International Cartographic Association
Association Cartographique Internationale

Deriving Multiple Representation Database: A Model Generalisation Approach

J. Boodala [a, *], O. Dikshit [a], N. Balasubramanian [a]

[a] *Geoinformatics, Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur - 208016, Uttar Pradesh, India.*
jaggubreddy@gmail.com, onkar@iitk.ac.in, nagaraj@iitk.ac.in

* Corresponding author

Abstract: The main aim of the research being conducted is to automatically derive multiple representation database through model generalisation. In this regard, this paper presents a detailed methodology of the derivation process of multiple representation database. The topographic information of three different levels of detail (LoDs) viz., LoD1 (1:10,000), LoD2 (1:25,000) and LoD3 (1:250,000) are considered for this research. The abstract LoDs, i.e., LoD2 and LoD3 are derived automatically by model generalisation. The method for linking corresponding features in all these LoDs is shown and thus deriving multiple representation database. The major component of the proposed methodology is the model generalisation process. The design of the model generalisation process which includes generalisation operators, classification hierarchies, geometric constraints, functional hierarchies and topological constraints are discussed in this article. The proposed methodology is currently in the implementation stage.

Keywords: multiple representations, level of detail, model generalisation, multiple representation database

1. Introduction

The change that may occur in geometry and topology of a feature when the level of detail (LoD) of its representation changes is termed as "multiple representations" (Buttenfield, 1993). The Multiple Representation Database (MRDB) is a database structure in which two or more representations of a same feature are stored individually at different LoDs and they are connected (Kilpeläinen, 1997).

The research problem on multiple representations was introduced by National Centre for Geographic Information and Analysis (NCGIA) (Buttenfield, 1993). Thereafter, many researchers have attempted this problem. Kilpeläinen (1997) proposed a two-fold model based on model & cartographic generalisation and described its application in the context of multiple representation. The model for MRDB system was also presented and it was applied to topographic data. Chaudhry (2007) used a model generalisation approach to derive the geographic phenomena at 1:250,000 through aggregation process. This work involved the modelling of the geographic phenomena at multiple levels of detail.

The benefits of MRDB are non-redundant data, possibility to automatically check the consistency and error, flexible maintenance, etc. (Sarjakoski, 2007). These advantages of MRDB form the motivation for this research problem.

The remaining part of this paper is organised as follows. Section 2 describes the structure of MRDB. The methodology to derive MRDB is explained in section 3. Lastly, conclusions and future work are provided in section 4.

2. Structure of MRDB

The Figure 1 demonstrates the structure of the target MRDB. The plan is to use ladder approach to derive LoD2 (1:25,000) from LoD1 (1:10,000) and LoD3 (1:250,000) from LoD2 through model generalisation process. The strategy and the results of model generalisation process helps in maintaining the bidirectional links between LoDs as shown in Figure 1.

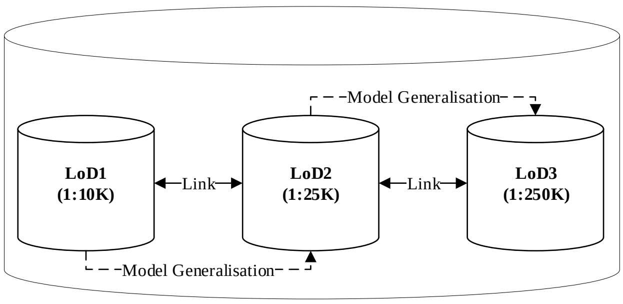

Figure 1. Multiple Representation Database (Boodala et al., 2018)

LoD1 (1:10K)		LoD2 (1:25K)		LoD3 (1:250K)	
From	**Product**	**From**	**Product**	**From**	**Product**
OS	OS Open Map – Local (OML)	OS	OS VectorMap District (VMD)	OS	Strategi
OS	Boundary-Line (BL)	OS	OS Open Roads (ORoads)	EG	EuroRegionalMap (ERM)
		ESDIN	ExM Large Scale (ExM-L)	ESDIN	ExM Small or Medium Scale (ExM-SM)

Table 1. Data products

3. Methodology for MRDB Derivation

3.1 Data products used in modelling phase

To design the MRDB structure at theme and feature level, the data products and their specifications listed in the Table 1 are used. These data products and specifications are obtained from organisations like Ordnance Survey (OS), EuroGeographics (EG) and also from European Spatial Data Infrastructure Network (ESDIN) project.

