Implementation of Data Alignment Measures for the Alignment of Planning, Lands and Public Works Data

Final Report (Volume 2E)
Specification and Explanatory Notes of TPU/SB CSU

March 2004

Amendment History

Change Number	Revision Description	Pages Affected on Respective Version	Revision/ Version Number	Date	Approval Reference
0	Initial Release		1.0	4 Mar 2004	

Table of Contents

1 Introduction

1.1 General Overview

1.1.1 TPU/SB Common Spatial Unit (TPU/SB CSU) is defined as the standard unit for exchange of geospatial data of TPU/SB, comprising spatial data and a set of common attributes, between participating departments, i.e. the Data Agent, Data Owners and Data Users.

1.1.2 This document gives the specification of TPU/SB Common Spatial Unit (TPU/SB CSU) which shall be followed by all participating departments in the data exchange process. Participating Departments are not obliged to adopt the same specification for their internal systems.

1.1.3 The CSU specification includes 4 major components:

(a) CSU Definition – to describe the scope covered in the CSU, the rules adopted for the delineation of CSU polygons, the CSU Identifier to uniquely identify a CSU, and the data custodianship;

(b) CSU Workflow – to describe the processes and workflows involved in the production and exchange of CSU data during different stages of a CSU lifecycle;

(c) CSU Data Interface Requirement – to describe the logical structure of CSU data exchanged between the interfacing systems of PDs;

(d) Maintenance of the CSU – to describe the regular mode and frequency of data provision by Data Owners, and dissemination by the Data Agent.

1.2 Enquires

1.2.1 Any enquires to the specification shall be referred to the DAM Management Committee, c/o HPLB.

2 CSU Definition

2.1 Overview

2.1.1 The Tertiary Planning Unit (TPU) is a geographic reference system demarcated by Planning Department for the territory of Hong Kong. Each TPU is sub-divided into a number of Street Blocks (SB) for urban areas and Village Clusters (VC) for rural areas.

2.1.2 Projected population is key information that is usually affiliated to the TPU/SB and is always requested by PDs for their business planning purpose. Currently, the Working Group of Population Distribution (WGPD) is responsible for producing projections of population distribution by TPU and SB annually. The projected population by TPU estimated for each base year and the following 6 years are open source data that can be provided to other government departments and the public for planning use. On the other hand, the projected population data by TPU (from 7th year onward) and SB are classified data with restricted access. PDs interested in such data are recommended to approach WGPD directly, as required. To facilitate data sharing, the projected populations by TPU for the first 7 years (including the base year) will be integrated in the TPU/SB CSU.

2.1.3 To comply with the data privacy law, the population figure for a TPU with an actual population size of less than 50 will not be disclosed. As WGPD opined that the averaging out of projected populations of such TPUs would produce misleading results to the data user, only the aggregated projections of population distribution will be provided for the sparsely populated TPUs in close proximity.

2.2 Scope

2.2.1 For town planning purpose, the whole territory of Hong Kong is hierarchically divided into 9 Primary Planning Units (PPUs) at the first level, 51 Secondary Planning Units (SPUs) at the second level and 282 Tertiary Planning Units (TPUs) at the third level. Each TPU is sub-divided into a number of Street Blocks (SBs) for urban areas and Village Clusters (VC) for rural areas. The hierarchy is summarized below:

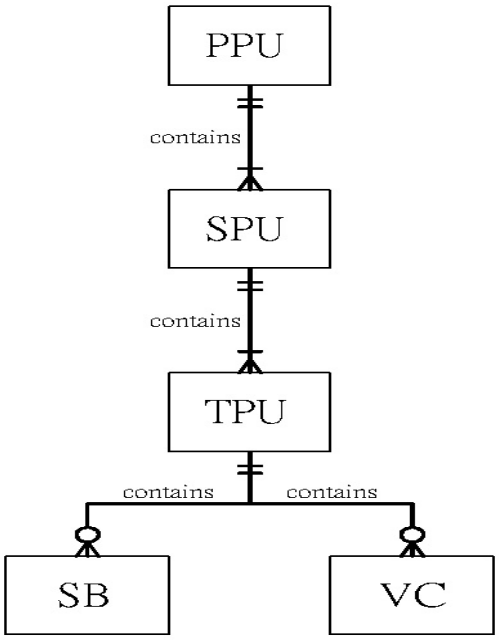


Figure 1 – PPU, SPU and TPU/ SBVC

2.2.2 The scope of TPU/SB CSU includes:

(a) Tertiary Planning Units (TPUs) are the third level of planning units under a geographic reference system demarcated by PlanD for the HKSAR, and

(b) Street Block and Village Cluster (SB_VC) are the fourth level of planning units, which are the smallest planning units in the TPU. SB is the unit used for urban areas and new towns, while VC is used in rural areas within the territory.

