# Section 2 General Data Specification

# 2 GENERAL DATA SPECIFICATION

## 2.1 Introduction

2.1.1 This section defines the General Data Specification and standards for spatial data. Specific or additional requirements are provided within the Specialised Data Specifications in Sections 3 onwards. The Specialised Data Specification for the type of data being created or updated must be read in conjunction with the General Data Specification. See Table 2-1 for a list of the Specialised Data Specifications.

**Table 2-1 List of Specialised Data Specifications**

| Specialised Data Specifications | |
|---|---|
| Section 3 | Master Site Plan Datasets |
| Section 4 | Spaces Plan Datasets |
| Section 5 | Reference Data |
| Section 6 | Engineering Detail Survey for Infrastructure and Assets |
| Section 7 | Land Management Data |
| Section 8 | Communications Data |
| Section 9 | Aerial Photography |
| Section 10 | Satellite Imagery |
| Section 11 | Master Planning Datasets |
| Section 12 | Contaminated Sites Data |
| Section 13 | Environmental Data |
| Section 14 | Hazards Data |
| Section 15 | GEMS Geo-Enabled Objects |

2.1.2 The SDMP applies to all E&IG business units and E&IG contractors. All are responsible for the appropriate application of standards and the implementation of the SDMP requirement accordingly.

## 2.2 Data Format

2.2.1 All E&IG spatial data shall meet the format requirements defined in Table 2-2.

Section 2 General Data Specification

**Table 2-2 E&IG Spatial Data Formats**

| Information Type | Preferred Format | Extension | Comments |
|---|---|---|---|
| CAD (Computer Aided Design) Data | AutoCAD Drawing file | DWG | Shall be compatible with AutoCAD Release 2013 |
| GIS Data | ESRI Shapefile | SHP, SHX and DBF | Shapefiles shall be supplied with the three basic file extensions as a minimum and content must comply with relevant SDMP specification requirements |
| | Geodatabase | oracle spatial<br>SQL server<br>SAP<br>ESRI file geodatabase | It is anticipated that database extracts would be provided by E&IG for amendment. |
| Point Cloud | Vendor neutral Point Cloud data format | LAS<br>Or<br>LAZ | For transfer of airborne, terrestrial and mobile laser scanned data. |
| Digital Elevation Models | Band Interleaved by Line<br>Band Sequential<br>TIFF and GeoTIFF<br>ArcGIS GRID<br>ERDAS IMAGINE<br>Comma separated value | BIL<br>BSQ<br>TIFF & TIF<br>Standard contents of ArcGIS GRID directory<br>IMG<br>CSV or TXT | JPG and TIF images shall be supplied with an associated 'world file' containing image header information (.JFW or .TFW file) |
| Data Attribute Tables | Dbase IV<br>Comma separated value<br>MS Excel 97 (*.xls) | DBF<br>CSV or TXT<br>XLS | |

## 2.3 File Names

2.3.1 File names shall succinctly summarise the data to allow users to quickly understand the content of the data. The file name should include the subject of the data and the data locality with reference to the Defence Property name and or EBI if appropriate.

2.3.2 Spaces shall not be used in file names, underscores shall be used to separate text.

2.3.3 File naming conventions specific to particular types of data are covered with the Specialised Data Specification sections of the SDMP. These take precedence over requirements in Section 2 General Data Specification

2.3.4 Specific file naming requirements for ESRI GIS Elevation Model 'Grid' file format: the ESRI Grid file format stores data within a set of directories that can be easily confused with standard directory folders in Windows Explorer.
Each ESRI Grid file shall be clearly named to identify the file as a Grid format, and shall provide a description of the file contents within the filename limit of 13 characters. The following filename format shall be adhered to:

Format: [Short description of file contents]_grd

Example: MBTAelev_grd (elevation grid for Mount Bundy Training Area)

## 2.4 Directory Structures

2.4.1 Directory structures used for spatial data deliverables shall be intuitive to users.

2.4.2 A recommended directory structure is defined within Appendix E and is available for download as a ready-defined structure from the following page on the E&IG Defence Estate Quality Management System (DEQMS) website

http://www.defence.gov.au/estatemanagement/

## 2.5 Spatial Metadata

2.5.1 Metadata is structured documentation that describes spatial data. Information stored within metadata includes the data currency, accuracy, method of capture, data author, and data quality.