3.2 Data models matching process

The derivation of MRDB starts from the source database, i.e., LoD1. The OS Open Map – Local (OML) and Boundary-Line (BL) data products are used in the source database. The feature types and their geometries in the source database are shown in Table 2. (Ordnance Survey, 2016a, 2016b)

The feature types and their geometries which are supposed to be derived from LoD1 to populate target database, i.e., LoD2, are modelled using OS VectorMap District (VMD), OS Open Roads (ORoads) and ExM Large Scale (ExM-L). (ESDIN, 2010a; Ordnance Survey, 2016c, 2016d)

Similarly, for LoD3, Strategi, EuroRegionalMap (ERM) and ExM Small or Medium Scale (ExM-SM) are used. (BKG-Germany, 2017; ESDIN, 2010b; Ordnance Survey, 2015)

The Table 2 presents the one to one mapping of the feature types at different LoDs. This mapping is done using the definitions of feature types and their attribute information. This gives an idea of how to represent a feature at different LoDs and how to model their relationships. The feature type such as "Urban or Built-upArea" in "Settlements" theme appears only at LoD3. This feature type can be modelled using "Building" feature type at LoD1 and LoD2. Thus, modelling the functional hierarchy.

3.3 Generalisation operators and their order of applicability

The model generalisation operators are essential to transform the database from LoD1 to other abstract levels (LoD2 and LoD3). The open data products (shown in Table 1) of different scales from OS are studied. This study helped in choosing the required model generalisation operators and also to decide their order of applicability. Table 3 is the outcome of this procedure and it will be refined after experiments by incorporating guidelines from ESDIN and/or EG.

The Add (C+), Eliminate (C-), Reclassify (Cc), Aggregate (Gg), Collapse (Gc), Merge (Gm) and Simplify (Gs) (Roth et al., 2011) operators are used in this research. The definitions and notations of these operators given by Roth et al. (2011) are followed.

The Table 3 illustrates the generalisation process. To derive the 'Building' at LoD2, Cc followed by Gm operators are applied. This in turn is followed by Gc and Cc operators to derive LoD3. Similarly, it is shown for all the feature types in the source database (LoD1). The C+ operator is used to indicate the selection of feature types to LoD1.

The attribute information of the feature types is available in the form of codelists. Some codelists are hierarchical in nature and they can be used to drive the generalisation process. These codelists are used to form the classification hierarchies. This attribute information is used by C+, C- and Cc operators. However, Gg, Gc, Gm and Gs operators use geometry information of the feature types.

The Figure 2 shows advantages of linking hierarchical codelists from different data models. In this example, the codelists from Infrastructure for Spatial Information in Europe (INSPIRE), ESDIN, EG, OS and European Location Framework (ELF) are used (BKG-Germany, 2017; ELF, 2016; ESDIN, 2010b; INSPIRE, 2013; Ordnance Survey, 2016b). The abstract level concept like 'Commerce and services' can be derived from the 'Secondary education' which is not possible by only using OS classification scheme. In the same way, other classification hierarchies can be linked.

3.4 Geometric and topological constraints

The geometric constraints which are essential to drive the generalisation are shown in Table 4. This information is organised according to LoDs and the sources of these values are also tabulated.

