LANDXML TO SCIMS:
TOWARDS AUTOMATED HARVESTING OF DP OBSERVATIONS

The NSW regulatory requirement for survey plans to connect to Permanent Survey Marks (PSMs) has the long-recognised potential to assist in the densification of the State Control Survey. On behalf of the Surveyor-General, DCS Spatial Services, a unit of the NSW Department of Customer Service (DCS), is responsible for the establishment, maintenance and improvement of the State Control Survey, which is made available to users via the Survey Control Information Management System (SCIMS).

SCIMS is the state's database containing more than 250,000 survey marks on public record, including coordinates, heights, accuracy classifications and other metadata, provided in the Geocentric Datum of Australia 2020 (GDA2020), its predecessor GDA94 and the Australian Height Datum (AHD).

LandXML is an international data standard for exchanging geospatial information. The Intergovernmental Committee on Surveying and Mapping (ICSM) has developed a national LandXML schema from which NSW has developed a subset NSW LandXML recipe.

Since 2015, DCS Spatial Services has captured and stored registered Deposited Plans (DPs) in the LandXML format. To date, this has resulted in the capture of more than 1 million DPs including over 4.4 million land parcels. DP LandXML files are stored in Cadastre NSW, a system to centrally store, index and track the lifecycle of cadastre-specific data assets.

This article outlines, in general terms, the innovative LandXML to SCIMS (LX2S) pilot project, which was initiated by DCS Spatial Services to automate the harvesting of State Control Survey observations from registered DPs, adjust the 'islands' of harvested observations, and publish the adjusted GDA2020 coordinates and their quality in SCIMS. The automated harvesting and adjustment for the LX2S pilot project was executed using Python code developed in-house to retrieve, test and process DP LandXML files from Cadastre NSW.

This initiative supports industry and community growth by extending the state's fundamental positioning framework, the State Control Survey, with a greater density of survey marks with GDA2020 coordinates of known quality. It also demonstrates to industry customers the clear benefit and outcomes of the regulatory requirement for survey plans to connect to the State Control Survey.

In recognition of its innovative design and the enormous benefits it provides, the LX2S pilot project was awarded joint winner of the 'Extra Dimension & Innovation' category at the 2023 Excellence in Surveying and Spatial Information (EISSI) awards.

NEW SPATIAL APPLICATIONS

The LX2S pilot project required several new spatial applications to be generated to support the automated workflow design. This was critical to enable the desired automated harvesting, Quality Assurance (QA), adjustment and SCIMS update for vast amounts of data.

DCS Spatial Services has been developing in-house utilities to automate and streamline its unique workflows for more than 50 years. To this end, a closed-source code repository called SurvPy has been developed using the interpreted, dynamically typed programming language Python.

The LX2S pilot project utilises innovative data structures and algorithms. As no off-the-shelf software existed for the project requirements, a custom-built code base was generated to further extend the SurvPy library in Python. LX2S adopted an object-oriented programming model, as this was considered the only viable option for a project of this size. The LX2S code base greatly increases the scope of the existing SurvPy library, with additions for the LX2S pilot project totalling approximately 25,000 lines of Python code. This number continually increases as processes are refined, expanded and optimised into the future.

In order to store, perform QA, query and output the relevant State Control Survey observations from registered DPs, a customised SQLite relational database was built to fulfil the project requirements. Customised LX2S database tables were configured for the storing of observations and metadata pertinent to each observation, including its parent DP, observation weightings, QA assessments and overrides (both automatic and manual). This customised LX2S database is the project-critical data structure in the transition from DP LandXML files to the production of least squares adjustment input files.

The open-source DynAdjust least squares network adjustment software distributed by ICSM was used to adjust the LX2S output data. DynAdjust was chosen due to the high performance of both its simultaneous and phased adjustment modes and for consistency and compatibility with the GDA2020 state and national adjustments, which also use DynAdjust. Customised LX2S Python modules allow the automated sequential execution of DynAdjust for each auto-generated input file comprising the LX2S adjustment islands.

MATHEMATICAL INTEGRITY REPORT

The Mathematical Integrity Report (MIR) was developed to speed up the examination of individual DP LandXML files when testing code development and manually assessing individual LX2S island adjustments. The report has large spin-off benefits to DCS Spatial Services as it reduces manual close-checking times from hours/days to minutes and empowers the Surveyor-General with better tools to uphold legislative responsibilities. The MIR also has potential to be of significant benefit to industry stakeholders external to DCS Spatial Services.

