THE CHANGING NATURE OF GEOSPATIAL DATA – CHALLENGES FOR A NATIONAL MAPPING AGENCY

D. A. Holland [1]*, I. Hurst [1], G. Heathcote[1], J. Horgan[1], D. Capstick [1]

[1] Ordnance Survey, Adanac Park, Southampton, UK (david.holland, izabela.hurst, guy.heathcote, jon.horgan, dave.capstick)@os.uk

KEY WORDS: Mapping Agency, 3D models, Automation, Mobile Mapping

ABSTRACT:

National Mapping and Cadastral Agencies have been creating geospatial products for customers for many decades and, in some cases, for over two centuries. During that time the nature of the products largely remained the same, consisting of cartographic representations of the world, usually generalized and projected in a two-dimensional form. Even when mapping agencies began to convert their mapping from paper to digital form, the products created were largely based on their paper map counterparts. In recent times, the general public has become far more aware of geospatial data due to global initiatives from Google, Bing, Apple, OpenStreetMap and others. While some users of geospatial data still require the same products as before, many other users need different kinds of geospatial data and products, ones which will provide new challenges to National Mapping and Cadastral Agencies. In this paper we discuss some of these new geospatial data users and illustrate some the challenges using an example from Ordnance Survey's recent experience of a project in the connected autonomous vehicle domain.

1. INTRODUCTION

1.1 A long tradition of mapping

Ordnance Survey, along with many other National Mapping and Cadastre Agencies, has been providing spatial information to its customers for over two hundred years. For most of this time, the information has been provided in cartographic form, usually as a paper map projected on to a local or national coordinate reference system such as, in Great Britain, the County Series projections of the 19th and early 20th Centuries or the OS National Grid reference system of more recent times (Matthews, 1976).

Towards the end of the 20th Century many mapping agencies had converted their old cartographic mapping into digital form. In many cases, however, the digital data bore a strong relationship to its cartographic forebears. Some changes did occur, for example the creation of grid-based digital terrain models to replace cartographic contours. Even here, there is a strong desire on the part of traditional customers to retain the old familiar cartographic representation, so most leisure maps and raster products still retain the cartographic contours and symbols of old.

1.2 Geospatial data today

It isn't long ago that geospatial data occupied a rather niche area of interest and was largely unknown to many sectors of society. Aerial photography, satellite imagery and photogrammetric interpretation techniques were used by specialists and were seldom encountered by the general public. This all changed with the advent of Google Earth, Google Maps and similar products from Bing, Apple, OpenStreetMap and others. Now almost everyone is a frequent user of geospatial data, accessing it via their computer, in-car navigation system or mobile phone. This has changed the perception of the discipline of geospatial data and has made the work of a mapping agency more familiar to the general populace. Before the ready availability of Google Earth, it wasn't easy to explain cartography, remote sensing and photogrammetry to people, whereas now everyone knows what "satellite imagery" looks like – even when what is being described is often actually aerial photography!

Some traditional customers of mapping data still require the same products or data, such as topographic objects, road and rail networks, land use and land cover maps. There are however, both new and existing customers who want something different, and it is these customers who pose a particular challenge to national mapping agencies. In this paper we discuss some of the new uses of geospatial data, the markets which are using geospatial data for the first time and the demands that new users are making on the suppliers of such data. This will be illustrated by a case study of the Ordnance Survey's role in OmniCAV, a multi-partner project to develop a simulator for autonomous vehicle testing (OmniCAV Consortium, 2019).

2. THE CHANGING NEEDS OF GEOSPATIAL DATA USERS

2.1 Products of a mapping agency

When an organisation has been in business for many decades, it will have built up a portfolio of products that have been developed to meet the needs of a wide variety of customers. In the case of a mapping agency these customers have traditionally been in the government and utilities sectors of the market, involved in activities such as land registration, planning, construction and asset monitoring. These users are well versed in the characteristics of geospatial data, knowing what type of data to use, what quality metrics to apply and which products they need to meet the requirements of their particular use cases.

Traditional geospatial data comes in several different types, but a large proportion of it consists of topographic mapping, depicting both natural and anthropomorphic features, captured using various survey methods. At Ordnance Survey much of this data is collected using aerial photography and photogrammetric processing techniques.