Kilpeläinen (1997) stated that "Topology does not change from scale to scale. An exception occurs only when the objects are omitted in context with scale transformation". In order to derive the consistent data at all LoDs, the topological relationships between the features at each LoD should be maintained. The topological constraints are defined at the feature level using ERM product specification (BKG-Germany, 2017) and it is shown in Table 5.

LoD1 (1:10K)			LoD2 (1:25K)			LoD3 (1:250K)		
Product	**FeatureType**	**Geometry**	**Product**	**FeatureType**	**Geometry**	**Product**	**FeatureType**	**Geometry**
Buildings and Structures								
OML	Building	GM_Surface	VMD	Building	GM_Surface	ERM ExM-SM	Building Building	GM_Point GM_Point
OML	Glasshouse	GM_Surface	VMD	Glasshouse	GM_Surface			
OML	ElectricityTransmissionLine	GM_Curve	VMD	ElectricityTransmissionLine	GM_Curve	ERM ExM-SM	PowerTransmissionLine PowerLine	GM_Curve GM_Curve
Communications - Rail								
OML	RailwayTunnel	GM_Curve	VMD	RailwayTunnel	GM_Curve	Strategi	Railway	GM_Curve
OML	RailwayTrack	GM_Curve	VMD ExM-L	RailwayTrack RailwayLink	GM_Curve GM_Curve	Strategi ERM ExM-SM	Railway Railway RailwayLink	GM_Curve GM_Curve GM_Curve
OML	RailwayStation	GM_Point	VMD ExM-L	RailwayStation RailwayStationNode	GM_Point GM_Point	Strategi ERM ExM-SM	Railway Point RailwayStation RailwayStationNode	GM_Point GM_Point GM_Point
Communications - Road								
OML	Road	GM_Curve	VMD ORoads ExM-L	Road RoadLink RoadLink	GM_Curve GM_Curve GM_Curve	Strategi ERM ExM-SM	A Road; B Road; Minor Road; Motorway; Primary Route. Road RoadLink	GM_Curve GM_Curve GM_Curve
OML	RoadTunnel	GM_Curve	VMD ORoads	RoadTunnel RoadLink	GM_Curve GM_Curve			
OML	MotorwayJunction	GM_Point	VMD ORoads ExM-L	MotorwayJunction MotorwayJunction RoadNode	GM_Point GM_Point GM_Point	Strategi ERM ExM-SM	Roads RoadIntersection RoadNode	GM_Point GM_Point GM_Point
OML	Roundabout	GM_Point	VMD ORoads ExM-L	Roundabout RoadNode RoadNode	GM_Point GM_Point GM_Point	Strategi ExM-SM	Roads RoadNode	GM_Point GM_Point
			ORoads ExM-L	Road RoadLinkSequence; Road; ERoad; RoadName.	NA NA	ExM-SM	Road; ERoad	NA
Land Cover								
OML	Woodland	GM_Surface	VMD	Woodland	GM_Surface	Strategi ERM ExM-SM	Woodland Woods; Forest. WoodForest	GM_Surface GM_Surface GM_Surface
			VMD	Ornament	GM_Surface	ERM ExM-SM	GroundSurfaceElement SoilSurfaceRegion	GM_Surface GM_Surface
Land Use								

OML	FunctionalSite	GM_MultiSurface	VMD ExM-L	FunctionalSite ApronArea; AerodromeArea	GM_Point GM_Surface	ERM ExM-SM	Airport; Airfield. AerodromeArea	GM_Surface GM_Surface
OML	ImportantBuilding	GM_Surface						
			VMD	HeritageSite	GM_Point			
						Strategi	Land Use Seed	GM_Point
Settlements								
						Strategi ERM	Settlement Seed Built-upArea	GM_Point GM_Point
						Strategi ERM ExM-SM	Urban Built-upArea BuiltupArea	GM_Surface GM_Surface GM_MultiSurface
Named Places								
OML	NamedPlace	GM_Point	VMD ExM-L	NamedPlace NamedPlace	GM_Point GM_Object	Strategi ERM ExM-SM	Cartographic text PopulatedPlace; NamedLocation. PopulatedPlace	GM_Point GM_Point GM_Point
Administrative Boundary								
BL	BoundaryLine (collection of all boundaries)	GM_Surface	VMD ExM-L	AdministrativeBoundary AdministrativeBoundary	GM_Curve GM_Curve	Strategi ERM ExM-SM	Administrative Boundary AdministrativeBoundary AdministrativeBoundary	GM_Curve GM_Curve GM_Curve