2.2.3 Primary and secondary planning units (PPU and SPU, respectively) are larger planning units used by PlanD for similar purposes, but these units will not be included under the TPU/SB CSU. Information referenced to the PPU and SPU could be derived by aggregation and dissolving the information from the TPU layer.

2.2.4 Also, other planning units such as District Council Constituency Area (DCCA) boundary, the Planning, Vision and Strategy (PVS) 2030, the Territorial Population and Employment Distribution Model (TPEDM) data matrix, and the Comprehensive Transport System (CTS) have been reviewed. Since these units serve different purposes, they are not included in the scope of TPU/SB CSU.

2.2.5 TPU/SB is updated at about every five-year interval. Since most of the PDs do not need to have historical TPU/SB information, the archive of TPU/SB boundary will

not be maintained in the CSU scope. Individual PD, e.g. C&SD can continue to maintain the historical TPU/SB information within department.

2.2.6 Within the scope, TPU/SB CSU retains the same definition as is currently adopted by PlanD.

2.3 Common Rules for Delineation and Shapes of Polygons

2.3.1 The current delineation rules and the current mode of maintenance of the geospatial data adopted by PlanD will be retained. In general, the TPU/SB polygons are delineated when the boundaries can be easily referenced by an adjoining feature which could be the boundary of a major development site, road centreline, coastal lines, natural ridges, village clusters, zoning boundaries in the Outline Zoning Plans and boundaries of District Councils. There are two sets of polygons to represent TPU/SB CSU.

(a) TPU polygon

(i) The TPU polygon boundary is determined by the nature of the geographic features in the area. The features include roads, railway lines, coastlines, contours, waterways, lot boundaries and zoning boundaries.

(b) SB_VC Polygon

(i) Generally, the SB_VC polygon in urban areas and new towns is defined along the grids of streets. In urban areas or new towns in the New Territories, each TPU is sub-divided into a number of SBs, usually following grids of streets. In rural areas, the remaining TPUs are sub-divided into VCs, following natural boundaries such as roads/streets and streams.

2.4 Common Attributes

2.4.1 The spatial data and non-locational (textual) attributes relevant to the TPU/SB CSU will be selected to form a set of common CSU data attributes.

2.4.2 Table 1 below summarizes the common attributes requested by the PDs.

Spatial Boundary
TPU & SBVC Polygon
Textual Attribute
TPU

a. TPU Number
b. TPU Area
SB_VC
a. SB_VC Number
b. TPU Number
c. Area
TPU Mapping
a. TPU ID
b. Population ID
c. Base Year
Projected Population
a. Population ID
b. Base Year
c. Base Year Projected Population
d. First Year Projected Population
e. Second Year Projected Population
f. Third Year Projected Population
g. Fourth Year Projected Population
h.. Fifth Year Projected Population
i. Sixth Year Projected Population

Table 1 Common attributes of TPU/SB CSU

2.5 CSU ID

2.5.1 The TPU ID and the composite key (TPU ID and SBVC No) currently established by PlanD will be adopted as the CSU ID for TPU polygon and SBVC polygon respectively.

2.5.2 Each TPU is identified by a unique 3-digit number (i.e. 133). The first digit identifies which Primary Planning Unit (PPU) the TPU belongs to, while the first and second digits together correspond to the Secondary Planning Unit (SPU) code. The 3-digit identifier is the TPU ID.

2.5.3 The SBVC polygons are uniquely identified using a combination of the TPU ID and the SBVC No, to form the SBVC ID. The SBVC No is currently assigned by a 2-digit code. In order to cater for future need and maintain data uniqueness, a 3-digit code will be adopted as the SBVC No. Hence, the SBVC ID would have 6 digits.