2.2 Data collection methods and positional accuracy

Topographic features may be captured using stereo imagery or via ortho-images, all to a specification originating in the mapping scales of the paper maps of the past. For example, in urban areas, traditionally depicted at 1:1250 scale by Ordnance Survey, the positional accuracy of the features is specified as

having an RMSE of 0.4 m. Over many years of survey practice this is an accuracy value that surveyors in the field, or on a photogrammetric workstation, can be confident of achieving. It is also a value well-known to the long-established users of topographic data.

To those new to the concept of geospatial data, however, such accuracy statements may come as a surprise. When they see a map or a spatial data product, they expect it to be "correct" and often do not consider the many ways in which error and uncertainty can enter into the data collection process. If the user has collected their own data using high-accuracy survey techniques, they want this data to fit perfectly with existing mapping data. It can come as a surprise that the mapping data may be of lower accuracy than their own surveyed data. If these users thought about the cost of capturing their own data, then extrapolated this out to a national scale, perhaps they would be more forgiving.

2.3 Modelling an imperfect world

A second issue that sometimes exasperates new users of geospatial data is the fact that the world is a complex and messy place and is not as neat and tidy as we would like it to be. As an example, depicting a kerb-line along a road sounds like a simple task, but in practice it is fraught with uncertainty. Kerbs are different shapes and sizes, at different heights above the road surface, they frequently drop away or disappear altogether at driveways, pedestrian crossings or entrances to car parks. In every town and city there will be some roads with pristine kerbs, while others will have kerbs which are cracked, broken or completely missing in places. Mud, dirt, leaves and other debris build up against the kerb, often making it difficult to determine where the debris ends, and the kerb begins. In a detailed survey of a road, captured using imagery or lidar sensors, these artefacts will all be captured in the data along with the kerb itself. It is often the job of the skilled photo-interpreter to determine where exactly the kerb is and how to depict it in the data.

The non-uniform nature of the world is also manifest in products based on imagery, be it from aerial, satellite, terrestrial or road vehicle-based platforms. If imagery is to be used in a simulation of a car journey, for example, the user will want the roads to be free of any traffic, perfectly lit but with no shadows or bright areas and with no overhanging vegetation or other occlusions. This is, after all, what the world looks like in a completely simulated model. In practice of course, the real world is far less perfect. It is up to data collectors such as national mapping agencies to make the data they provide as representative as possible of the real world, while also attempting to meet the needs of their users. This is not an easy task to achieve.

2.4 Unfamiliar data formats

Another consideration that must be taken into account is that many users are unfamiliar with GIS formats such as GML, Shapefiles, GeoJSON or GeoPackage. Some customers in emerging markets, such as the autonomous vehicle industry, require data in formats more familiar to the gaming industry than to a mapping agency. It is often the case that traditional GIS cannot cope with the very high resolution, complex models required in these emerging markets. Data providers must either convert their data to the user's format of choice, or make the data available as an API, so the user can extract the data they need in the form they require.

3. THE OMNICAV PROJECT

3.1 Description

In 2018 Ordnance Survey joined a consortium to bid for funding to develop a simulator which could be used to test Connected Autonomous Vehicle (CAV) software. The project succeeded in gaining funding from Innovate UK, the UK Government's innovation agency, and work was started on the project in November 2018. The OmniCAV consortium consisted of eleven partners representing different aspects of the autonomous vehicle field, including simulation experts, traffic monitoring and modelling organisations, an autonomous vehicle manufacturer, an insurance company and a local authority.

Each of the partners brought their own area of specialist knowledge to the consortium and the role of Ordnance Survey was to build a digital model of the OmniCAV test route, in a mainly rural area of Oxfordshire in the UK. The model had three aspects: 1) a vector dataset defining the locations of kerb edges, lane markings, stop lines and other road markings; 2) a network model defining the road lanes, junctions, turns and traffic restrictions; and 3) a 3D model depicting the ground surface and the features on that surface in the neighbourhood of the road.

3.2 Data sources and specifications

The concept of data collection for non-traditional users is not new to Ordnance Survey. For example, recent projects have involved research into the geospatial data required by telecommunications companies for 5G network planning (Ordnance Survey, 2018) and the capture of detailed street-side data for the Manchester Internet of Things project, Cityverve (Cityverve consortium, 2020). Each of these required data from multiple sources and platforms, including high resolution aerial photography, UAV imagery and mobile mapping data from lidar and image-based sensors.