Table 2. Data models matching

Source Database FeatureType	**Generalisation order →**	
LoD1 (1:10K)	**LoD2 (1:25K)**	**LoD3 (1:250K)**
Building (GM_Surface), C+	Cc, Gm	Gc, Cc
Glasshouse (GM_Surface), C+		C-
ElectricityTransmissionLine (GM_Curve), C+		Cc
Road (GM_Curve), C+	Cc	Cc, Gs
RoadTunnel (GM_Curve), C+	Cc	C-
Roundabout (GM_Point), C+	Cc	Cc
RailwayTunnel (GM_Curve), C+		Cc, Gm, Gs
RailwayTrack (GM_Curve), C+		Cc, Gm, Gs
RailwayStation (GM_Point), C+		Cc
FunctionalSite (GM_MultiSurface), C+	Gc, Cc	C-
Woodland (GM_Surface), C+		Cc, Gm, Gs
AdministrativeBoundary (GM_Curve), C+		Cc, Gs

Table 3. Generalisation operators and their order

Proceedings of the International Cartographic Association, 2, 2019.
29th International Cartographic Conference (ICC 2019), 15–20 July 2019, Tokyo, Japan. This contribution underwent
single-blind peer review based on submitted abstracts. https://doi.org/10.5194/ica-proc-2-10-2019

FeatureType(s)	Level of Detail
Buildings, Administrative boundaries, Woodlands	LoD3 (1:250K) • Area > 60,000 m^2 = feature is represented as a polygon. (ERM) • Area > 40,000 m^2 = feature is represented as a polygon feature. (ExM-SM)
Glasshouses	LoD1 (1:10K) • Area > 5000 m^2. (OML) LoD2 (1:25K) • Area > 5000 m^2. (VMD)
Electricity transmission lines, Railways, Coastlines	LoD3 (1:250K) • Length of edge between two connected points ≥ 50m. (ERM & ExM-SM) • Weed & Fuzzy tolerance = 5m. (ERM) • Weed & Fuzzy tolerance > 5m. (ExM-SM)
Roads	LoD1 (1:10K) • If dual carriageway is closer than 32.5m then it is collapsed to a single centreline. (OML) • Dead-end < 36m = Remove them. (OML) • Weed & Fuzzy tolerance = 4m. (OML) LoD2 (1:25K) • If dual carriageway is closer than 32.5m then it is collapsed to a single centreline. (VMD) • Dead-end < 36m = Remove them. (VMD) • Weed & Fuzzy tolerance = 4m. (VMD) LoD3 (1:250K) • Length of edge between two connected points ≥ 50m. (ERM & ExM-SM) • Weed & Fuzzy tolerance = 5m. (ERM) • Weed & Fuzzy tolerance > 5m. (ExM-SM)
Roundabouts	LoD1 (1:10K) • Area < 450 m^2 = roundabout is represented as point feature and extend the roads to meet it. (OML) LoD2 (1:25K) • Area < 450 m^2 = roundabout is represented as point feature and extend the roads to meet it. (VMD)
Settlements	LoD3 (1:250K) • Urban area ≥ 1 sq.km. = large urban area. (Strategi) • Urban area < 1 sq.km. = small urban area. (Strategi) • Open areas within urban area < 0.5 sq.km. = remove those open areas. (Strategi) • Area > 60,000 m^2 = settlement is represented as a polygon feature. (ERM) • Area > 40,000 m^2 = settlement is represented as a polygon feature. (ExM-SM) • Area ≥ 10,000 m^2 = settlements is represented as a polygon feature. (Chaudhry, 2007)

