STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

Working Paper No. 5
English only

CONFERENCE OF EUROPEAN STATISTICIANS

UN/ECE Work Session on Statistical Metadata
(Washington, DC, USA, 28-30 November 2000)

Topic (i): Statistical metadata for dissemination

METADATA AND DATA QUALITY

Submitted by U.S. Bureau of Labor Statistics[1]

Invited paper

I. INTRODUCTION

1. The purpose of this paper is to discuss statistical metadata (hereinafter called metadata) from the survey methodologist's perspective. In particular, metadata as it relates to data quality is the main theme. We develop this by examining the metadata needs for a methodologist from the perspective of the design and evaluation of a specific measurement objective in a specific survey: the collection and classification (or coding) of industry and occupation (I&O) data for persons in the Current Population Survey (CPS) (Abraham and Prewitt, 2000).

2. We have chosen I&O data to illustrate the role of metadata in establishing data quality for several reasons. First, its collection involves the quality issues inherent in the collection of any survey data (e.g., respondents may not understand the question as intended or may forget relevant experiences when answering). Second, coding open-ended answers to questions into formal classification schemes (such as industry and occupation systems) involves a layer of processing where error can be introduced by someone other than the respondent or interviewer. There are particularly vexing problems in evaluating the quality of I&O data. For example, it is hard to establish what is "truth" for a particular open occupation response because its classification into a coding system is usually a matter of judgment and degree. Thus the construct of validity is less useful here than for evaluating other types of data. Instead, methodologists usually rely on reliability or the agreement among coders as a measure of quality.

3. Statements about data quality are examples of metadata (Dippo and Gillman, 1999), so in the I&O example, measures of reliability and validity can be considered metadata; their existence, in fact, adds to the quality of data. Since metadata are data, there is quality associated with them, too. For example, coding reliability may be affected by keying or transcription errors. These errors may result in unintentional agreements or disagreements between coders, leading to lower quality reliability measures. The purpose of metadata is to describe data within some context, thus, how well the metadata describes the underlying data is a quality issue, too.

4. There are two facets of data quality associated with metadata:

---

[1] Prepared by Cathryn S. Dippo, Frederick G. Conrad, and Daniel W. Gillman.

- ACCURACY – The degree to which a metadata element value is close to the truth. This is often reported as a number but is not restricted to numerical measures.
- RELEVANCY – The degree to which a metadata element bears upon or is connected to its intended use.

These facets of metadata quality are analyzed in this paper.

5. In general, coding in surveys is a classification process, i.e., it is the assignment of a survey response to one of a pre-defined set of categories. In the case of I&O coding, there are two classification steps, one for industry coding and the other for occupation. The choice of occupation code in the CPS is usually dependent on the industry code. Coding is based on responses to four open-ended questions plus two multiple-choice questions.

6. From the survey methodologist's perspective, however, there is more to coding than just classifying responses. The construction of the classification schemes, evaluation data sets, and alphabetical indexes of coded phrases have an effect on the coding process. The structure of the schemes, definitions of concepts, completeness and non-overlap of the categories, completeness and relevance of the phrases, accuracy of assigned truth codes, and other factors are all metadata pertinent to the coding process.

7. The interview, that is the context in which respondents provide industry and occupation descriptions, is also crucial to the quality of the metadata. The design of the questionnaire, the design of the specific questions, the interview instructions, and the interviews themselves all have an effect on the coding process and there is metadata associated with each of them.

8. Finally, accuracy of classification decisions by coders, use of the classification schemes, ease of use of clerical computer assisted coding software, accuracy of automated coding software, accuracy and completeness of reference manuals for hard-to-code cases are all metadata about the coded responses.

9. The paper is organized into five sections: the introduction; metadata quality for the coding structures and definitions; metadata quality for the interview (i.e., data collection); metadata quality for coding (classification); and a conclusion, including some recommendations.

II. DEFINITIONS and CODING STRUCTURES

10. I&O coding relies on a set of pre-specified lists, structures, and databases in order to conduct the operation during survey processing. The quality of these resources has a large impact on the quality of the codes that are produced during processing. The major resources for I&O coding in the CPS are developed for the US Decennial Census and current surveys: the I&O Classification (containing industry and occupation classifications); the I&O Alphabetical Indexes (an industry list and an occupation list of phrases with associated codes); the expertly coded evaluation and test sets of responses; and I&O coding software. We will briefly describe

- each resource
- how each resource is used in I&O coding
- how the quality of the resource affects the quality of coding
- what quality measures are needed.