2.5.4 For the SBVC ID of SBVC polygon – 144161, as an example:

(a) "1" represents the PPU ID;

(b) "14" represents the SPU ID;

(c) "144"represents the TPU ID;

(d) "161" represents the SBVC No.

2.6 Data Custodianship

2.6.1 Please refer to Volume 2F – Data Custodianship and License Agreement, for details about the defined roles and responsibilities for the Data Agent, Data Owner and Data User.

Data Agent

2.6.2 PlanD is the Data Agent responsible for assigning CSU ID and data dissemination of TPU/SB CSU.

Data Owner

2.6.3 PlanD is the only Data Owner contributing all the attributes of TPU/SB CSU[1].

Data User

2.6.4 Data Users of TPU/SB CSU include: BD, C&SD, DSD, LandsD, PlanD, RVD and TDD.

[1] Source of projected population data is originated from WGPD. With consent from WGPD, PlanD is delegated to be the Data Owner of projected populations data.

3 CSU Workflow

3.1 Overview

3.1.1 The flow charts below are used to describe the data exchange processes among the PDs in the context of TPU/SB CSU. Only those processes that are directly relevant to update or retrieval on CSU data are indicated. Internal processes within a PD, and data exchange processes between a PD and other organization (e.g. developers, government departments other than the PDs) are not included.

3.2 Data Exchange Processes

3.2.1 The proposed workflow process will cover the data exchange in the Preparation Stage. It is illustrated by the chart below:

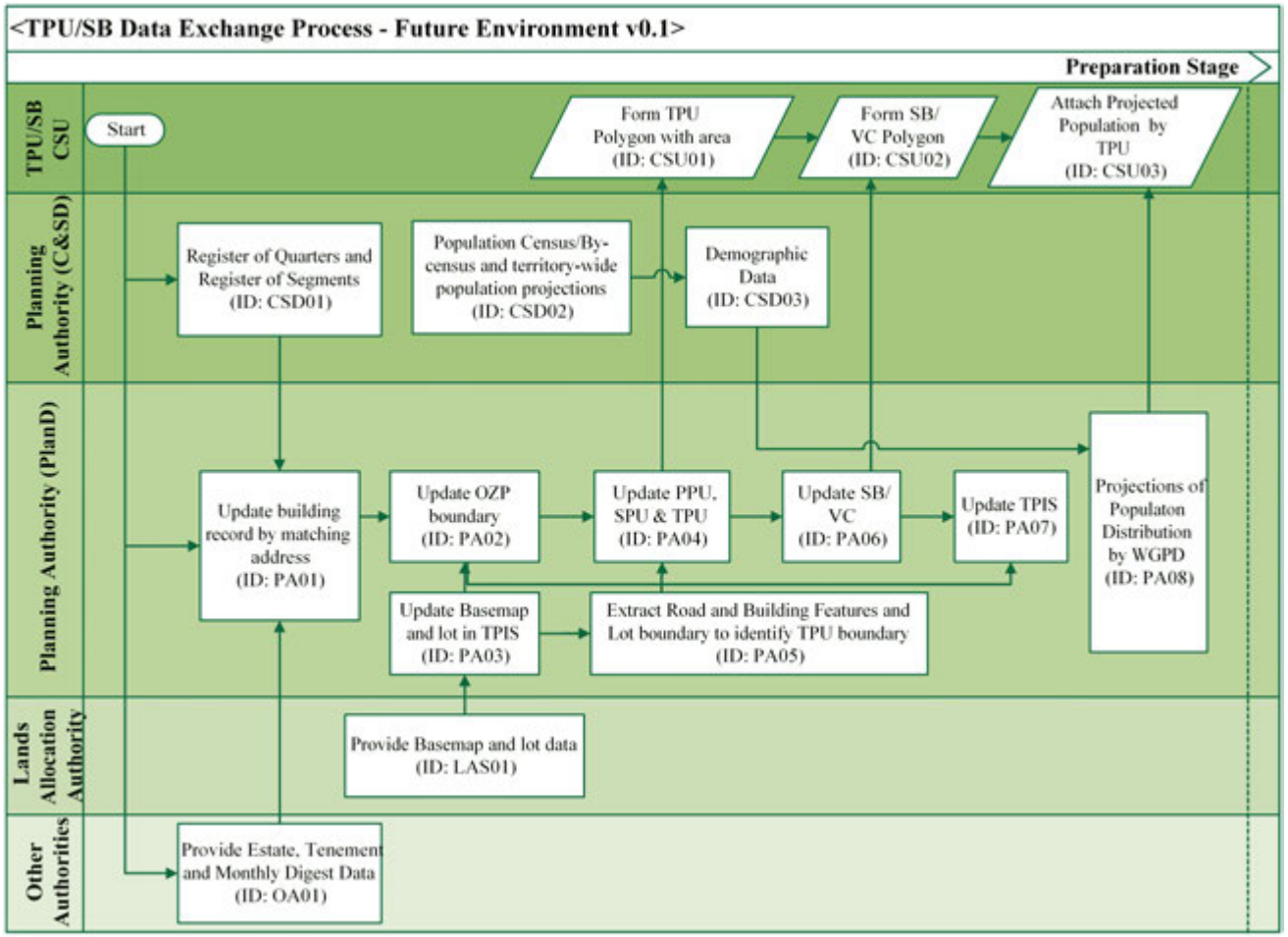


Figure 2 Workflow of TPU/SB CSU – Preparation Stage

3.2.2 Development of TPU/SB would rely on supporting information from several sources including data from C&SD, base map and lot boundary data from LandsD, and estate, tenement and monthly digest data supplied by other BD and other PDs (OA01).

3.2.3 PlanD extracts the required information from these sources and updates the building record by matching addresses. PlanD makes use of data from Register of Quarters and Register of Segments from C&SD to keep its building records updated (CSD01, PA01). PlanD uses the building records to update the Outline Zoning Plan (OZP) (PA02).

3.2.4 Also, PlanD requests updated base map and lot data from LandsD to keep their Town Planning Information System (TPIS) up to date on the base map information (LAS01, PA03). PlanD extracts road and building features from the TPIS base map and lot boundaries from the LandsD's lot data in order to help form TPU boundaries (PA05).

3.2.5 The delineation of a boundary is based on the alignment of major developments, road centerline, coastal lines, natural ridges, village clusters, zoning boundaries in the Outline Zoning Plans and boundaries of District Councils. At every five years interval, before population census and by-census, the PPU, SPU and TPU boundaries will be updated. The TPU polygons with their respective areas are now produced (CSU01).

3.2.6 Each TPU boundary is further sub-divided into Street Blocks or Village Clusters (SB/VC) and their boundaries would be updated by making use of OZP boundaries from updated building records, and updated TPIS base map and lot data with information supplied by LandsD (PA07, CSU02). The TPU and SBVC polygons are updated at every five years interval.

3.2.7 On the basis of the census/by-census figures (which would be chosen as the benchmark in the base year), C&SD would update mid-year and end-year population figures, made on territory-wide level by age and by sex. Such mid-year population projections figures would then be provided to the Working Group on Population Distribution Projections for compiling estimates of Population Distributions (CSD03, PA08). The projected populations by TPU are produced (CSU03) yearly.

3.2.8 Having received the projected populations from WGPD, the Data Agent, PlanD, would update the mapping table between the TPU ID and the Projected Population ID, and assign the base year information to the TPU Mapping and the Projected Population entities. Finally, the TPU mapping and the Projected Population entities would be released on a yearly basis.

4 CSU Data Interface Requirement

4.1 Overview

4.1.1 A logical model for CSU data exchange is defined for PDs' exchange of the TPU/SB CSU. It describes the logical structure of CSU data exchanged between the interfacing systems of PDs. PDs, as either Data Owners or Data Users, are not required to adopt the same logical data structure in their own departmental systems. However, each PD is recommended to maintain a mapping between the Common Spatial Units and their departmental records in their respective core departmental systems(s) such that:-

(a) Data Owner can extract data from her departmental system according to the definition of CSU; and

(b) Data User can import CSU data into her departmental system for further processing/ analysis.

4.1.2 Mapping of the unique CSU ID with the departmental ID may be a one-to-one, one-to-many or many-to-one relation (but a many-to-many relation is not recommended) depending on the data definition of PDs' departmental records.