Table 4. Geometric constraints

FeatureType	Topological Relationship	Related FeatureType	Description
Buildings (GM_Point)	Must not overlap with	Buildings (GM_Point)	Buildings must not overlap between them.
	Must not be covered by	Settlements (GM_Point)	Buildings must not be covered by settlements as nodes.
	Must not be covered by	NamedPlace (GM_Point)	Buildings must not be covered by named location node.
	Must not be covered by	RailwayTrack (GM_Curve)	Buildings must not be covered by railway lines.
	Must not be covered by	Road (GM_Curve)	Buildings must not be covered by road lines.
ElectricityTransmissionLine (GM_Curve)	Must not intersect or touch interior		Electricity transmission lines can only touch at their ends and must not overlap each other.
Road (GM_Curve)	Must not intersect or touch interior		Road lines can only touch at their ends and must not overlap each other.
	Must not overlap with	RailwayTrack (GM_Curve)	Road lines must not overlap with railway lines.
RailwayTrack (GM_Curve)	Must not intersect or touch interior		Railway lines can only touch at their ends and must not overlap each other.
	Must not overlap with	Road (GM_Curve)	Railway lines must not overlap with road lines.
Settlements (GM_Surface)	Must not overlap with		Settlements as area feature must not overlap between them.
	Must not overlap with	Woodland (GM_Surface)	Settlements as area must not have overlap with vegetation area features.
Settlements (GM_Point)	Must not overlap with		Settlements as nodes must not overlap between them.
	Must not be covered by	Buildings (GM_Point)	Settlements as nodes must not be covered by buildings.
	Must not overlap with	Settlements (GM_Surface)	Settlements as node feature must not overlap settlements as area feature.
Woodland (GM_Surface)	Must not overlap with	Woodland (GM_Surface)	Vegetation areas as polygons must not overlap with themselves.
	Must not overlap with	Settlements (GM_Surface)	Vegetation area features must not be overlap with settlements as area features.
	Must not be covered by	Settlements (GM_Surface)	Vegetation area features must not be covered by settlements as area features.
AdministrativeBoundary (GM_Curve)	Must not intersect or touch interior		Administrative boundaries can only touch at their ends and must not overlap each other.
	Must not have isolated start node and/or end node		Administrative boundaries lines must touch one other administrative boundary line and cannot be isolated.
	Must not have pseudo-nodes		The end of a line cannot touch the end of ONLY one other line but several.
NamedPlace (GM_Point)	Must be inside	Settlements (GM_Surface)	Named location node feature must be inside a settlements as area feature.
	Must not be covered by	Buildings (GM_Point)	Named location node feature must not be covered by buildings.
	Must not be covered by	RailwayTrack (GM_Curve)	Named location node feature must not be covered by railway lines.

Table 5. Topological constraints

Proceedings of the International Cartographic Association, 2, 2019.
29th International Cartographic Conference (ICC 2019), 15–20 July 2019, Tokyo, Japan. This contribution underwent single-blind peer review based on submitted abstracts. https://doi.org/10.5194/ica-proc-2-10-2019