The MIR is an attention-focuser that highlights mathematical integrity problems within a DP LandXML file via the use of juxtaposed textual and graphical reports. The MIR converts a DP LandXML file from data to meaningful diagrams and analytical reports of the many survey closures within the LandXML file in a manner fit for human interpretation. At a fundamental level, it checks the closes in a plan in the same manner that a professional surveyor would manually check them with a hand-held calculator, whilst also detecting and highlighting potential problem observations.

The report is delivered as a colour pdf with navigational bookmarks for ease of use. Summary tables and graphics explicitly show those closes which exhibit a misclose outside the relevant legislative tolerances (Figure 1).

Figure 1: Excerpt of a MIR-extracted cycles health diagram example, indicating problematic cycles (green = good, orange = suspect, red = possible error).

INNOVATIVE METHODOLOGY

The scale of the LX2S project vision required innovative approaches to each stage of the project workflow, which can be broken into four broad categories:

1. Retrieval, parsing and processing.
2. Internal verification.
3. Datum connection analysis.
4. Adjustment and SCIMS update.

The LX2S methodology for retrieving bulk quantities of DP LandXML files from Cadastre NSW requires a focused balance between retrieval speed and hardware capacity limitations. A strategy of chunked asynchronous retrieval using 'Elasticsearch' queries was adopted, where the overall query is (if required) broken into smaller chunks so that hardware capacity limits are not breached in downstream processing (Figure 2). Each chunk is then submitted asynchronously to maximise the speed of retrieval of each Cadastre NSW DP bundle. Bundles are assessed to determine the best available LandXML file for each DP.

The process of parsing (breaking up the LandXML file into its components) and storing required the LX2S team to custom-build

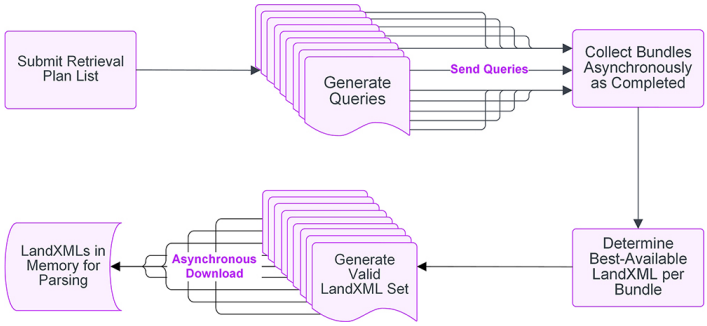

Figure 2: Retrieval overview.

a complete object-oriented data structure to fulfil current and future processing requirements of DP data from LandXML sources, other existing sources as well as sources and formats yet to be defined. LX2S parses retrieved LandXML data and places the parsed data into 'DepositedPlan' Python objects. The DepositedPlan Python object contains attributes, some of which themselves can be objects with attributes (think of a set of drawers where a drawer can itself be broken into a set of drawers). It is into these compartments and sub-compartments that the parsed data is stored.

LX2S processing passes the DepositedPlan object to custom-built Python processing objects. These populate the component objects and attributes within the DepositedPlan object and calculate any attributes required in the downstream processing. One of the main priorities in processing is assessment of the mathematical integrity of the DepositedPlan object, as this comprises the critical pre-QA assessment.

This is essentially a misclose check of the LandXML data for the subject DP and leverages sophisticated applications of mathematical graph theory. Simply put, graph theory examines the relationships of graphs, which are mathematical objects consisting of vertices (or nodes) connected pairwise by edges (or lines). It is a branch of mathematics used to represent relations and networks and is widely used in network analysis.

As a result of the mathematical integrity assessment, each observation is given a weighting, the Misclose Weight Ratio (MWR), which is applied downstream in adjustment weighting. Observations involved in closes outside legislative tolerances are down-weighted unless the MWR value is below a floor value, at which point they are discarded. Observations only involved in closes that meet legislative tolerances are given a ceiling MWR value. Once processing is complete, the relevant data is extracted from the processed DepositedPlan objects and written to the custom-built LX2S SQLite database.

