

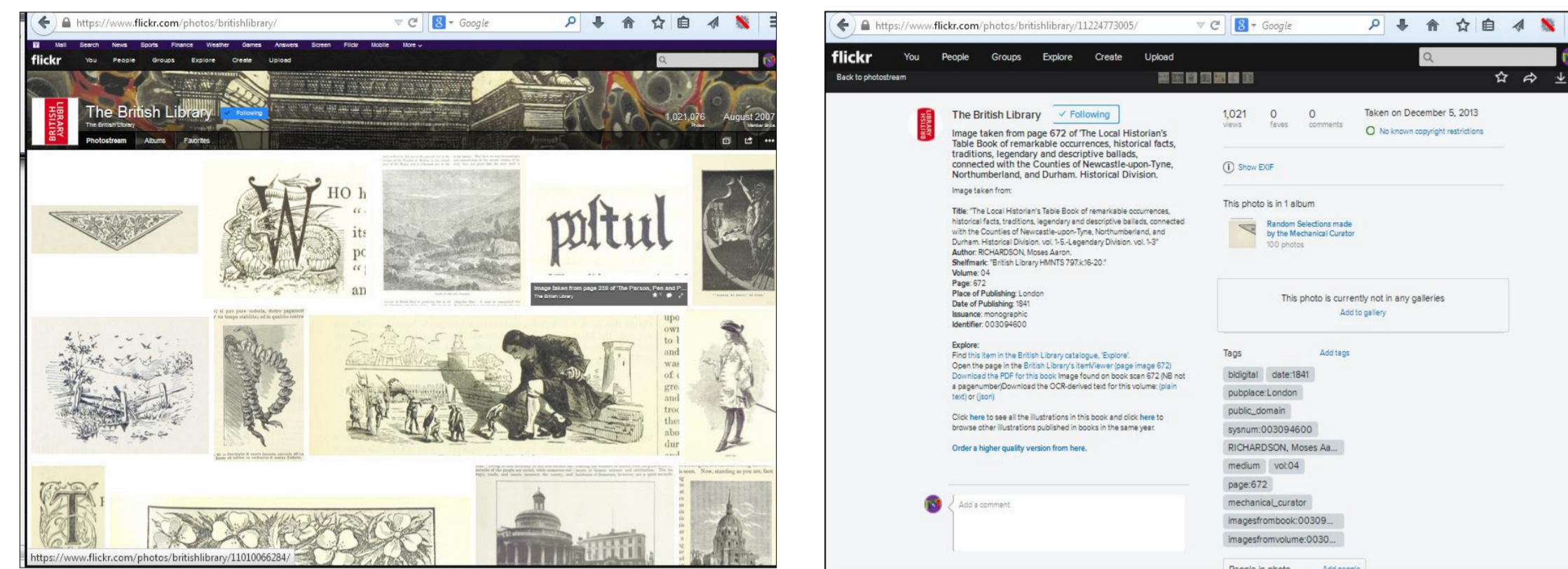
Case Study: Mapping the Maps

How to find 50,000 maps in a haystack of 1,000,000 images; geolocate them, and categorise them...on a budget of \pounds not many Euros