4.1.3 For TPU/SB CSU, some Data Users have already adopted the TPU number in their departmental system. Effort required in performing this mapping for these Data Users would be negligible.

4.1.4 The following sections will provide details for the logical model in terms of :-

(a) CSU status – possible statuses of a CSU;

(b) Themes - thematic layers; and

(c) Logical data structure, including a detailed description of the data items and assignment of ownership.

4.2 Entity Description

4.2.1 Some data items may be left as null due to time lag, but must be filled once the CSU is completely defined. Such rules will be described in the Description column of the affected data item. Please also refer to section 3 for more information on the CSU lifecycle.

4.2.2 Some data items are applicable for particular records only. For non-applicable CSUs, such data item(s) are always left as null. For applicable CSUs, such data item(s) is left null for a particular CSU only when the information is not yet available

4.3 CSU Status

4.3.1 The TPU/SB is updated at about every five years. Since most of the PDs do not need to have historical TPU/SB information, archive of TPU/SB boundary will not be maintained in the CSU scope. Only active status for the TPU/SB CSU is available.

4.4 Themes

4.4.1 According to the current mode of maintenance of the TPU/SB CSU geospatial data by PlanD, the TPU/SB CSU encompasses two thematic layers:

(a) TPU Polygon layer – a geospatial layer representing the latest version of tertiary planning units.

(b) SBVC Polygon layer – a geospatial layer representing the latest version of Street Block (SB) and Village Cluster (VC) units.

4.4.2 All polygons in the above layers are closed polygons. The polygons are two-dimensional. The polygons represent geographic features stored as a series of segments that enclose an area. No overlapping polygon within a single layer is allowed.

4.4.3 For each individual polygon, it should not be split even if the polygon lies along the tile border of a survey sheet.

4.4.4 For the full set of polygons, it is recommended that they should be maintained in a seamless manner, that is, the splitted polygon located along the tile border has to be merged, and the border line has to be dissolved, thereby, other PDs' post processing effort on merging the polygons can be minimized, also, this facilitates the spatial query and filtering operation.

4.4.5 The polygon layer should be confined within the HK 1980 Grid Coordinate System, that is, the minimum spatial extent and the maximum spatial extent are 800000, 800000 and 867500, 848000 respectively.

4.4.6 It is recommended to establish a topological relationship that defines the behaviour of features, when resource is available. This would define the rule for features to share geometry with other features (e.g. TPU cannot overlap) in a given layer or between multiple layers (e.g. the SBVC polygon should completely fall within the TPU polygon.)

