

A Study on the Need of Longitudinal Micro Data Archive System

Jeong Im Ahn

Data Mgt. Division, National Statistical Office

#1205, Bd3, Gov't Complex Daejeon, Dunsandong, Seogu

302-701 Korea

jiahn@nso.go.kr

It has been around for one, two decade(s) that the NSO disseminate researchers or the public micro data holdings. Now, it seems to be reasonable to build a new LMA(Longitudinal Micro data Archive) that serves a more complicated needs of users. This paper is about to design a design process in LMA model.

. Micro data use

The National Statistical Office have begun to disseminate some of their survey data holdings, so called micro data, as well as released statistics figures. When users ask sets of micro data of certain subject matter over the decades, the independent yearly-data sets are supposed to disseminate together with the codebook or documentation of the original survey respectively. So users shall handle those set of micro data, integrating an inconsistent terms or items over long period of times.

(Unit: users)

	1998	1999	2000
Total	381	368	441
Governmental Use			
Central gov' t	30	39	51
Local gov' t	21	16	19
Non-governmental Use			
Academic & Public institute	135	136	128
Individual	34	30	74
Others	161	147	169

. Need of longitudinal micro data set

The above dissemination system have a few problem, for example, lack of standard metadata directory to convert the different codes used in a certain-year survey, poor inter-connectivity between micro data sets, and a lack of common terminology on statistical data. Now, there is a need to construct the new longitudinal micro data archive files that can supply users with much more useful information than are available from the independent micro data set. Here, supplement of missing data based on the suppositions concerning that period situations will be even better to analyze the micro data.

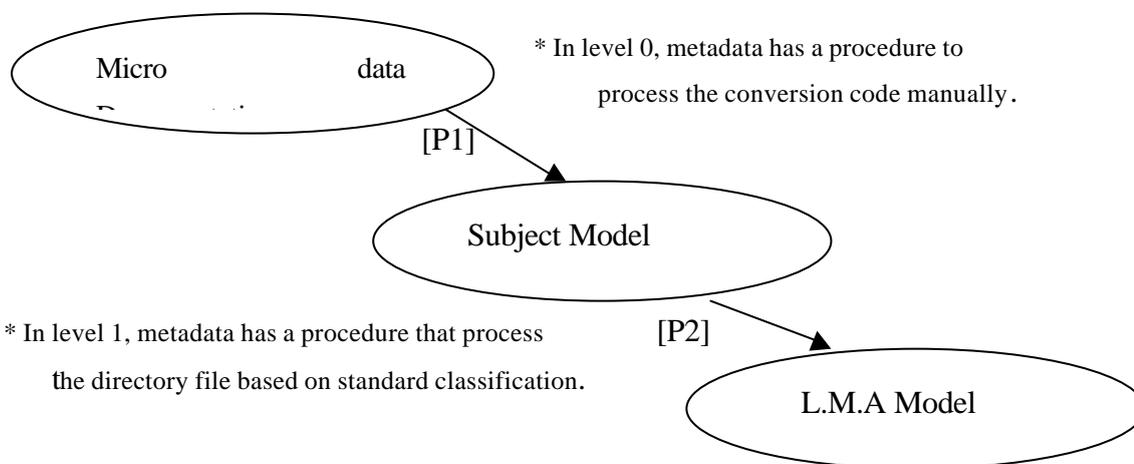
. Strategies to build model on LMA (longitudinal micro data archiving)

3.1 Degree of intensity of metadata¹ integration

Control of the metadata is the most important thing when we connect the different micro data set over decades. However, descriptive information on statistical data would not be enough to show common specifications among several dataset. So let me define the level of degree of metadata integration intensity as below.

- Level 0: Micro data archives independently, such as, yearly volume,
- Level 1: Micro data archives with metadata directory, which has conversion matrices between different classification scheme
- Level 2: Micro data archives with metadata directory which includes extrapolation or filling in the missing data processing function

3.2 Development in longitudinal micro data archiving model



- [P1] Procedure is to make a rule of refining the survey items over decades, to check integrity of data in the same subject, and to create Superset Key among several several micro data sets.
- [P2] Procedure is to integrate and reorganize similar tables on relational scheme, which composes “Superset table” and “Subset table”. Longitudinal data in Subset table by subject will have inserted columns(i.e., temporal data, derived data) for querying or deleted column(i.e., unnecessary data but surveyed)

• Conclusion

The new reconstructed LMA model can solve the problem of getting longitudinal data with metadata conversion manually. Moreover, it will be possible to tell what was not known at the time the original survey was conducted.

¹ Statistical metadata is descriptive information or documentation about statistical data, which facilitate understanding of statistical data over the lifetime of that data.