James Heald, Wikimedia volunteer
@heald_j

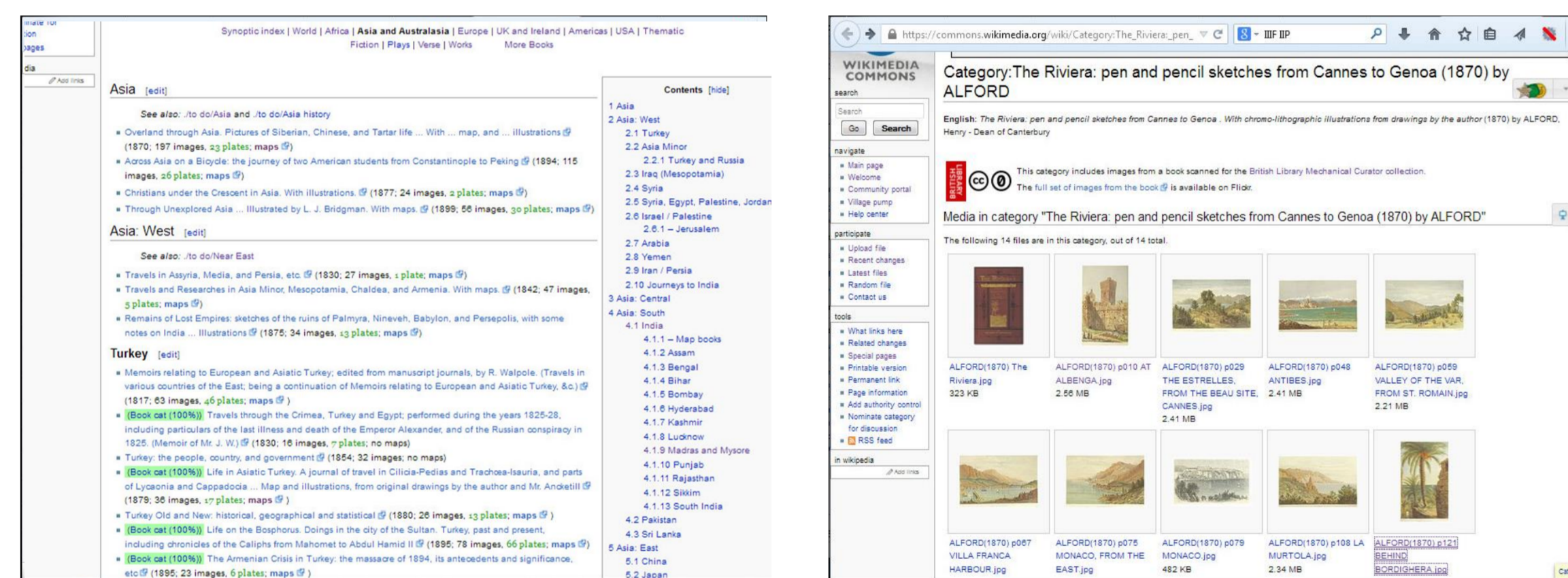
Kimberly Kowal, British Library
Kimberly.Kowal@bl.uk

Our Challenge



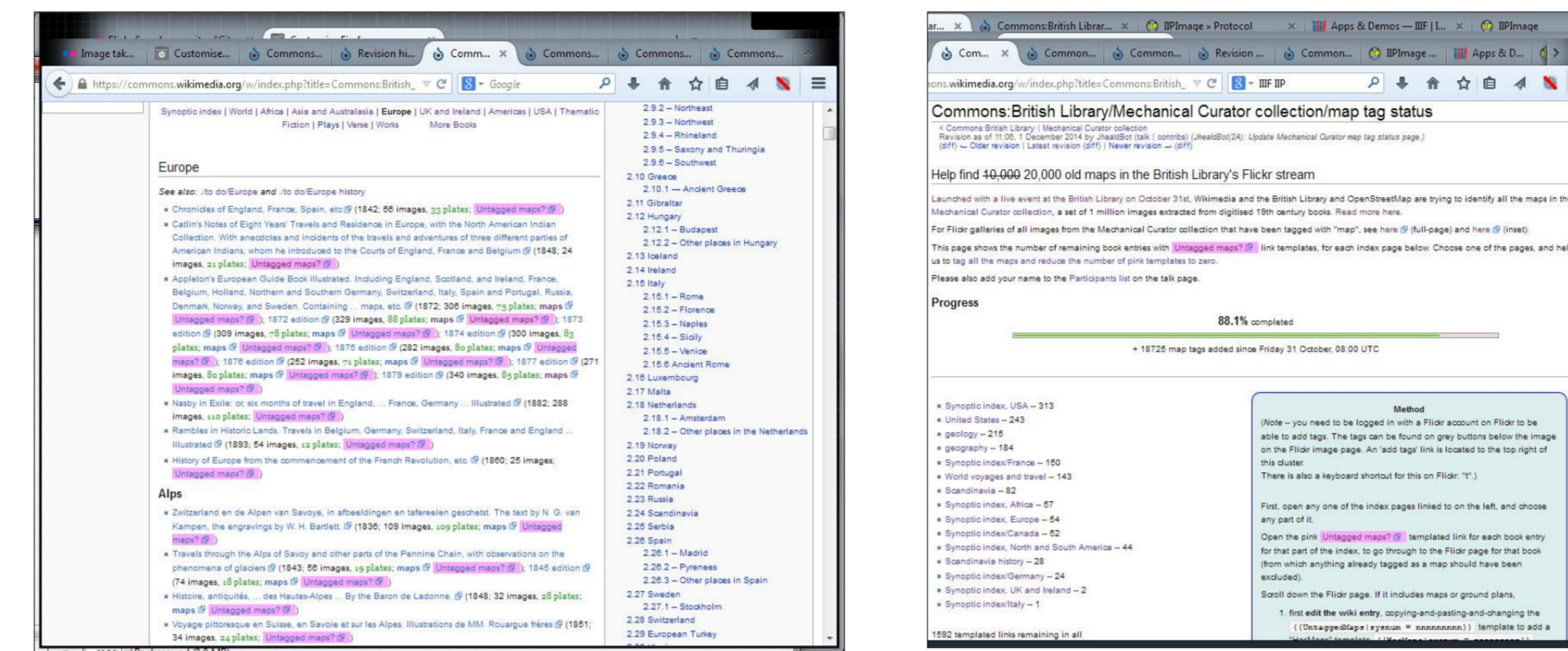
- 1,000,000 out-of-copyright images released by the BL
- Extracted as a by-product from mass-scanning 60,000 books
- But...No metadata finer than the whole-book level!