2.5.2 Metadata allows data users to make an informed decision on the suitability of data for a given purpose, to understand how the data was captured and the currency of the data.

2.5.3 All E&IG spatial data shall have a metadata record that accompanies the dataset. A valid metadata record is required to allow data to be loaded into the E&IG National Spatial Information Management System (NSIMS).

2.5.4 Spatial metadata shall be created and edited using the E&IG Metadata Entry Tool (MET). The MET allows users to create, view and edit metadata records within the defined metadata structure. Instructions on installation and use are provided on the E&IG Defence Estate Quality Management System website:

http://www.defence.gov.au/estatemanagement/

2.5.5 The MET is a simple tool designed for non-specialist users to create spatial metadata that meets the E&IG Spatial Metadata Specification. The tool provides an interface that leads a user through dialog steps and options to create a metadata record. Users do not need to be familiar with the full E&IG Spatial Metadata Specification (Appendix D) to create metadata that meets this specification.

2.5.6 The MET creates a metadata file with the same filename as the data but with an .XML file extension. This metadata file should remain in the same directory location as the dataset and shall always have the same filename as the data.

2.5.7 Keywords are used to improve efficiency in locating relevant information. A set of keywords has been developed to align to E&IG's business structure. Multiple sets of keywords can be assigned to a dataset to assist users in locating a dataset in the future.

2.5.8 Specific Spatial Metadata Requirements

2.5.8.1 The following metadata requirements shall be stored in metadata records:

- Data capture methodology shall be specified in enough detail for the data capture to be repeated or for the methodology to be reviewed.
- Horizontal and vertical accuracy of captured data shall be recorded.

2.5.8.2 The full E&IG Spatial Metadata Specification is provided within Appendix D. The Metadata Specification is written for reference by the SDMP Technical Authority and Contractors designing Directorate of Strategic Information Management (DSIM) architectures. It is not expected that E&IG Data Suppliers or E&IG Business Managers shall be familiar with the contents of the Spatial Metadata Specification.

## 2.6 Datum

2.6.1 A datum is the reference point against which spatial position measurements are taken. Detailed information on datum can be found on the following website: www.icsm.gov.au.

2.6.2 Horizontal Datum

2.6.2.1 A Horizontal datum is used to describe the position of a point on the Earth's surface in association with a Coordinate System.

2.6.2.2 Defence has adopted the World Geodetic System 1984 (WGS84) as the standard for all spatially referenced data created for E&IG.

2.6.2.3 Although WGS84 has been adopted across Defence to support Defence operations, E&IG recognises that industry partners who support the Defence Estate only have access to GDA and as a consequence will accept the use of GDA2020 for spatial data used to support the Defence Estate as specified in the Spatial Data Management. During the transition to GDA2020, GDA94 may be used but must be clearly marked as such.

2.6.3 Vertical Datum

2.6.3.1 A Vertical datum is used to describe a position vertically as the elevation or depth of a point.

2.6.3.2 E&IG has adopted **Australian Height Datum (AHD)** as the standard vertical datum.

2.6.3.3 Where heights have been derived from GNSS the geoid model used to derive AHD must be stated.

## 2.7 Coordinate Systems

2.7.1 A coordinate system divides the Earth's surface using a set of coordinates to identify the position of a point.

2.7.2 E&IG has adopted the Geographic Coordinate System and the Map Grid of Australia (MGA) as the standard coordinate systems.

2.7.3 E&IG will accept data in the MGA coordinate system where data is provided in a format that does not support the use of Geographic Coordinates. The metadata accompanying these data must clearly state the datum and coordinate system used.

2.7.4 Geographic Coordinate System

2.7.4.1 The Geographic Coordinate System defines the locations of points on the surface of the Earth using Latitude and Longitude.

2.7.4.2 Latitude and longitude coordinates shall be recorded in the format Decimal Degrees and shall be recorded to a minimum of 5 decimal places (approximate equal to a precision of 1 m).
Example: 31.99235 115.88157

2.7.4.3 Lines of latitude are measured as an angle from the equator (0°) to either Pole, 90° South and 90° North.

2.7.4.4 Lines of longitude intersect both the North and South poles. They are numbered using degrees beginning at the Royal Greenwich Observatory in England, which is designated as 0°, and continue both East and West until they meet at 180°.

