processing and tile pyramids together vastly simplifies the logistics. Fibre-optic (to the premises) broadband delivers adequate upload speeds for the web server, at least for the near-term traffic levels.

Futures

OpenTiles.org as a project has lots of raw material and a maturing processing capability. Automated collar detection is still work-in-progress, and the layers online today still have their collars in place. The Allied maps of Germany 1:25,000 have unusually large collars, which means the geolocation errors are considerable. But the tile-making processes work well and a scalable infrastructure is in place.

There are several areas where additional practical contributions are sought. The author has taken the project to its current place almost single-handed, but it needs to transition to be a collaborative open-source project. People can help in several ways but these are the main perceived gaps:

1. The project would benefit from more cartographic knowledge. Applying the correct re-projections, and assembling the code libraries necessary, needs improvement. Users will expect the quality of the layers to improve over time from the 'good first attempt';
2. A good Javascript developer would help drive the value of the finished layers but also help with crowd-sourced engagement;
3. Setting up a steering committee of people familiar with the world of open-source, map librarianship, application of cartographic information would be a real bonus.

Oh, and more raw material is always welcome.