Measures to Facilitate Address Data Capture and Validation (Paper No. 2)

Introduction

1. Address CSTF Paper 2004/04 has set out our strategy for facilitating address validation. The strategy relies on data to be contributed by the Lands Department (LandsD) and the Rating and Valuation Department (RVD) under the Data Alignment Measures (DAM) project, plus floor and unit information from RVD's departmental address database. LandsD and RVD are currently designing their approach to implement the DAM recommendations.

2. This paper serves to walkthrough the mechanism of :
   - address validation;
   - spatial analysis based on validated addresses;
   - exchange of validated and fail-to-be-validated address information

   from an application's point of view, and lays down the assumptions underlying these mechanisms with a view to align common understanding on how the proposed mechanism may work. This helps assure that the development work to be undertaken by LandsD[1], RVD[2] and the Office of the Government Chief Information Officer (OGCIO)[3] will fit together in the future.

3. In addition, this paper suggests a mapping to the address standard being developed by the Organization for the Advancement of Structured Information Standards (OASIS).

4. In the context of this paper, validating a captured address means checking the

---

[1] *In the context of address validation, the building address data to be contributed by LandsD under the DAM project will be used to form part of the Common Address Database.*

[2] *In the context of address validation, the building address data to be contributed by RVD under the DAM project, together with floor and unit information of the corresponding buildings from RVD's address database, will be used to form part of the Common Address Database.*

[3] *OGCIO is bidding for government funds to prototype a common service for address data capture and validation. This prototype service will utilize data in the Common Address Database.*

captured address against a comprehensive set of local addresses (i.e. a local address database) to better assure that the captured address refers to a location that physically exist. The accuracy of such checking depends on the quality (e.g. timeliness) of the data in the address database. (Another interpretation of "validating an address" is checking whether a person genuinely resides at the reported location, or whether an organization genuinely exists at the reported location. This interpretation is not applicable in the context of this paper.)

5. Readers are encouraged to read Address CSTF Paper 2004/04 in conjunction with this paper. Paper 2004/04[4] sets out our strategy on how to facilitate address data capture and validation.

6. Readers are also encouraged to read Volume 2B[5] of the Final Report on Implementation of Data Alignment Measures (DAM) for the Alignment of Planning, Lands and Public Works Data. Volume 2B contains specification and explanatory notes of the Building Common Spatial Unit (CSU). Sections 4.5.15 - 4.5.21 are relevant in particular.

The Address Validation Process

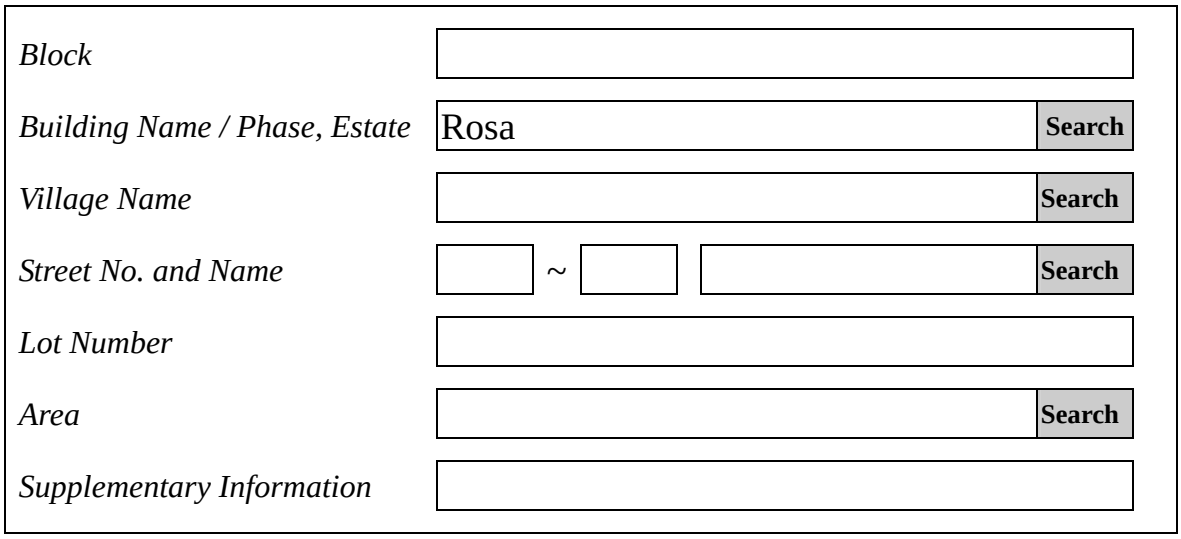
7. The mechanism described below is one of the many ways to perform address data capture and validation. The described mechanism is for illustration purpose only and may not correspond to the way address validation is to be implemented in the future. The address validation mechanism to be designed should be based on good understanding of the address related information to be associated with the DAM Building CSUs.