2.7.5 Map Grid of Australia

2.7.5.1 Map Grid of Australia is a projected coordinate system that defines the locations of points on the surface of the Earth by the measurement of metres from a point of origin.

## 2.8 Units of Measurement

2.8.1 Units of measurement shall be recorded using metric standard scientific (SI) units. The units of measure shall be defined within the dataset attributes or metadata.

2.8.2 The unit of length measurement shall be metres unless specified within the Data Type Specifications in Sections 3 onwards.

2.8.3 In ongoing projects, the unit of measurement within the dataset attributes shall be kept consistent.

## 2.9 Spatial Accuracy

2.9.1 Spatial accuracy measures how closely a position in a dataset is located to its true position on the Earth's surface.

2.9.2 Spatial data shall be captured at an accuracy that is appropriate for the data use. For example, Master Site Plans and Spaces Plans require a high spatial accuracy for use in asset and infrastructure management that is specified within the Data Type Specifications in Sections 3 onwards.

2.9.3 Where a spatial accuracy is not specified within the SDMP the E&IG Project Manager shall provide guidance on an appropriate level of accuracy.

2.9.4 Spatial accuracy shall be recorded in metadata records for all data capture and derived data products.

2.9.5 Global Navigation Satellite Systems (e.g. GPS) are used by E&IG in a variety of business purposes. The right type of GNSS must be used to obtain data of a suitable accuracy for the business purpose for which it is being collected. Hand-held GPS cannot collect data at the accuracy required for the Master Site Plan Dataset. An Engineering Detail Survey using Survey Control Stations must be used for Master Site Plan Data collection. Differential GNSSS techniques can be used for applications that require higher accuracy than a hand-held GNSSS, but don't require the accuracy of an Engineering Detail Survey.

2.9.6 GNSS receivers record the accuracy (or precision) of the GNSS reading as a Dilution of Precision (DOP) or Estimated Precision Error (EPE). These values can vary dependent on the satellite position and signal quality. Where an accuracy of <30m is required, these readings should be recorded with the attributes of the GNSS positional data and stored against each record or in the metadata record.

## 2.10 Spatial Scale

2.10.1 Scale is the ratio of the distance on a hardcopy map or electronic data viewer to the true distance on the ground (e.g. 1:50,000). Data or maps with smaller scales (1:250,000) are less accurate and show less detail compared to data or maps at larger scales (1:25,000).

2.10.2 The accuracy of a dataset or map is equal to ± ½ mm (0.0005m) at map scale. A scale accuracy table is provided in Appendix F.

Example: ± 0.0005 m at 1:50,000 = ± 25m

± 0.0005 m at 1:10,000 = ± 5m

Section 2 General Data Specification

## 2.11 Data Capture

2.11.1 Data products derived from other data sources shall be captured at a scale appropriate for the data use and where possible defined in Australian industry standards.

2.11.2 Data created from multiple data sources takes on the smallest dataset scale. For example, if data is combined from 1:250,000 vegetation mapping and 1:100,000 vegetation mapping the resulting dataset will have a scale of 1:250,000.

## 2.12 Data Viewing

2.12.1 Software functionality enables users to zoom into a dataset and use or print information at very large scales. Caution shall be taken when using data beyond the scale at which it was collected as the accuracy of the data remains the same (see accuracies defined in Appendix F).

## 2.13 Spatial Integrity of Features - Topology

2.13.1 All data captured or provided for E&IG shall be topologically clean and free of errors. Data shall be corrected for overshoots and undershoots, polygons shall be closed, and slivers removed. Figure 2-1 and Figure 2-2 provide examples on how correct topology shall be applied.

2.13.2 Topologically clean data allows users to link attributes to features more accurately and to undertake various types of spatial analysis within a Geographical Information System. For example, analysis can be undertaken to calculate the area or parameter of a feature automatically, analyse proximity of features to other features, and automatically quantify change over time.