4.5 Logical Data Structure

4.5.1 Please refer to Appendix A for conventions used in this section.

4.5.2 The logical relationships among entities are illustrated below.

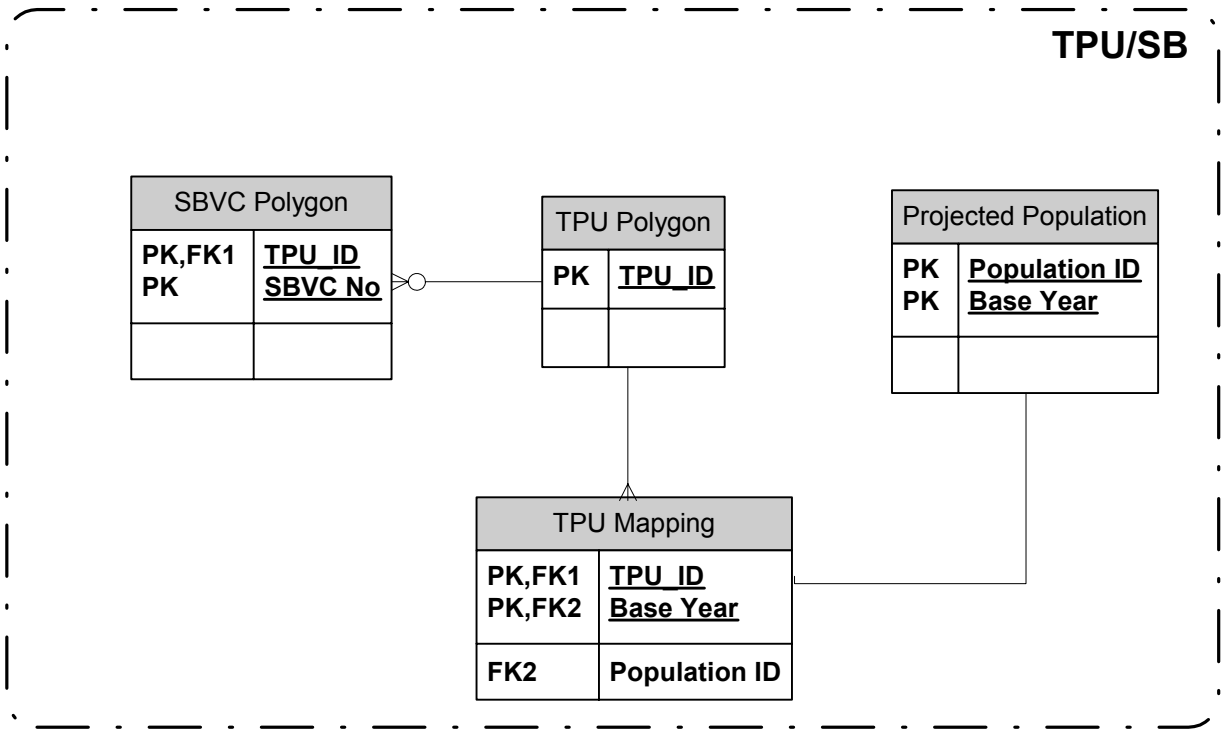


Figure 3 LDS for the TPU/SB CSU

4.6 Entity Description

4.6.1 TPU Polygon

– Spatial feature representing the latest version of Tertiary Planning Unit boundaries, which is updated at about every five years interval.

(a) Data Item Description

Data Item	Description	Format	Mandatory
TPU ID	Unique identifier for the TPU features	N(3)	Y
Geometry	Geometry of TPU polygon	Polygon type	Y
TPU Area	The TPU area is derived from the TPU's geometry in square meter unit.	N(14, 6)	Y

(b) Constraint Description

Type	Data Item	Validation	Reference Entity	Reference Data Item
PK	TPU ID			
CP	Geometry			

(c) Data Ownership

All records	PlanD

4.6.2 SBVC Polygon

– Spatial feature representing the latest version of Street Block and Village Cluster boundaries, which are updated at about every five years interval.

(a) Data Item Description

Data Item	Description	Format	Mandatory
TPU ID	Unique identifier for the TPU features	N(3)	Y
SBVC No	Identifier for the SBVC feature in the corresponding the TPU	N(3)	Y
SBVC Area	The SBVC area derived from the SBVC 's geometry in square meter unit.	N(14, 6)	Y
Geometry	Geometry of SBVC polygon	Polygon Type	Y

(b) Constraint Description

Type	Data Item	Validation	Reference Entity	Reference Data Item
PK	TPU ID, SBVC No			
FK	TPU ID		TPU Polygon	TPU ID
CP	Geometry			

(c) Data Ownership

All records	PlanD

4.6.3 TPU Mapping

– A textual entity containing the relationship between the TPU ID and the Population ID, which is used to cater for the many-to-one relationship between TPU and the projected population for a particular base year. Since the projected population and its corresponding Population ID are updated yearly, to ensure the data uniqueness, a composite key (Population ID + Base Year) is used to link with the Projected Population entity.

(a) Data Item Description

Data Item	Description	Format	Mandatory
TPU ID	Unique identifier to identify the TPU features	N(3)	Y
Population ID	Unique identifier to identify the projected population stored in the Projected Population entity. For ease of understanding and minimize the updating work, it is suggested to adopt the TPU ID as the Population ID for those TPU which has one-one relationship with the Projected Population entity, whereas for a TPU which has many-to-one relationship with the projected population, a new ID (with 4-digit) is assigned.	N(4)	Y
Base Year	The base year of population projection, e.g. 2002	N(4)	Y

(b) Constraint Description

Type	Data Item	Validation	Reference Entity	Reference Data Item
PK	TPU ID Base Year			
FK	TPU ID		TPU Polygon	TPU ID
FK	Population ID Base Year		Projected Population	Population ID

(c) Data Ownership

All records	PlanD

4.6.4 Projected Population

– A textual entity containing the projected population by TPU for the first 7 years (including the base year). Since the projected population for the 7th year and onwards is classified as restricted data, and can only be made available to PDs under restrictive status, thus, they are outside the context of the TPU/SB CSU. For ease of use, the field name for the Projected Years is recommended to contain the 4-digit year information. Population figures are rounded to the nearest 100. Due to such rounding, figures may not add up to the respective totals. This entity is produced and released annually.