8. One way to offer the address data capture and validation service is to operate a 7x24 service where authorized applications can redirect their clients (e.g. a PC accessing a Web application offered by a B/D) to when an address needs to be captured through the client. This approach relies on the URL redirection mechanism used by some payment gateways. By implementing this as a 7x24 callable service, we can hide the complexity of the address validation

[4] *http://www.xml.gov.hk/download/address_cstf_200404.pdf*

[5] *http://www.hplb.gov.hk/eng/publication/pdf/frv2B.pdf*

logic from applications. All that applications need to know is how to call this service and what parameters the service will return. When the calling application needs to capture and validate an address through the client, it will redirect the client to the address validation service which displays a window on the client to prompt for the building address. For example :



9. The user enters “Rosa” against “Building Name / Phase, Estate” and presses the associated “Search” button. The address validation module searches the common address database’s “Building Address” table, “Building Development Name / Development Name” table and “Building Name” table for all buildings and estates whose name contain “Rosa” and displays their addresses for the user to select. The user selects the address “Block E, Villa Monte Rosa, 41A Stubbs Road”.

10. The address validation service displays the full building address and further prompts for floor and unit (flat) information which cannot be validated until our common address database is equipped[6] to validate 3-D addresses. Upon user confirmation, the address validation service returns to the calling program the following information :
    - DAM CSU identifier = the CSU identifier of the selected building
    - Area = Hong Kong (derived by the address validation module based on a Building CSU to Area mapping table that can be generated periodically by

[6] *Address CSTF Paper 2004/04 proposed to validate floor and unit information against RVD’s property unit information which can be linked to the DAM Building CSUs after RVD has done a mapping between RVD’s building records and the DAM Building CSUs*

overlaying the DAM building polygon layers on top of a new "Area" layer where "Area" refers to Hong Kong, Kowloon, New Territories, Lantau Island, Cheung Chau, Ping Chau, etc.)
- Street = Stubbs Road
- Start building number = 41A
- Estate = Villa Monte Rosa
- Block = Block E (derived by manipulating RVD's data in the "Building Address" table)
- Floor descriptor = Floor
- Floor number = 18
- Unit descriptor = Flat
- Unit number = A
- Validation indicator = Validated up to building level

11. The calling program stores all of the above captured information, regardless of whether the address can be successfully validated or not.

12. If the concerned location is addressed by lot number, then it may not be possible to validate its address despite the fact that the full address from RVD in the "Building Address" table contains the lot number, owing to the inconsistency in lot number descriptions.

13. The above is just one of the many ways to do address validation. We can also consider making use of address aliases to facilitate the search of address components. For example, the user enters the Cantonese pronunciation " Lai Kong Sing" and the address validation module prompts if he means "Laguna City", etc. This approach will require additional alias data on top of the DAM building addresses.

14. Existing address validation applications like ESD's "Change of address" transaction can continue to make use of the building addresses associated with LandsD's B1000 maps if they do not need the flexibility[7] offered by RVD addresses, and they do not wish to benefit from the addresses' association with

---

[7] *The RVD addresses in DAM have finer breakdown. For example, "Whampoa Garden (Site 2) Cherry Mansions Block 2" on the B1000 map is broken down into the following components in RVD's address format :*
- *Building name line 1 = Block 2*
- *Building name line 2 = Cherry Mansions*
- *Building name line 3 = Whampoa Garden*

persistent DAM Building CSU identifiers.

15. To summarize, we choose to have our common address database based on DAM addresses because we want to benefit from the addresses' linkage with persistent DAM Building CSU identifiers, as well as the identifiers' linkage with the DAM building polygons. These linkages will facilitate spatial analysis. We will elaborate this point in the next section.

16. The assumption made here is that RVD[8] and LandsD[9] will contribute address data against the DAM buildings as far as possible. ***[Assumption No. 1]***

17. Our common address database can also be based on other address databases provided that the address database has a good coverage of person and organization addresses and its address data are attached with a DAM CSU identifier where applicable. In view of the Property Information Hub (PIH)[10] initiative, it is possible that the PIH operator may offer an address validation service based on "enhanced" PIH addresses if the operator sees a business case in doing this. Nevertheless, government bureaux and departments (B/Ds) are concerned about the cost of using address validation service through the PIH.