For the OmniCAV project, detailed spatial data was required for the area in the neighbourhood of the main route, an approximately 32 km loop around Oxfordshire (see Figure 1). Initially, field survey methods were used to capture ground control coordinates around the loop and aerial survey was carried out using a Vexcel Ultracam Eagle camera. Third party survey companies were commissioned to capture more detail in the vicinity of the road. Korec provided mobile mapping data from a Trimble MX9 system, while GeoXphere captured aerial imagery using their XCAM oblique camera system, flown in a progressive circular flight pattern. The mobile mapping was driven in both daytime and night-time, each in both clockwise and anticlockwise directions around the loop. This was done in order to provide a good chance of capturing as much of the scene as possible – for example the night-time laser data suffered fewer occlusions as there was far less traffic on the road; while driving in both directions allowed the capture of detail on both sides of the carriageway.

3.3 Data overload

During the early stages of the project, the plan was to use data from all the mobile mapping and airborne sensors, to get the most detailed model possible. This soon proved to be too ambitious a target. Dealing with detailed point clouds over a small area is perfectly feasible, but over a 32 km route the number of points involved (more than six billion) was too much for the processing hardware and software available. From the

data available it would have been theoretically possible to create a 3D terrain model with an average spacing of less than 1.5 cm but, in addition to taking too long to process, such a model would create far too much data for the CAV simulator to be able to process.
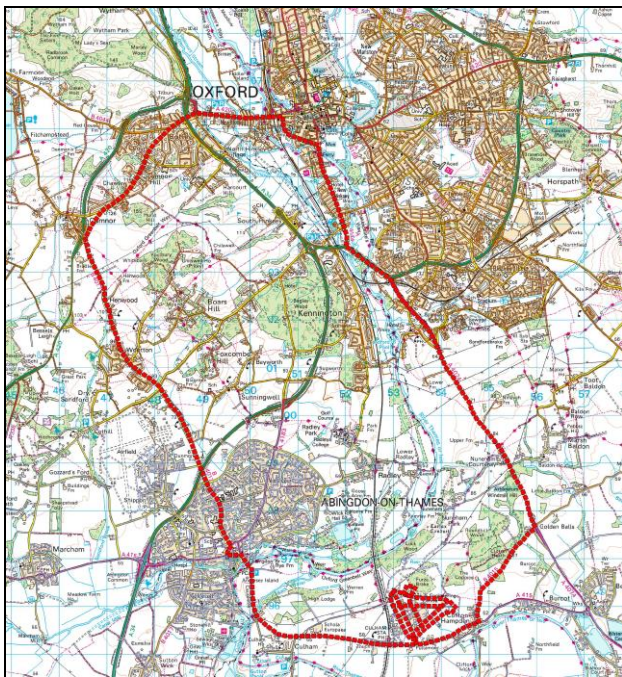

Figure 1. The OmniCAV route in Oxfordshire, UK

Various point densities were experimented with, before a compromise resolution with a point-spacing of around 15 cm was chosen. Even this reduced resolution would be too dense for the simulator if applied to the whole scene, so a variable-resolution terrain model was created, using a high density on the road surface and a lower density (of around 2 m point spacing) in all other areas.

3.4 Collecting features

It was not possible to collect the many different types of features using a single method or single software package. Some of the features could be measured using aerial imagery, others required the use of a mesh model, while some of the more detailed objects could only be captured using the original point clouds. For some of the features, different combinations were tested until the most appropriate method and software package were identified.

3.4.1 Network data:
The capture of the network data was one of the first tasks undertaken as it only depended on data from the aerial imagery already flown at the start of the project using Ordnance Survey's standard production system, a Vexcel UltraCam Eagle camera. ArcGIS software was used to capture the network lines representing the trajectories that could be taken by traffic on the route. The network data included the traffic lanes, bus and cycle lanes, junctions and turns. Lane information also included the width of the lane, the junctions it connected to and any traffic restrictions which applied to it.

3.4.2 Kerbs and roadside edges:
One of the most time-consuming tasks was the extraction of the tops and bottoms of kerbs and roadside edges. This had to be

done to a high level of accuracy as the kerb is a highly important feature in the model within the CAV simulator. The kerbs were extracted from the point-cloud as 3D vectors, depicting the bottom of the kerb (where it meets the road surface), the top of the kerb (where it meets the pavement) and, where present, the position of dropped kerbs (e.g. at entrances to driveways and at pedestrian crossings). Along the more rural parts of the route there was often no obvious kerb present in the point cloud. In these areas an edge-of-road feature was captured instead.

