Image: Visual3Dfocus / Shutterstock

Enhancing geospatial content through data conflation

Stefano Cavazzi and Gobe Hobona explain how crowdsourced geoinformation can be conflated with authoritative datasets to realise enhanced defence-oriented products

In our rapidly-changing world, information has never been more strategic to decision making. The new information landscape is associated with social media, high-speed networks and distributed information sharing from people all around the world. Smartphones with built-in GPS and survey capabilities enable anyone to create open source geospatial datasets, transforming the way in which information is produced.

Addressing concerns

Thousands of people already collect geospatial data, as in the case of the OpenStreetMap (OSM) initiative where volunteers contribute to the creation of a global map. This collective approach, defined as crowdsourcing, might raise concerns over the quality of the information collected.

To address such concerns, it is useful to define quality in terms of currency, precision and completeness. Conventional mapping is updated at periodic intervals while the OSM depends on users' intervention. The latter can confer a significant advantage where, thanks to their local expertise, users are capable of detecting and recording change almost in real time.

The same advantage is also valid in the case of precision although, again, OSM can be more variable than conventional mapping depending on user input. Finally, in terms of completeness, OSM has

contributed significantly in placing areas of the developing world finally on the map.

One can also note the crowdsourcing community's rapid and effective response to specific emergency events such as the 2010 Haiti earthquake where Non-Governmental Organisations and volunteers collaborated in generating real-time crisis mapping. Such work continues to this day wherever disaster strikes. With regard to reliability, OSM ultimately benefits from a vast user community that not only contributes to the creation of the map but also uses it on a daily basis.

Authoritative data such as those produced by National Mapping Agencies (NMAs) are tried- and-tested and integrate well with applications and systems developed over the years by third parties. However, NMAs are under pressure to trim their budgets and find new cost-saving ways of delivering information to the same standard but with fewer resources. This ultimately impacts the collection of geospatial data and limits its currency and completeness.

Potential solution

A potential solution to this challenge is to integrate authoritative and crowdsourced datasets. According to the Open Geospatial Consortium (OGC), data conflation is the process of unifying two or more separate datasets that share certain characteristics into one

integrated, all-encompassing result.

In simple terms, data conflation aims to combine geospatial data from separate sources to create a dataset that is better than either source on its own. Conflation consists of several sub-processes. The first involves data discovery, analysis and comparison to ensure its suitability for further processing. Secondly, the data needs to be adjusted to accommodate such operations as map alignment and spatial or thematic generalisation.

Only at this stage can features be matched using geometrical, topological and semantic attributes to achieve unambiguous mapping. This is one of the biggest challenges in data conflation as it poses a number of problems including different coordinate reference systems, representations, resolutions or classifications. After the features have been matched it is possible to join or transfer the required attributes between the datasets to complete the data conflation process. However, if the process is not properly managed, the actual or perceived reliability of the data could be undermined.

Consolidating research

One development aimed at consolidating research on this topic is the Advanced Geospatial Information and Intelligence Services Research (AGIS) project commissioned by the Defence Science and Technology Laboratory (Dstl). Its remit: to focus and consolidate Geospatial Information and Intelligence (GI2) research from across central government, academia and industry, and allow the UK Ministry of Defence (MoD) to achieve - and be seen to achieve – a maximum return on its research investment.

Earlier research by the MoD, and specifically the Geospatial Intelligence Integrated Reference Architecture (GI2RA) research project, proved the feasibility of using a pan-domain harmonised data model to deliver a software architecture able to support the delivery of coherent and consistent Geospatial Intelligence (GEOINT).

Under contract to Dstl, West Sussex-based Envitia has taken this research step further by developing a proof of concept to support the conflation of data from different data sources. The adopted conflation workflow, implemented within the AGIS research project, is shown in Fig.1.