Spatial Analysis Based on Validated Addresses

18. To facilitate spatial analysis, an application should record the DAM CSU identifier in addition to the textual address details. To do spatial analysis, a GIS application simply uses the DAM CSU identifier to locate the DAM building or other relevant spatial object.

19. The assumption made here is that the DAM building polygons and their addresses and CSU identifiers can be made available to B/Ds. If we want to facilitate the private sector to use the same address validation and spatial

---

[8] *The addresses to be contributed by RVD, which contain building and estate names, will reside in the "Building Address" table*

[9] *The addresses to be contributed by LandsD will be distributed in the "Development Name"/"Building Development Name", "Building Name" and "Building Address" tables*

[10] *http://www.info.gov.hk/rvd/content/pdf/eoi.pdf*

analysis mechanism in the future, then the DAM building polygons, their CSU identifiers and the common address database should also be made available to the public. ***[Assumption No. 2]***

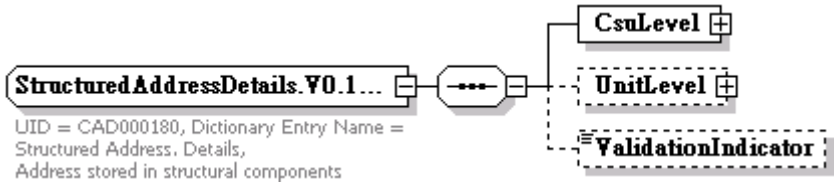
20. Unless there is a major system revamp of LandsD's Computerized Land Information System (CLIS), the building polygons on LandsD's B1000 maps currently sold to the public cannot be aligned with the DAM building polygons.

21. Comparing this with the traditional way of doing spatial analysis, which uses an absolute x,y coordinate to associate objects (e.g. a dangerous goods storage location) on a map, this approach of using the CSU identifier has the advantage of attaching objects permanently to a CSU, rather than an absolute x,y. If, due to one reason or another, the x,y of the CSU is shifted, all objects associated with that CSU will be shifted simultaneously.

22. Due to operation need, some B/Ds have created their own spatial objects by modifying those provided by LandsD. Since the DAM project has aligned a number of CSUs, these CSUs should be used by B/Ds as far as possible. In other words, B/Ds should create their own spatial objects only when they are not covered in DAM, and as DAM evolve to cover more CSUs, B/Ds are recommended to switch to use the DAM CSUs, where appropriate, instead of the spatial objects they created earlier

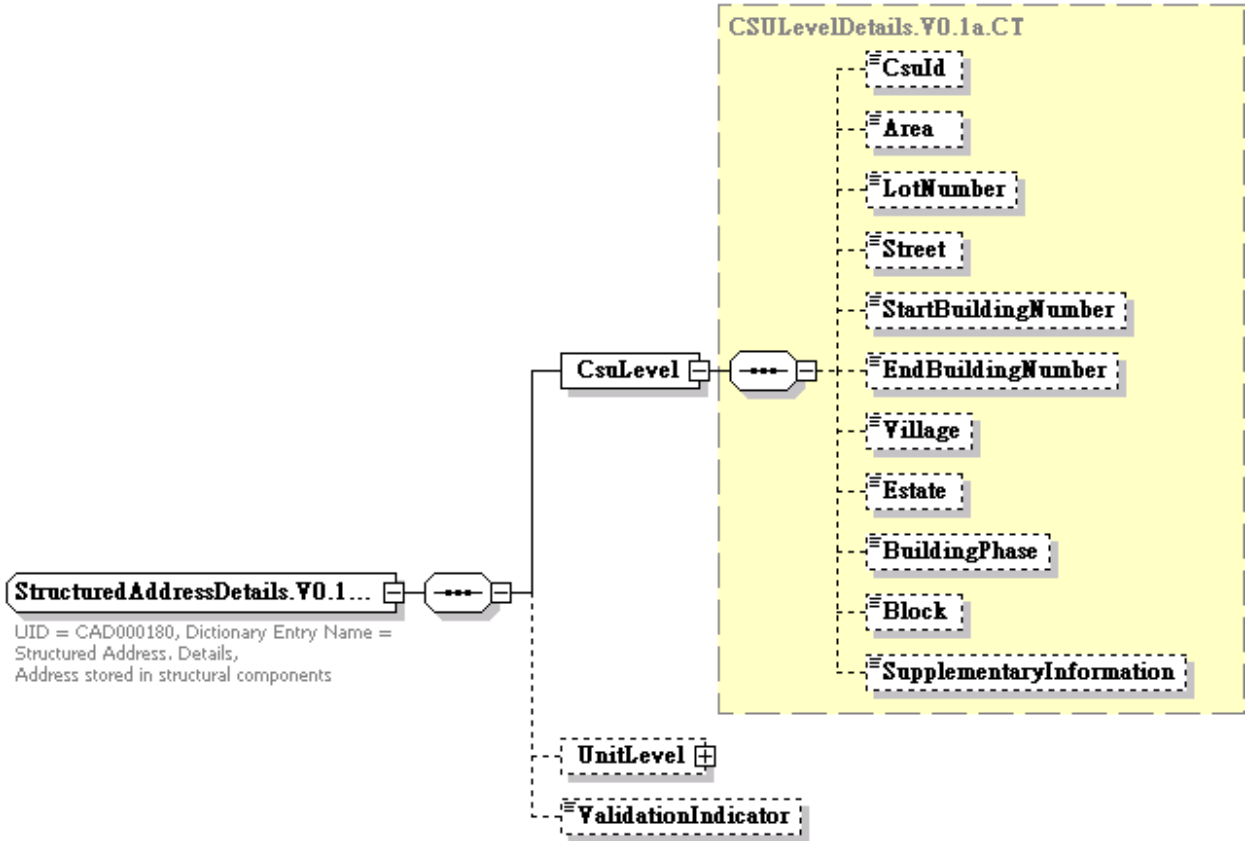
Exchange of Validated and Fail-to-be-validated Address Information