Semi-automatic kerb-finding software was tested at the start of the project. This involved the manual selection of a point on the kerb, from which point the software took over and identified the kerb as it extended along the road. It was found that this method worked in ideal conditions, but often failed in the less-than-ideal conditions found in the real world of urban Oxford and rural Oxfordshire. The kerb following algorithms would start to follow the kerb, then at some point would mis-identify part of the road surface as a kerb and would veer off in seemingly random directions. The effort involved in the rectification of these errors took so much time that it rendered the automatic part of the process redundant. Further testing showed that a greater amount of kerb extent could be captured manually rather than semi-automatically over the same time period. It was reluctantly conceded that all the kerbs and road-edges for the Oxfordshire loop had to be captured by manual methods.

3.4.3 Terrain surface:
The terrain was composed of a detailed 3D mesh on the road surface, a high-definition delineation of the kerb or the road edge, plus a low-resolution mesh in areas beyond the road. To create this model, mobile mapping point clouds, 3D kerb vectors and point clouds derived from airborne sensors were combined to produce a single variable-resolution mesh. While it only takes a short paragraph to describe this, the actual process involved several weeks of research, development and testing, followed by many more weeks of data processing.

3.4.4 Buildings:
Capturing and texturing the many building features along the edge of the OmniCAV route was a task requiring specialist skills and equipment which Ordnance Survey did not possess to the level required of this project. Consequently, a third-party provider, PLW ModelWorks, specialising in the creation of 3D data, was commissioned to generate building models from the XCAM oblique aerial imagery. They successfully delivered textured models of all the buildings in the project region, despite being evacuated for several days during a major weather event, when Hurricane Dorian threatened their premises on the Florida coast.

3.4.5 Road markings, traffic lights, lamp posts and other street furniture:
Road lane markings, stop lines, give way lines, box junctions and other carriageway markings were captured as vector features from the mobile mapping point cloud, using Skyline's TerraExplorer software. This was also used to record the position and height of all the 3D features, including walls, hedges, gates, lamp posts, traffic lights, signposts, bollards, benches and bus shelters. Several different software packages were used during the project to capture the positions of trees and bushes, using both automatic and manual techniques.

3.5 Modelling features

Once the roadside features had been captured, they then had to be placed into the 3D scene within the model. For this task the Blender Open Source software was used as we were fortunate to have a Blender expert working on the team. A set of 3D objects was crafted, to represent all the features that were to be found along the OmniCAV loop. At first, this doesn't sound too difficult, but we soon learned that there are many more different types of roadside object than we originally expected. For example, all the traffic lights on the route had to be modelled, so a traffic light object was created in Blender with a red, amber and green light on a dark pole. Detailed observation of the actual traffic lights on the route revealed that this model was far from sufficient. Some traffic lights had filter arrows on the right or left of the main pole; some had forward arrows instead of circular lights; others had double sets of traffic lights to control different lanes; while still others had extra information such as "no right turn" signs built in to the traffic light structure. Then there were the pedestrian and cycle traffic lights at pedestrian crossings, which also came in different guises. Each of these traffic light types had to be created as a separate 3D object in the model, made up of a pole object, a head object and a set of lamp objects. These separate lamp objects were required to enable the CAV simulation software to control the state of the traffic lights during a simulated journey.

Road signs also come in different shapes, sizes, and combinations; each modelled as a 3D object. Figure 2 shows examples of 3D objects in the model, including a road sign made up of a pole, two separate signs and an overhead lamp.

Figure 2. Examples of features in the 3D model. A road sign, a bollard and a sign indicating a sharp deviation to the left.

Separate Blender objects were created for bus shelters, bus stop signs, benches, fences, gates, walls and all the many other items of street furniture present in the scene. For the trees and bushes, simple "lollipop" models were created, depicting the position, height and crown diameter of each item of vegetation. These were later replaced by more realistic vegetation models in the simulation software.

3.6 Placement of objects

When all the object models had been constructed, the 3D scene could be populated. Blender was again used for this task. All the points and lines collected in the vector data were replaced by 3D objects in the model. These objects were positioned automatically in the scene, then manually adjusted to ensure that they were sited directly on the terrain and were oriented at the correct angle. The route was split up into manageable chunks and for each chunk the terrain, buildings, traffic lights, signs, lampposts, street furniture and vegetation were all brought together into a single model (Figure 3)

3.7 Delivery

The data was delivered in stages, beginning with the network data for the entire route. For the topographic vector and 3D data, the route was split into approximately 1km chunks. The first chunk, within the town of Abingdon, was supplied to the consortium as a prototype dataset. Feedback from the consortium members, especially the designers of the 3D simulator, XPI Simulation Ltd., was used to refine the requirements and inform the methodology used to produce subsequent sections.