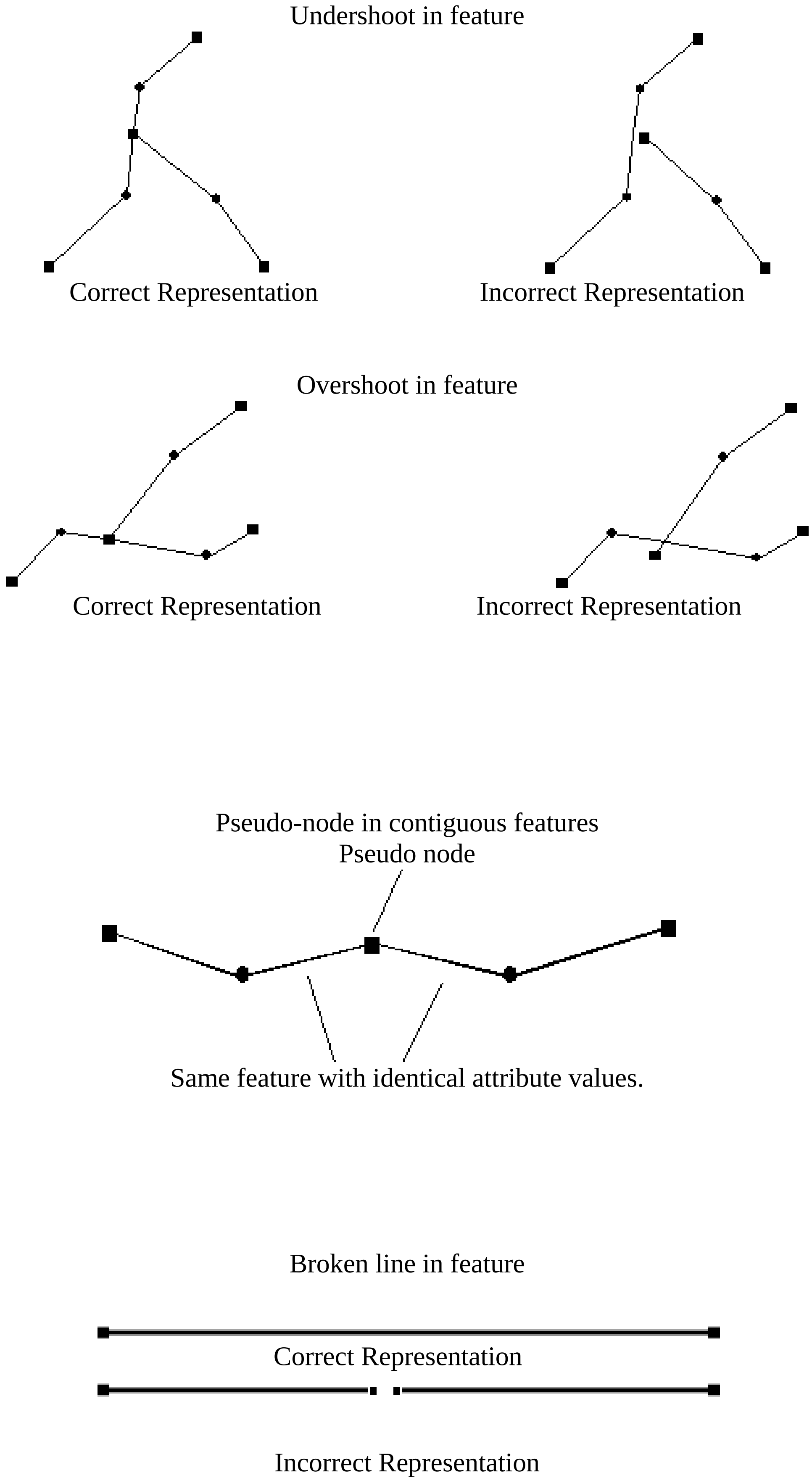


**Figure 2-1 Examples of Correct Topological Representation**

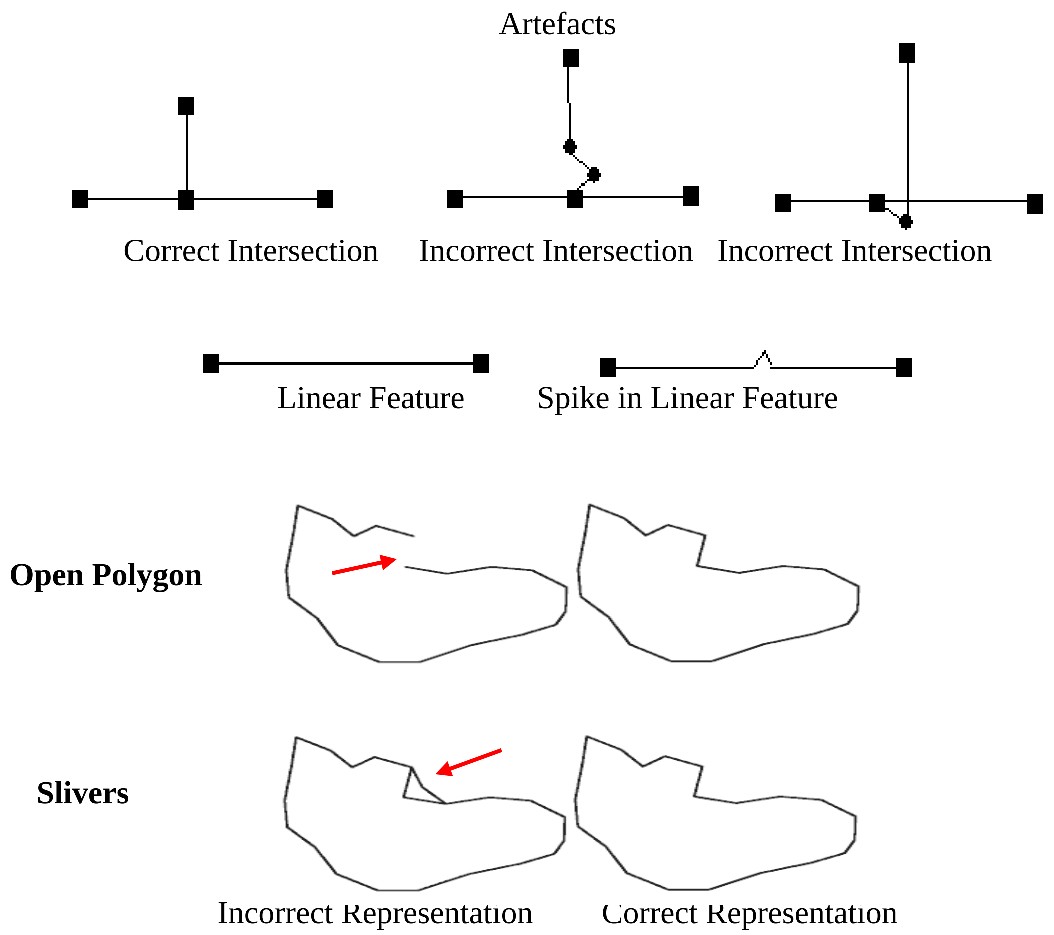


**Figure 2-2 Examples of Correct Topological Representation**

## 2.14 Spatial Integrity of Features – Topology in a CAD Environment

2.14.1 The following information will assist in maintaining topologically clean data in a CAD environment.

- Poly-lines (polygons) close – their start and end points are the same;
- Line features join or snap together – they share endpoints;
- Line feature vertices can also join with point features – they share nodes;
- Point features provide the vertices for the connecting linework; and
- Text is associated with the appropriate polygon, line or point feature.

2.14.2 The capturing/editing of the CAD line work will include:

- "Snapping" of line/poly-line nodes to appropriate point features;
- "Breaking" utility line work at appropriate point features, e.g. where a pipe/cable connects to above ground features;
- Not "Breaking" utility line work where it does not intersect,

- e.g. where pipes/cables are at different height levels; and
- Point/Symbol features are represented as a 'block'.

The following examples provide an overview of the requirements for producing topologically correct Master Site Plan data within a CAD environment.