(a) Data Item Description

Data Item	**Description**	**Format**	**Mandatory**
Population ID	Unique identifier to identify the projected population stored in the Projected Population entity. To ease of understanding and minimize the updating work, it is suggested to adopt the TPU ID as the Population ID for those TPU which has one-one relationship with the projected population entity, whereas for those TPU which has many-to-one relationship with the projected population, a new ID (with 4digits) is assigned.	N(4)	Y
Base Year	The base year of population projection, e.g. 2002	N(4)	Y
Base Year Projected Population	The projected population by TPU for the Base year. 8 digits are allocated for projected population to cater for future needs.	N(8)	Y
First Year Projected Population	The projected population by TPU projected from the Base year population for the 1st year. 8 digits are allocated for projected population to cater for future needs.	N(8)	Y
Second Year Projected Population	The projected population by TPU projected from the Base year population for the 2nd year.	N(8)	Y

Data Item	Description	Format	Mandatory
Third Year Projected Population	The projected population by TPU projected from the Base year population for the 3rd year.	N(8)	Y
Fourth Year Projected Population	The projected population by TPU projected from the Base year population for the 4th year.	N(8)	Y
Fifth Year Projected Population	The projected population by TPU projected from the Base year population for the 5th year.	N(8)	Y
Sixth Year Projected Population	The projected population by TPU projected from the Base year population for the 6th year.	N(8)	Y

(b) Constraint Description

Type	Data Item	Validation	Reference Entity	Reference Data Item
PK	Population ID Base Year			

(c) Data Ownership

All records	PlanD

5 Maintenance of the CSU

5.1 Data Provision Frequency

5.1.1 In general, Data Owners should provide the latest changes on CSU relevant data to the Data Agent on a regular basis. All changes since the last data extraction must be included in the interfaced data file in each data uploading.

5.1.2 PlanD is the only Data Owner contributing data to TPU/SB CSU. The same arrangement will apply to PlanD.

5.1.3 Currently, the spatial boundary of TPU/SB is updated every five years and the updated data are released a year before conducting the Survey of Census and By-Census. For TPU/SB CSU, its spatial boundary would be identical to that of the TPU/SB currently maintained by PlanD. The same updating frequency, i.e. 5 years shall be retained for the spatial boundary of TPU/SB CSU.

5.1.4 With respect to the textual attributes, PlanD will disseminate the first 7 years of population projections (by TPU), related mapping table, together with the TPU spatial boundary on yearly basis (September of the year).

5.2 Data Dissemination Frequency

5.2.1 Data downloading from the Data Agent to Data Users will be supported regularly. PlanD, as the Data Agent, will pre-pack the CSU data according to the recommended File Format Standard (from DAM3) and make the packed files available for download by Data Users in a pre-defined location. Details of disseminating CSU will be discussed in Section 5.3.1.

5.2.2 PlanD is the only Data Owner and also PlanD is the Data Agent of the TPU/SB CSU. The data provision frequency could be made equal to that of the data dissemination frequency.

5.2.3 Since the data attributes of TPU/SB CSU are relatively static, members (Data Users) agreed that this TPU/SB CSU data will be updated as per the following frequency schedule:

	Data Provision Frequency (also Data Dissemination Frequency)
TPU & SBVC Polygon	Every 5 Year
TPU	

	Data Provision Frequency (also Data Dissemination Frequency)
c. TPU Number	Every 5 Year
d. TPU Area	Every 5 Year
SB_VC	
d. SB_VC Number	Every 5 Year
e. TPU Number	Every 5 Year
f. Area	Every 5 Year
TPU Mapping	
a. TPU ID	Yearly
b. Population ID	Yearly
c. Base Year	Yearly
Projected Population	
a. Population ID	Yearly
b. Base Year	Yearly
c. Base Year Projected Population	Yearly
d. First Year Projected Population	Yearly
e. Second Year Projected Population	Yearly
f. Third Year Projected Population	Yearly
g. Fourth Year Projected Population	Yearly
h.. Fifth Year Projected Population	Yearly
i. Sixth Year Projected Population	Yearly