Throughout the project we found that there were differences in the requirements of the various consortium members. As several of the partners were using their own modelling and simulation software, they each needed data that could be easily imported into their systems. Sometimes this proved difficult, especially if the modelling software treated data in slightly different ways. For example, the network data described the road layout as depicted by the road markings on the ground. This was exactly what was required by some members of the consortium. Other partners, however, required a model which showed the trajectories that vehicles made as they travelled along the road. It was found that the position of a Give Way line on the road does not always mark the point at which a vehicle begins to manoeuvre when making a turn. In fact, in some cases, if the vehicle had only started to turn at the point marked by the line, it would not actually be able to complete the manoeuvre without mounting the kerb or veering into another lane. This and other such differences in requirements made it difficult to create a model which satisfied every user. A compromise had to be made in such cases, in which some extra data was added to the model by the domain experts involved, to enable it to be used in their systems.

Processing the data for the Abingdon section highlighted the complexity of the model and the requirement for resources beyond the expectations established at the start of the project. It was clear that to deliver the entire OmniCAV route at the same level of detail would take much more time and resources than were available to the project. A meeting with the consortium members was held in October 2019, at which it was decided that seven "areas of interest" would be captured at the detailed level, while a reduced specification would be used on the remaining "linking areas" of the route. The areas of interest were processed first, then the remaining chunks, which were delivered at the end of March 2020.

Figure 3: A view of part of the 3D model showing a zebra crossing, hedges, trees and lampposts

Apart from fixing a few minor issues in the data (such as the occasional tree floating above the ground surface of the model) the role of Ordnance Survey from a technical standpoint is now complete. The project will continue, with the consortium partners adding dynamic objects to the model, to represent autonomous vehicles, conventional vehicles and pedestrians; and running simulations which take into account traffic flows, traffic light phasing and the behaviours of the various simulated road users.

3.8 Lessons learnt

This project has highlighted some of the issues a mapping agency will have to deal with if it wants to provide products to new and emerging markets such as those in the autonomous vehicle domain. We will have to develop new skills and methodologies and find ways to build these into production systems more rapidly than we do at present. One of the key lessons learnt on this project is that not everyone thinks of the world in the same way. One organisation's idea of what a 3D model looks like may be entirely different to another's, and even different groups within the same domain may have quite different requirements. Concentrating on the requirements at the start of a project and determining the specification of the final product at an early stage will be crucial to the successful delivery of projects like OmniCAV in future. National Mapping and Cadastre Agencies will have to adapt or change their production processes, enabling them to deliver data, in different formats and with different attributes, to a more diverse set of customers.

4. CONCLUSIONS

The experience of working on OmniCAV has been at times difficult and frustrating, but it has also been an interesting and rewarding one. We now have a very detailed 3D model of rural Oxfordshire, which will pay a major part in the CAV simulators developed by the other project partners.

The final stages of this project were undertaken during the COVID-19 global lock-down, making data access and data processing much more difficult for the team while they worked from home on their laptops. During the project, the lead partner in the consortium had to leave the project altogether, requiring a change of leadership and a change in some of the technical aspects of the work. In addition, one of the contractors had to abandon their processing for a time due to the imminent arrival of a major hurricane. So, in the end we were quite fortunate that we were able to complete any part of the data collection process, in the face of extreme weather events, a global pandemic and the loss of a leader along the way. We have gained new skills and learned how to deal with new types of customers, and we are looking forward to the next challenge that comes our way. Next time we hope it will be a slightly smoother ride.

5. ACKNOWLEDGEMENT

6. REFERENCES

Cityverve Consortium, 2020: Cityverve website, https://cityverve.org.uk/ (3rd May 2020)

Matthews, A.E.H., 1976: Revision of 1: 2500 Scale Topographic Maps. *The Photogrammetric Record*, 8: 794-805. doi.org/10.1111/j.1477-9730.1976.tb00855.x

OmniCAV Consortium, 2019: OmniCAV Project website, https://omnicav.com/ (22 April 2020)

Ordnance Survey, 2018: 5G planning and mmWave environment reports, on UK Government website, https://www.gov.uk/government/publications/ordnance-survey-5g-planning-and-mmwave-environment-reports (2nd May 2020)

.