23. The data structure for exchanging structured address is suggested to comprise a CSU level description, a unit level description, plus a validation indicator. The CSU level should be pitched at a building (or other relevant DAM CSU) level. The unit level then describes the floor and unit information. The validation indicator indicates whether the address is :
    - Not validated;
    - Validated up to building level;
    - Validated up to building and floor level; or
    - Validated up to building, floor and unit level.

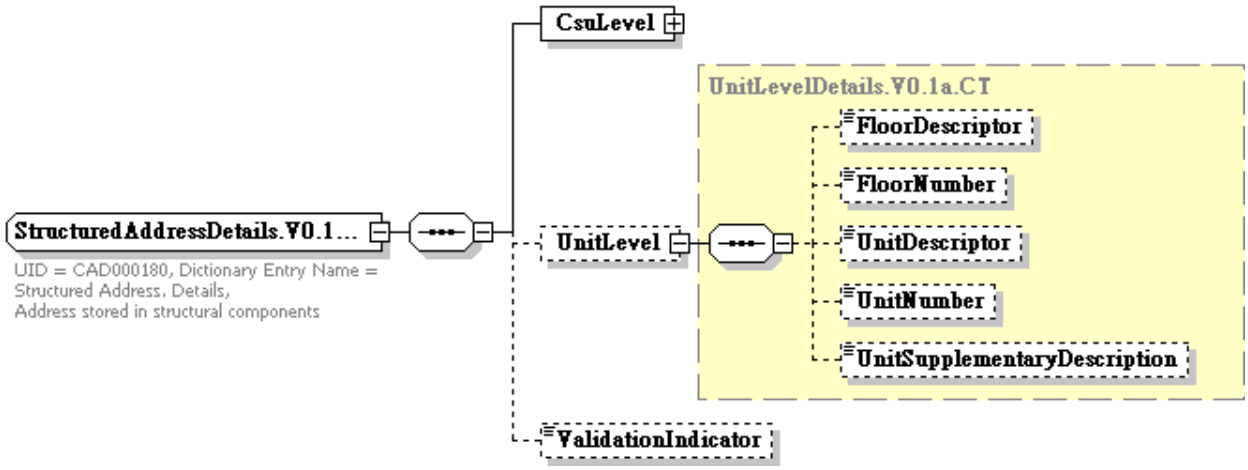


24. The CSU level description is suggested to comprise the following data elements (which include the full textual building address[11], in addition to the CSU identifier, if applicable) :
    - DAM CSU identifier (applicable for validated address only)
    - Area
    - Lot number
    - Street
    - Start building number
    - End building number
    - Village
    - Estate
    - Building / Phase
    - Block
    - Supplementary information

[11] *This point is different from that proposed in Address CSTF Paper 2004/01.*



25. The unit level description is suggested to comprise the following data elements :
    - floor descriptor (e.g. Basement Level 1, Floor)
    - floor number
    - unit descriptor (e.g. Shop, House, Boat parking lot, Flat, Room, Suite)
    - unit number (e.g. B1-20)
    - unit supplementary description (e.g. Rear)



26. The data structure suggested above should be reviewed when we prototype the address data capture and validation service.