Table 1 Updating Frequency of data attributes of TPU/SB CSU

5.3 Mode of Dissemination

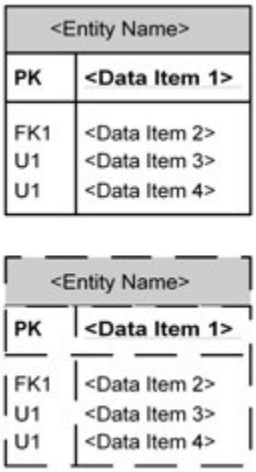
5.3.1 PlanD is required to prepare the data conforming to the CSU logical data model, and then convert them into the standard file formats (recommended in Final DAM3) prior to data dissemination.

5.3.2 The ultimate solution for disseminating the TPU/SB CSU is via the department portal programme (DPP) of PlanD, which will be implemented by mid 2004. This portal will be launched as an Intranet portal with authentication and other advanced features such as download by criteria. Instead of collecting the data by CD-ROM (current practice), Data Users can access the Intranet and download the CSU data from this single source. Only the authorized government officers can browse, search or download TPU/SB CSU information from the website.

5.3.3 To streamline future data exchange process, PDs are encouraged to use the new dissemination facility, thus the existing channel(s) through written requests will no longer be supported. However, should PDs require population data (7+ years), which is classified as sensitive data, and not incorporated in the scope of the TPU/SB CSU, such requests, are still supported by PlanD on individual request.

Appendix A. Conventions for Data Interface Requirement

A.1.1 Logical Data Structure Diagram

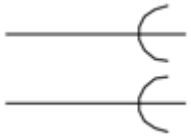


Entity
The upper part in grey shading shows the name of the entity. The lower part lists only the data items involved in the primary key, unique key(s) and foreign key(s) of the entity, while the other data items of the entity will be elaborated in Entity Description. Composite keys are represented by same key name in multiple data items (e.g. two data items marked as "U1" means a composite unique key composed of two data items)
Mandatory data item(s) will be printed in bold text.
For diagrams spanned across multiple pages, the first occurrence of each entity is shown in solid-line border while all repeated occurrences in later pages are shown in dotted-line border.

Line with crow's foot
"many" end of a relation

Normal solid line
Mandatory end of a relation

Line with small circle at the end
Optional end of a relation. That is, a record on the left may not have any associated record on the right.



Line with a curve at the end
Exclusive relation. i.e. only 1 among the connected entities on left is linked to the entity on right at a time

A.1.2 Entity Description - Data Item Description - Format

(a) X(n) : Variable-length character strings of max. length n

(b) CX(n): Variable-length character strings of maximum n Chinese characters. Maximum number of bytes will depend on the character set used by the CSU dataset. E.g. CX(5) occupies at most 10 bytes if data are stored in Big5 character set.

(c) N(m,n) : Fixed and floating point numbers, where m is precision (total number of digits before and after decimal point) and n is scale (number of digits to the right of decimal point). The n part is omitted, i.e. denoted in N(m), for integers.

(d) Date : Point-in-time values (date and time)

(e) Polygon : Closed polygon representing a spatial area.

A.1.3 Entity Description - Constraint Description - Type

(a) PK : Data item is part of primary key, which is used to uniquely identify a record in the entity.

(b) FK : Data item is part of foreign key, which means the data item values, if not null, must match the unique identifier of another entity.

(c) UK : Data item is part of alternate key, which is used as an alternate way to uniquely identify a record in the entity.

(d) CK : The value of data item should be checked ensuring that it falls within or meets the predefined values/ranges/rules. Hence, non-mandatory data items can be left as null, or otherwise must meet the criteria.

(e) CP : The value of geometry type data item should be a closed polygon.

A.1.4 Entity Description - Data Ownership - Condition

(a) RelatedEntity.DataItem : Reference to value of DataItem of the linked RelatedEntity record. For example, "CSU Feature.Status" means the Status data item of the related CSU Feature record. Unless specified, relation and linked key is determined by the concerned FK constraint as defined in Constraint Description.