Sophisticated, highly detailed internal verification objects form the bulk of the automated LX2S QA processes, whose broad categories can be summarised as follows:

1. A-priori coordinate generation: The LandXML CgPoint attributes stored in the DepositedPlan object are transformed to the best approximation of datum via sequential transformation method attempts, including a Euclidian least squares transform method.
2. Permanent Survey Mark (PSM) label verification: The PSM labels are verified using a range of techniques, including string matching (exact and combination matching), progressive radial searching for near stations, mark type switching, mark number integer substitution and Trigonometrical Station (TS) name matching.
3. Measurement QA: The validity of each PSM-to-PSM observation is assessed against multiple criteria, including MWR threshold testing, self-looping observations, compiled status and mark status.
4. PSM heights: It is ensured that all PSMs have a valid ellipsoidal or orthometric height for reduction of ground distances to the ellipsoid. PSMs without a height are given an orthometric (AHD71) height interpolated from the NSW Digital Elevation Model (DEM).
5. Measurement reduction: All observations are rigorously reduced to the ellipsoid surface in preparation for the adjustment step.

6. Measurement outlier detection: Outliers are detected and removed using a range of detection techniques, including modified z-score distribution analysis and comparison with SCIMS values.

Datum connection paths are regressively determined for all PSMs and adjustment islands generated (where datum connection is available) using mathematical graph theory. To be classified as 'connected to datum', an island must be connected to two or more established PSMs with Horizontal Positional Uncertainty (HPU) in SCIMS.

The adjustment files are written from the LX2S database to the DynAdjust input files using the LX2S PlanAdjustmentWriter object. The input observations are collated into the islands formed in the previous datum connection step and written to separate adjustment input files. The PSMs with HPU in SCIMS (as determined during the datum connection analysis) are used as the constraints for each island adjustment. Observations that retain an active status in the LX2S database are sent to the input files with a-priori weightings calculated from a combination of legislative tolerances and the MWR.

Once the automated adjustment process and a report of the adjustment outcomes (including convergence status, number of outliers, etc.) is complete, the outcomes are examined for any adjustments requiring manual intervention. Where manual intervention is required, staff assessment is carried out and the required data fields in the LX2S database are manually edited (e.g. mark name changed or observation de-activated) and the adjustment files re-generated.

The results of all successful island adjustments are collated into a single SCIMS bulk update file. The strategy for the SCIMS update required detailed attention from the LX2S team so that survey mark coordinates of superior quality were not accidentally overwritten (Figure 3).

A major strategic decision made was not to update hanging communities of survey marks. Extensive analysis of the pilot project

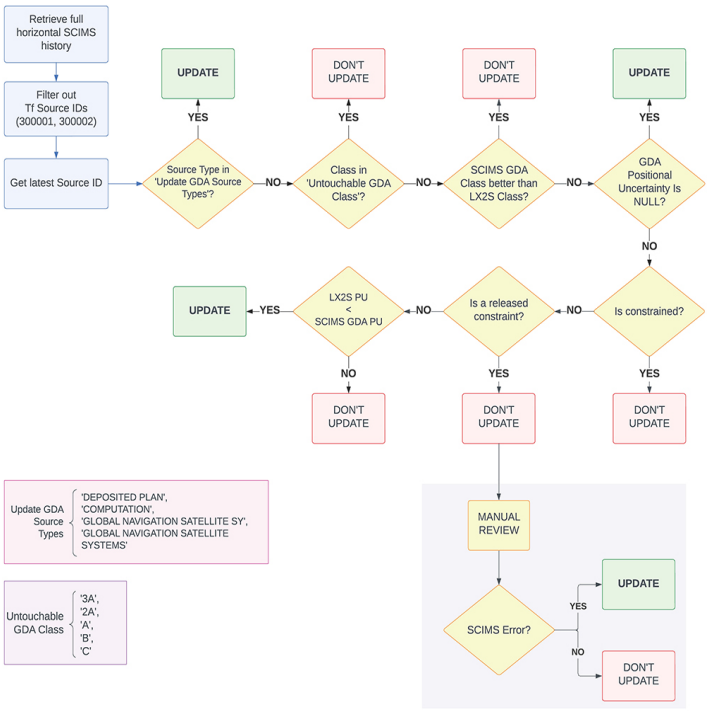

Figure 3: SCIMS update strategy.

areas showed that the veracity of the observations downstream from the articulation point (pivot) could not be ascertained (Figure 4). For all other survey marks, SCIMS is updated with GDA2020 coordinates, HPU and Horizontal Class according to the strategy shown in Figure 3. The minimum HPU that can be attained by a mark being updated is that of the smallest constraint HPU in the subject adjustment island. Horizontal Class is rigorously assessed using horizontal relative uncertainties, and the best Horizontal Class that can be obtained is Class D.