The Mapping from a Local Textual Address to xAL

27. OASIS, through its Customer Information Quality Technical Committee, is defining an XML standard for address data management called eXtensible Address Language (xAL)[12]. xAL is designed to encompass all types of addresses in the world. xAL data elements can represent the hierarchical structure used in addresses. For example, if we treat a multi-storey building as a "premise", then each floor is a sub-premise within the building premise, and each unit on the same floor is a sub-premise within a floor sub-premise. For some data elements, xAL offers flexibility in the hierarchy of the data elements, e.g. :
    - A thoroughfare can be under a dependent locality (for cases like "the Fifth Street in Hong Lok Yuen"); or
    - A dependent locality can be under a thoroughfare (for cases like "Villa Monte Rosa on Stubbs Road).

28. We could have adopted xAL when we exchange local addresses between B/Ds. However, since xAL is flexibly designed to encompass all types of addresses in the world, its structure and vocabulary are generalized and does not resemble what we often see in local addresses. The structure for exchanging address suggested in the previous section uses more localized terminology.

29. Some industries may incorporate xAL into its industry standard in the future. To facilitate B/Ds to map local addresses to xAL, we suggest a mapping in this section.

30. A local address can be mapped to the Locality element of xAL. The mapping of CSU level descriptions is as follows:

[12] *http://www.oasis-open.org/committees/ciq/ciq.html#6*

Data element in a local address	Corresponding xAL element	Remarks
Area	Locality	
Lot number / Village / Estate and Phase within an estate	DependentLocality	"Lot number", "estate" and "village" will not appear simultaneously in a local address
Street	Thoroughfare	If both start and end building number are present, then they map to a ThoroughfareNumberRange sub-element. If only the start building number is present, it is mapped to a ThoroughfareNumber sub-element instead
Start building number		
End building number		
Building / Block	Premise	
Supplementary information	AddressLine	

31. The mapping of unit level descriptions is as follows:

Data element in a local address	Corresponding xAL element	Remarks
Floor descriptor	SubPremise (at one level below premise)	"Type" attribute of the SubPremise element
Floor number		SubPremiseNumber of the SubPremise element
Unit descriptor	SubPremise (at one level below premise for single-storey premise, otherwise at one level below the floor subpremise)	"Type" attribute of the SubPremise element
Unit number		SubPremiseNumber of the SubPremise element
Unit supplementary description		SubPremiseNumberSuffix of the SubPremise element

32. The earlier example on "Flat A, Floor 18, Block E, Villa Monte Rosa, 41A Stubbs Road, Hong Kong" translated in xAL using the above mapping is illustrated below for reference :

```
<AddressDetails>
    <Locality>
        <LocalityName Type="Area">Hong Kong</LocalityName>
        <Thoroughfare Type="Street">
            <ThoroughfareNumber>41A</ThoroughfareNumber>
            <ThoroughfareName>Stubbs Road</ThoroughfareName>
            <DependentLocality Type="Estate">
                <DependentLocalityName>Villa Monte Rosa</DependentLocalityName>
                <Premise Type="Block">
                    <PremiseName>Block E</PremiseName>
                    <SubPremise Type="Floor">
                        <SubPremiseNumber>18</SubPremiseNumber>
                        <SubPremise Type="Flat">
                            <SubPremiseNumber>A</SubPremiseNumber>
                        </SubPremise>
                    </SubPremise>
                </Premise>
            </DependentLocality>
        </Thoroughfare>
    </Locality>
</AddressDetails>
```

Summary

33. The address validation mechanism illustrated in this paper shows how the address data in DAM can help address validation. The validated addresses will lay a good foundation to facilitate spatial analysis. Assumptions no. 1 and 2 must hold for these mechanisms to function.

34. Detail design of the address validation mechanism must be based on good understanding of the address related data to be associated with the DAM Building CSUs.

35. The data structure for exchanging structured address drafted in this paper should be reviewed when we prototype the address validation mechanism.

Comment Sought

36. Members are requested to comment on the address validation, spatial analysis and data exchange mechanisms suggested in this paper.

Office of the Government Chief Information Officer
December 2004