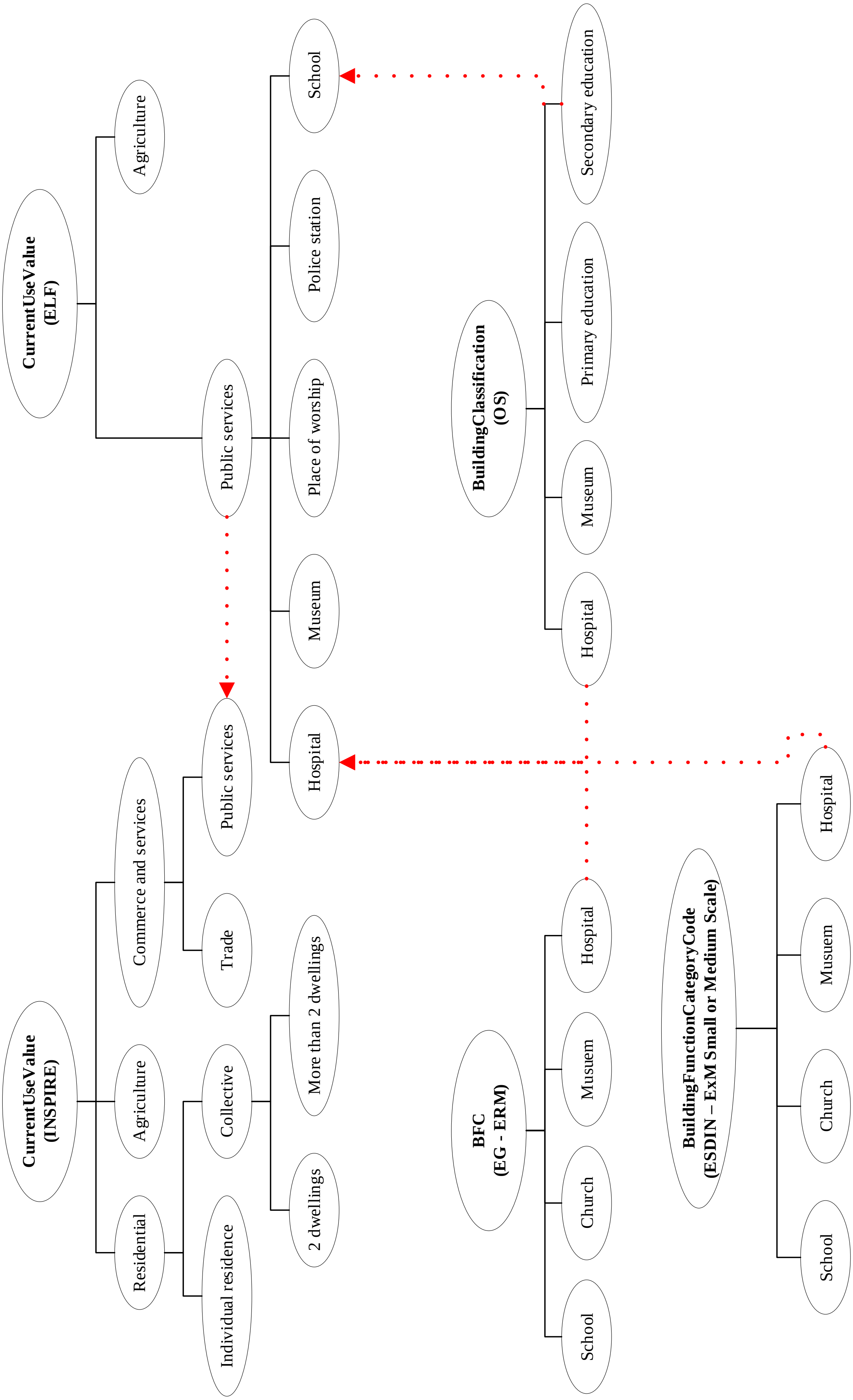

Figure 2. Codelists linking

Proceedings of the International Cartographic Association, 2, 2019.
29th International Cartographic Conference (ICC 2019), 15–20 July 2019, Tokyo, Japan. This contribution underwent
single-blind peer review based on submitted abstracts. https://doi.org/10.5194/ica-proc-2-10-2019 |

3.5 Selection of generalisation algorithms

The model generalisation operators used in this research are implemented using the suitable generalisation algorithms. These algorithms for different feature types and contexts are chosen from the recommendations given in Automated Generalisation New Technology (AGENT) project reports (AGENT, 1999a, 1999b).