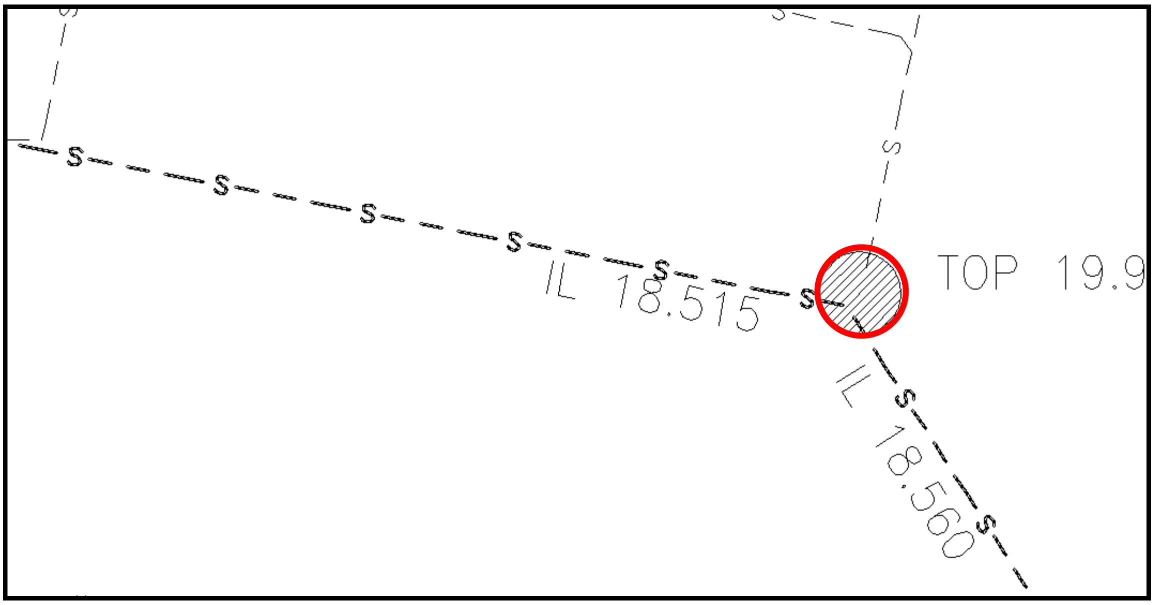


**Figure 2-3 Incorrect Topology - Unconnected line and point features**

**Figure 2-3** shows CAD line features that intersect a point (symbol) feature but the line features do not meet at the node within the point feature.

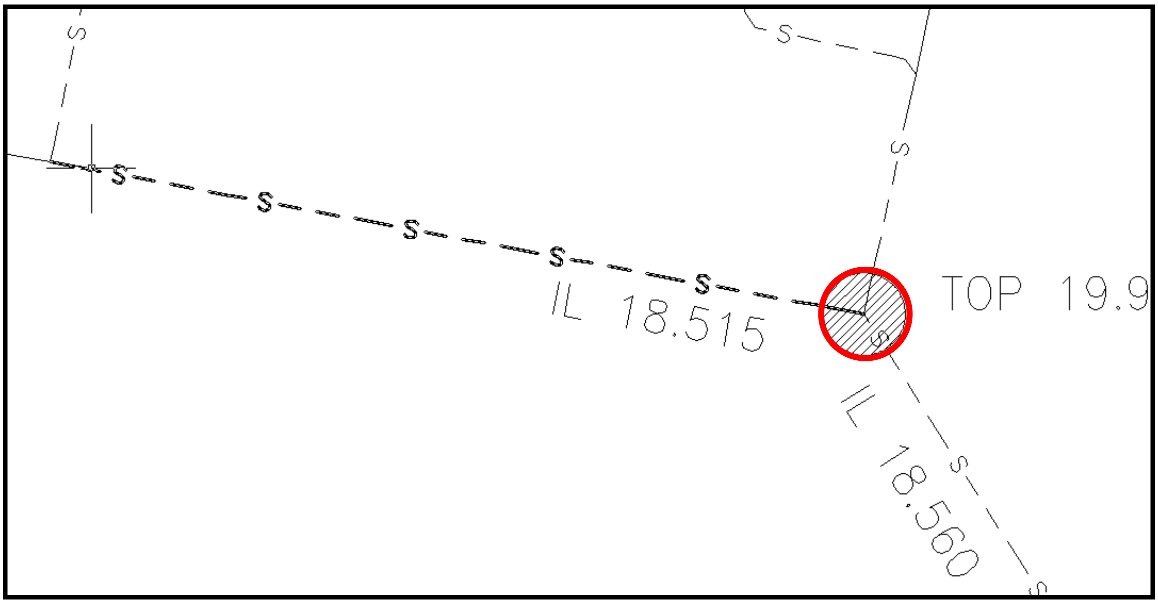


**Figure 2-4 Correct Topology - Connected line and point features**

The correct topology (**Figure 2-4**) represents the line features intersecting a point feature by inserting a new vertex to the line and "snapping" it to the insertion point of the point feature. It is assumed that the line feature should be split at the point feature and that the relevant end points of the lines should be at the point feature.