Wikimedia Volunteer Action



- Build an index by book subject (primarily geographical) -- from shelfmarks, with some hand refinement
- Hand-curated uploads -- 20,000 images hand-uploaded in a year by one user
- Maps identified as a candidate for mass-upload & reuse -- ready search & categorisation possible given co-ordinates

Identifying the maps

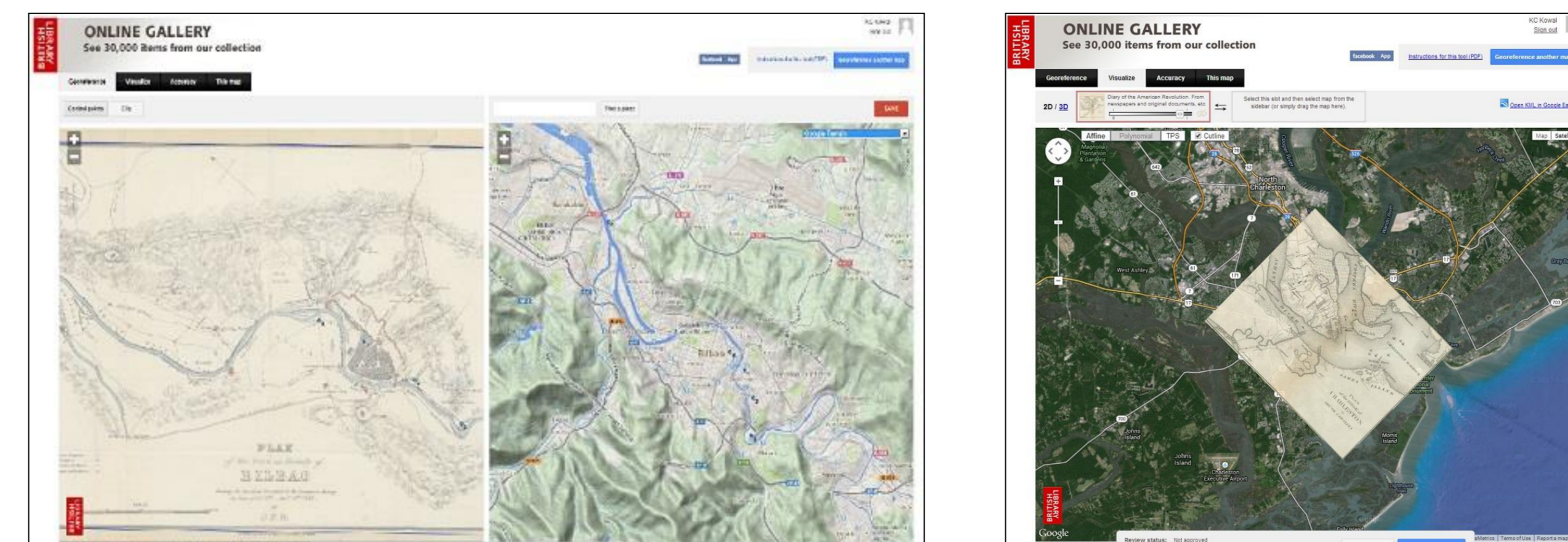


- Crowdsourced effort, based on Flickr tagging & then removing a to-do template for each book from the wiki index
- Kicked off with an open-day event in London, then online
- Some automated pattern-recognition contributions from one user
- Completed in eight weeks.