3.6 Implementation and evaluation

The research is currently in the implementation stage. The Python programming language is used to implement the proposed methodology and PostgreSQL with PostGIS extension is used as a backend database.

The OS OpenData products and their specifications are used to design the methodology. The results of the implementation will be tested against the same data products as well as the different products like OS Open Zoom Stack. The methodology will also be compared with the products of Survey of India (SoI).

4. Conclusions

This paper proposed the methodology to derive the MRDB. In line with that, the data modelling aspects for different LoDs, generalisation operators and their sequence, geometric and topological constraints are presented in this paper. This information is used to realise the prototype of the proposed concept. Eventually, the findings of this research will be transferred to derive MRDB for Indian scenario.

5. Acknowledgements

The authors are grateful to Ordnance Survey, United Kingdom for their OS OpenData products. The authors would also like to thank the anonymous reviewers for their suggestions.

6. References

AGENT (1999a). *Selection of Basic Algorithms*. (DD2): AGENT Consortium.

AGENT (1999b). *Strategic Algorithms Using Organisations*. (DD3): AGENT Consortium.

BKG-Germany (2017). *EuroRegionalMap: Pan-European Database at Medium Scale*. (Specification and Data Catalogue for Data Production). BKG, Germany: EuroGeographics.

Boodala, J., Dikshit, O., & Balasubramanian, N. (2018). Conceptual Framework for Multiple Representation Database. In : *Vol. 38. Indian Cartographer, Emerging Technologies in Cartography.* Indian National Cartographic Association.

Buttenfield, B. P. (1993). *Multiple Representations: NCGIA Research Initiative 3*. Buffalo NY.

Chaudhry, O. Z. (2007). Modelling Geographic Phenomena at Multiple Levels of Detail: A Model Generalisation Approach Based on Aggregation (PhD Thesis). The University of Edinburgh, Edinburgh.

ELF (2016). *ELF Data Specification*. (Data Specification for topographic and administrative reference data at all levels of detail): ELF Project.

ESDIN (2010a). *ExM Data Specification - Large Scale*. (Implementation Version, D7.2): EuroGeographics.

ESDIN (2010b). *ExM Data Specification for Medium/Small scale*. (Data Specification for topographic and administrative reference data at medium/small scale): EuroGeographics.

INSPIRE (2013). *Data Specification on Buildings*. (Technical Guidelines, D2.8.III.2): European Commission Joint Research Centre.

Kilpeläinen, T. (1997). Multiple Representation and Generalization of Geo-Databases for Topographic Maps (PhD Thesis). Suomen Geodeettisen Laitoksen julkaisuja. Geodeettinen Laitos., Kirkkonummi.

Ordnance Survey (2015). *Strategi*. (User Guide and Technical Specification). Southampton: Ordnance Survey, United Kingdom.

Ordnance Survey (2016a). *Boundary-Line*. (Product Guide and Technical Specification). Southampton: Ordnance Survey, United Kingdom.

Ordnance Survey (2016b). *OS Open Map - Local*. (Product Guide and Technical Specification). Southampton: Ordnance Survey, United Kingdom.

Ordnance Survey (2016c). *OS Open Roads*. (User Guide and Technical Specification). Southampton: Ordnance Survey, United Kingdom.

Ordnance Survey (2016d). *OS VectorMap District*. (User Guide and Technical Specification). Southampton: Ordnance Survey, United Kingdom.

Roth, R. E., Brewer, C. A., & Stryker, M. S. (2011). A typology of operators for maintaining legible map designs at multiple scales. *Cartographic Perspectives*. (68), 29–64. https://doi.org/10.14714/CP68.7

Sarjakoski, L. T. (2007). Conceptual Models of Generalisation and Multiple Representation. In *Generalisation of Geographic Information* (pp. 11–35). Elsevier. https://doi.org/10.1016/B978-008045374-3/50004-1