Within the workflow, the crowdsourced data is discovered, its data model analysed, and mapping rules created for its transformation into a harmonised information model. Once the crowdsourced data has been harmonised, the matching process can be implemented with the harmonised authoritative dataset. The matched features are integrated according to predefined integration rules related to
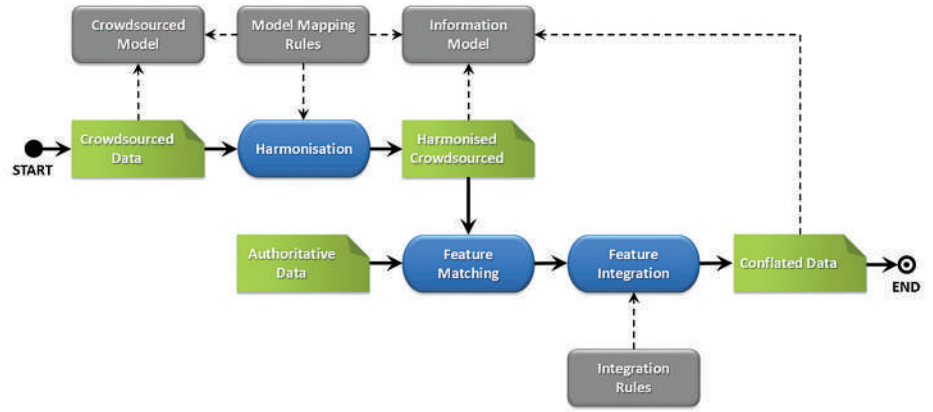

Fig.1: Conflation workflow implemented within the AGIS research project

data content and geometric characteristics, thus creating a new conflated dataset. The enriched dataset includes lineage and quality metadata to ensure traceability and conformity to conflation requirements.

Testing times

In order to test the proposed conflation workflow, an experiment was conducted using authoritative data from the Multinational Geospatial Co-production Program (MGCP) and other data from the OSM crowdsourcing initiative The selected data sources represent a typical reference data set, i.e., a well-structured data model (MGCP) and a crowdsourced semi-structured model but more up to date dataset (OSM).

Using mapping rules that convert them from their original data models, the datasets were harmonised into the NATO Geospatial Information Model (NGIM) that is provided by the NATO Geospatial Information Framework (NGIF). The matching operation within the conflation process included a combination of attributes and geometric feature properties. The geometries of the source linear features were densified and each new vertex was spatially joined to the nearest vertices of the crowdsourced features. All matches were resolved algorithmically.

The results, were extremely promising, with 87% of source data features successfully matched and conflated. Fig.2 illustrates the

feasibility of integrating authoritative and crowdsourcing data sources.

Data conflation benefits

The benefits that data conflation of crowdsourced and authoritative information bestows can be summed-up as:
• Lower survey costs per sq. km of data
• Potential for near real-time updates
• High density observations from a large number of sensors (users)
• Coverage of global regions where authoritative data is sparse, or not available
• A growing crowdsourcing community
• The ability to enrich authoritative datasets with information pulled-through from crowdsourced sources

In conclusion, it is increasingly important for organisations that generate or use geospatial information to create and maintain current, precise and complete databases that support a broad range of spatial analysis and mapping needs.

Data conflation can help in this by enhancing geospatial content. Envitia continues to build on this work by pioneering methods and standards for managing and conflating different types of geospatial datasets. Several of these methods are making their way into the company's MapRite product … one based on more than a decade of experience in asset location alignment and correction.
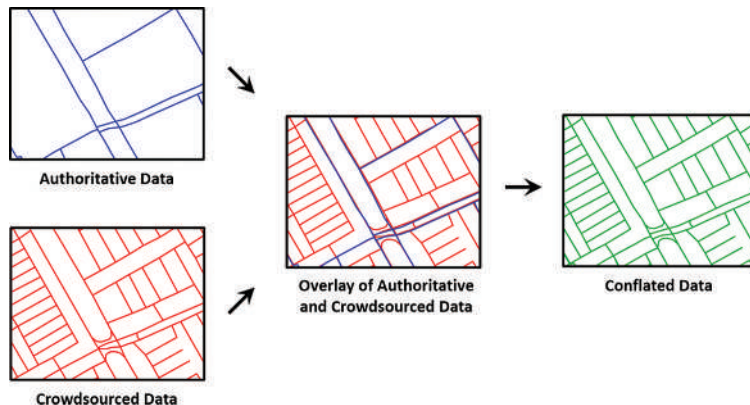
Dr. Stefano Cavazzi is a Geospatial Intelligence Consultant and Dr. Gobe Hobona is the Head of Applied Research, both at Envitia Ltd., in Horsham, Sussex (www.envitia.com)

Fig.2: Integrating authoritative and crowdsourced data