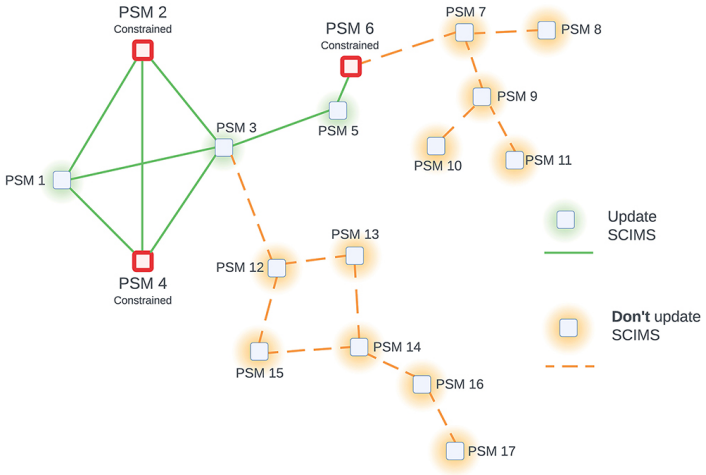

Figure 4: Hanging community strategy.

PILOT PROJECT AREAS AND RESULTS

Eleven major pilot areas were chosen, covering a diverse range of geographical areas in NSW (Figure 5). Overall, across the 11 test areas (and the suburbs within these), the LX2S pilot project retrieved 7,099 DPs, extracted and analysed 33,981 measurements from these DPs, and updated 4,188 survey marks in SCIMS. Of these, 2,059 survey marks (49%) were newly established with Horizontal Class D: 1,367 survey marks were upgraded from Horizontal Class U, and 692 survey marks were upgraded from Horizontal Class E. The distribution of HPU for the upgraded survey marks is impressive, with a median value of 0.035 m and a standard deviation of 0.024 m.

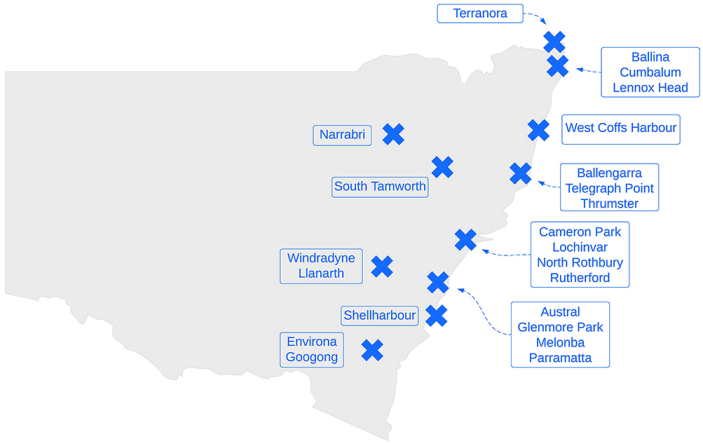

Figure 5: Location of the LX2S pilot areas and their suburbs.

As an example, Figure 6 illustrates the results for the Cameron Park / Edgeworth pilot area, showing the horizontal constraints, the measurements between marks and the resulting Horizontal Class after the SCIMS update. The vastly increased density of established survey marks is clearly evident, indicating the immense improvement made to the NSW State Control Survey through the LX2S pilot project.

FUTURE DIRECTION

The future vision of the LX2S initiative is a fully automated incremental feed of registered DPs updating the State Control Survey on a regular basis, with only minor manual intervention when required. The intent of the LX2S pilot project was to develop the base code structures and workflows necessary to realise this vision for the future, demonstrate the very substantial productivity gains for DCS Spatial Services as well as for internal and external stakeholders, and to determine areas of refinement required for the production environment.

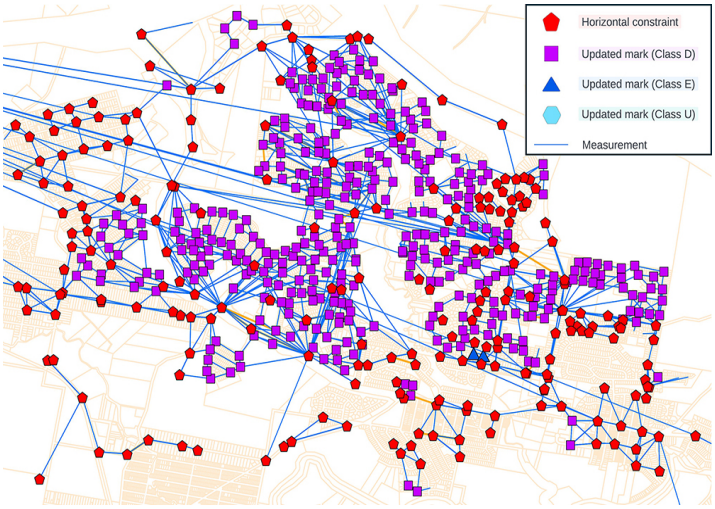

Figure 6: Resulting survey mark updates in the Cameron Park / Edgeworth pilot area (380 marks established, of which 218 were newly established).