## 2.15 Data Attributes

2.15.1 Data attributes store descriptive information about particular spatial features within a dataset. Data attributes can include descriptions, dates of capture or update, area, position coordinates, the value of a measurement taken at the location, or a classification or category.

2.15.2 Data attributes are described within the SDMP Detailed Data Specifications within Section 3 onwards for specific types of data.

2.15.3 Where no Specification exists for a data type or where additional attributes are required, the additional attributes shall be reviewed to confirm that there is a clear business requirement. Attributes shall be intuitive to users and shall be allocated intuitive attribute names. Attributes codes and classifications shall follow established industry standards where possible. A description of the business requirements for the information, the attributes and attribute classifications shall be provided within the associated metadata record for the data. If data attributes are provided in an attribute table that is separate to the spatial data then an Estate Business Identifier (EBI) shall be included in both the spatial data and the attribute table to enable the spatial features and attributes to be linked.

## 2.16 Imagery Management – Image Tiling

2.16.1 Digital ortho imagery and LiDAR data are generally ‘tiled’ (divided into smaller images) to manage file size.

2.16.2 It is recommended that the tiling process is undertaken by the data provider. It is also recommended that the data is delivered in both an uncompressed and compressed format.

2.16.3 The recommended tile size to the pixel resolution or Ground Sample Distance (GSD) is detailed in Table 2-3, in all cases the uncompressed file size should be under 300Mb.

**Table 2-3 Recommended Tile Sizes**

| Pixel resolution (GSD) (m) | Tile size (km) |
|---|---|
| 0.05 | 0.5 x 0.5 |
| 0.1 | 1 x 1 |
| 0.2 | 2 x 2 |
| 0.5 | 5 x 5 |

2.16.4 A tile layout dataset or diagram shall be supplied. It is recommended that a layout is provided in GIS format to allow it to be overlaid with the tiled images.

2.16.5 It is recommended that for most users images are used in compressed formats. Imagery files in excess of 300mb on standard PCs will result in slow display rates. There are numerous proprietary software programs available that can be used to tile or compress data. These include ERDAS Imagine and ERmapper.

2.16.6 Digital imagery file size can be reduced by applying a compression algorithm. The compression ratio can be varied, however the greater the compression the more the quality of the image will be degraded. There are lossless image compressions algorithms available, but the amount of compression is limited. In most cases it is recommended that files are compressed to no more than a 10:1 ratio.

2.16.7 Tile File Names

2.16.7.1 The naming convention for tiles cut from a large data is generally based on the co-ordinates values of the lower left corner of the tile. This can be amended, as required, for smaller tiles.

EXAMPLE: Lower left corner: 298000E 6230000N

Tile Name: 2986230

2.16.7.2 If appropriate additional information can be added to the tile name e.g. M6298623010110 where:

M=MGA
6=Zone 56
298=Easting
6230=Northing
1=0.1m resolution
01=January
10=2010

2.16.7.3 Tile metadata should record the following information in addition to E&IG Standard Metadata requirements:

- Release version;
- Area of tile;
- Date flown;
- Camera/sensor type;
- Photo scale;
- Image Resolution; and
- Digital Elevation Model (DEM) details (if relevant).

## 2.17 Hardcopy Archival

2.17.1 Where electronic documents do not exist, and hardcopy files are to be scanned for archival, the following requirements shall apply unless otherwise specified in the Statement of Work:

- The plans shall be scanned at a minimum of 200dpi;
- Raster images shall be saved as compressed tiff images (LZW compression);
- Any colour plans shall be scanned as 24 bit jpeg images;
- Images shall be cropped to minimum plan extents;
- Images shall be rotated where necessary;

- Skewed images shall be de-skewed (original dyelines may not be straight); and
- Poor quality originals shall be enhanced (de-speckled and/or sharpened).

## 2.18 Vectorisation of Scanned Information

2.18.1 Vectorisation of hardcopy information and/or electronic images has been undertaken by Defence in the past to capture vector features relating to Defence Properties. Vectorisation of information is not recommended due to the low return on the financial investment and should only be undertaken for a specific business purpose where the financial value of undertaking the work is greater than the cost of vectorisation.