50,000 maps and plans found

	totals	classmark index	detailed index
misc	16074	14091	1983
Europe	13136	6254	6882
British Isles	7191	269	6922
North America	6758	1524	5234
USA	5782	1209	4573
Asia	2736	1280	1456
Africa	2300	1075	1225
South America	895	659	236

Our next steps



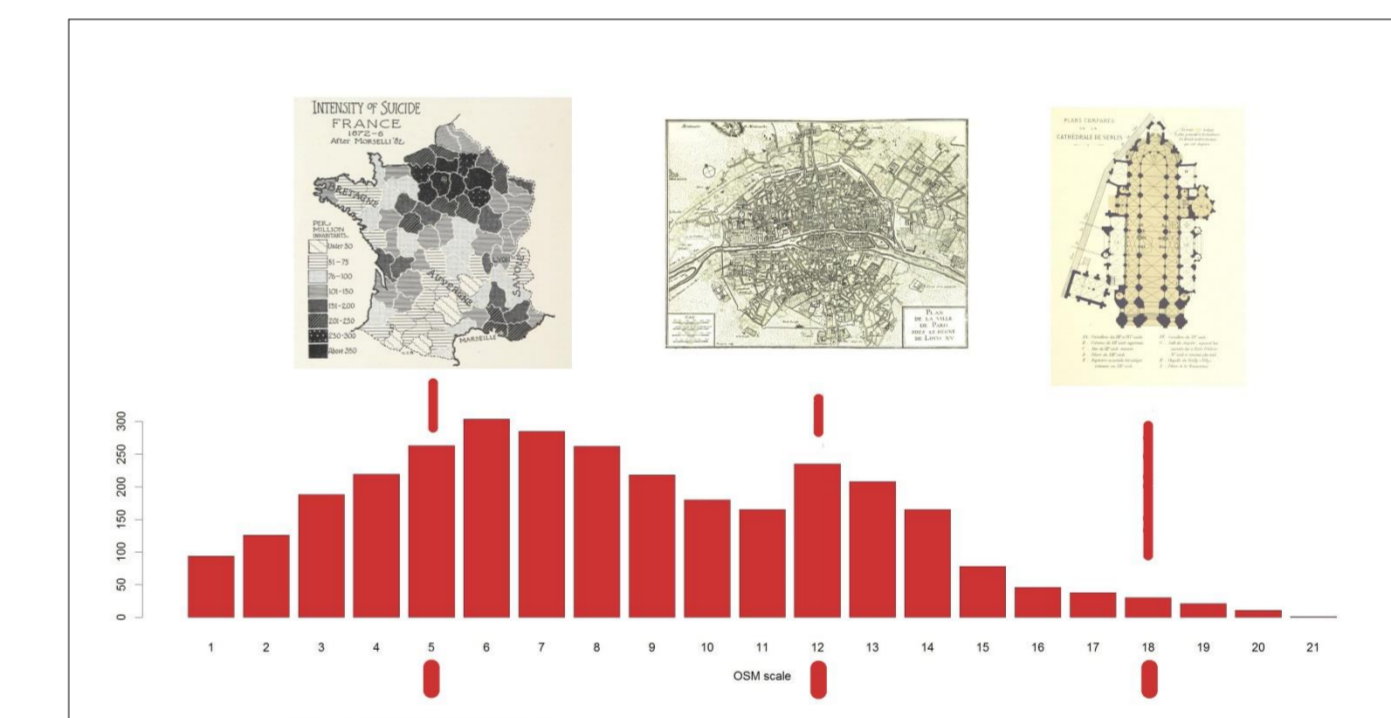
- Geolocate maps via crowdsourcing (using Klokan BL Georeferencer or free alternatives)
- Disseminate maps and geographical coordinate data as widely as possible
- Pilot of 3,000 maps already completed (& uploaded to Europeana)

Results and Re-use

- The map images can now be plotted on a map of the world for search (and laid over present-day maps for comparison)



- Co-ordinates allow maps to be grouped by scale (eg building / town / county / country / continent) and location



- Groupable using co-ordinates + Nominatim + title keywords + Wikidata into sets suitable for hierarchical Wikimedia Commons categories of about 20 to 200 images.

- Mass uploadable to Wikimedia Commons using GlamWiki Toolset (Thank you Europeana!)
- Hope is to have a pipeline that will upload a map automatically to Wikimedia, fully categorised, within 24 hours of a volunteer having geo-referenced it.