The pilot project has identified the next steps required to realise the LX2S vision:

- Realise an area-based (e.g. town or suburb) harvesting model prior to an incremental-feed workflow.
- Refine the automatic QA process.
- Augment reporting to focus the targeting of future field work by DCS Spatial Services surveyors.
- Investigate the potential for access to tools, including the Mathematical Integrity Report, visualisation of LX2S networks and provision of metadata.
- Increase awareness through training and education of DCS Spatial Services staff and customers.

BENEFITS

The LX2S pilot project provides substantial benefits to the surveying profession in particular and the people of NSW in general, including industry and community growth, economic benefits, regulatory benefits and increased community awareness. Based on the pilot project, it is anticipated that LX2S will have the greatest benefit in urban areas, as the active nature of development in those areas engenders a higher density of PSMs with viable paths to datum.

It supports industry and community growth by extending the state's fundamental positioning framework, the State Control Survey, with a greater density of survey marks with GDA2020 coordinates of known quality. For instance, this leads to improved positional accuracy of the NSW cadastre, hence improved positional accuracy of the Foundation Spatial Data Framework (FSDF) and the Digital Cadastral Database (DCDB).

Densification of geodetic networks, being the fundamental outcome of the LX2S pilot project, has long-recognised economic benefits, including the interoperability and standardisation of disparate spatial datasets, improved integration of digital mapping, planning and infrastructure management, improved natural hazard and disaster management, and further enablement of digital government outcomes.

Specific economic gains for NSW include easier compliance with the Surveying and Spatial Information Regulation (thus reducing time and cost of acquiring datum for the preparation of DPs), faster harvesting and ingestion of DPs into the DCDB (enabling faster state and local government outcomes), and greater integrity of the State Control Survey densification from DPs (reducing time loss due to outlier resolution reported by industry customers).

The LX2S pilot project uses the State Control Survey information required to be shown on DPs by the Surveying and Spatial Information Regulation and does so in a timely manner. This demonstrates to industry customers the clear benefit and outcomes of the regulatory requirement for DPs to connect to the State Control Survey.

CONCLUSION

This article has reported on the innovative LandXML to SCIMS pilot project, which was initiated by DCS Spatial Services to automate the harvesting of State Control Survey observations from registered DPs, adjust the 'islands' of harvested observations, and publish the adjusted GDA2020 coordinates and their quality in SCIMS. The automated harvesting and adjustment was executed using sophisticated Python code and innovative workflows developed in-house to retrieve, test and process DP LandXML files from Cadastre NSW.

The LX2S pilot project retrieved more than 7,000 DPs, extracted and analysed almost 34,000 measurements from these DPs, and updated 4,188 survey marks in SCIMS (49% of these were newly established with Horizontal Class D). The distribution of HPU for the upgraded survey marks was impressive, with a median value of 0.035 m and a standard deviation of 0.024 m.

The LX2S pilot project has successfully developed the base code structures and workflows necessary (and identified the required refinements for the production environment) to realise the vision of a fully automated incremental feed of registered DPs updating the State Control Survey on a regular basis, with only minor manual intervention when required. It has also demonstrated the substantial productivity gains for DCS Spatial Services, its stakeholders and customers once this vision is realised.

In July 2024, following further code structure and workflow improvements, the LX2S team initiated a second update to SCIMS, which covered six geographical areas of interest: Bathurst, City of Sydney, Melonba-Kellyville-Box Hill, North Richmond, Shellharbour-Dapto-Kiama and Pambula-Eden-Kiah. Acknowledging a small overlap with pilot project areas, the process retrieved 38,000 DPs, extracted and analysed 76,000 measurements from these DPs, and updated 7,800 survey marks in SCIMS (with 27% of these newly established at Horizontal Class D).

Simon Hine, Jonathon Smith and Dr Volker Janssen work at DCS Spatial Services, a unit of the NSW Department of Customer Service.