11. The US Census Bureau maintains the I&O Classification (Scopp, Priebe, and Earle, 2000) for the US, and, historically, has constructed a new one just prior to each Decennial

Census. The latest I&O Classification is based in the North American Industrial Classification System (NAICS) (Ambler, 2000) and the Standard Occupational Classification (SOC) (Johnson, 2000). The NAICS is a North American standard, developed jointly by Canada, Mexico, and the United States. It is a hierarchy (taxonomy) with 4 levels, and the US extended it to 6 levels for its use. The SOC is similarly constructed, but it is a US standard only. The I&O Classification is developed from NAICS and the SOC, but due to limitations of the questions used in the Decennial Census and current surveys and the expected small size of some categories, the I&O Classification is a three digit hierarchy for both industry and occupation. Some of the categories are not one-to-one mappings to NAICS or SOC categories, because several NAICS or SOC categories are joined into one I&O category. Many of the occupation categories are associated with particular industry categories, so in the I&O coding operation, an industry code is assigned first.

12. The categories defined in the I&O Classification must follow certain principles. The categories must be mutually exclusive and exhaustive. These categories must be well defined so a coding expert can classify accurately and consistently. Also, the developers of coding software must be able to understand the distinctions between categories in order to produce accurate codes. The rules that associate occupations with certain industries must be complete, consistent, and understandable, because the occupation category so often depends on the industry category. The categories are selected to correspond to the 3 digit level of NAICS and the SOC, so each category must be selected with knowledge of the size of the labor force in advance. Since the coding structure is used for tabulation purposes in reporting I&O data, the size of categories down to small geographic areas is important for avoiding disclosures. It makes little sense to define a category that has few respondents assigned to it.

13. The quality of the I&O Classification system affects the quality of the coding process. The two main quality issues are

- Do the definitions adhere to accepted rules and guidelines for writing definitions, i.e., are the definitions clear? Do they adequately describe the concepts represented in the classifications? (ACCURACY)
- Are the categories or concepts mutually exclusive and exhaustive with respect to the sets of all industries and occupations? (RELEVANCY)

14. The I&O Alphabetical Indexes (Scopp, Priebe, and Earle, 2000) are two alphabetized lists of phrases obtained from responses to the I&O questions from Censuses and demographic surveys. One index contains phrases describing industries with the appropriate industry code from the I&O Classification assigned, and the other index is similarly constructed for occupations. The main purpose of the I&O Alphabetical Indexes is to enable a human coder or software to associate I&O codes with survey responses. The Indexes must be kept up to date with respect to new descriptions of industries and occupations as the characteristics of labor force change over time, and they must have the most current I&O Classification codes assigned to each phrase.

15. The I&O Alphabetical Indexes are used by clerical coders to assign codes. They are crosswalks between the responses and codes. They are used by developers of automated (batch) coding systems for designing rules to incorporate into their systems. The most important quality considerations and their metadata quality types are:

- The extent to which the Alphabetical Indexes are effective crosswalks (RELEVANCY);
- The extent to which codes assigned to each phrase are agreed to by experts (ACCURACY).

16. The I&O coding operation requires the construction and maintenance of expertly coded sets of responses. These files, so called truth sets, are used for training clerical coders, testing computer assisted clerical coding systems, and training and evaluating automated (batch) coding systems (Gillman and Appel, 1994). These truth sets are most effective when they are representative samples of the cases in the population being measured. Each truth set must have the following quality characteristics (with the metadata quality type included):

- The extent to which cases in the truth set represent the population to be measured (RELEVANCY);
- The extent to which there are enough cases within each code category in the truth set for evaluation purposes (ACCURACY);
- The extent to which cases in the extent to which truth set are collected through the same interview mode as the survey or census under consideration (ACCURACY).

17. The software used to help assign codes to cases is also subject to quality measures. There are three major types of software, but only clerical assisted and automated are currently in use. The types are

- Clerical assisted coding software – software that assists clerks in assigning codes to cases;
- Interactive coding software – software that is used to assign codes at the time of the interview;
- Automated (batch) coding (Speizer and Buckley, 1998) software – software that assigns codes to cases without human interaction.

Each type of software has its own characteristics, but for I&O coding, all make use of the resources described above.

18. Evaluation techniques need to be used to measure the accuracy and productivity of automated (batch) coding software (ACCURACY), and some evaluation techniques are more effective than others (RELEVANCY) (Chen, Creecy, and Appel, 1993). Interactive coding will make use of an automated (batch) coder in the field, either CATI or CAPI. So, the same requirements as for automated (batch) coders are in place, but additional requirements such as the ability to ask for more information and effective user interfaces are needed (both of which introduce their own quality issues). Finally, the use of the I&O Alphabetical Indexes are required for computer assisted clerical coders. The indexes serve as the crosswalk between responses and codes for the clerks. Searching the indexes appropriately and quickly helps the clerks assign codes faster, more accurately, and more consistently.

19. No matter the quality of the classification, alphabetical indexes, test files, and software, the I&O responses and codes depend on the quality of the interview and the questions asked. Metadata quality issues related to the interview are described in the next section.

III. THE INTERVIEW

20. Given a well-specified concept, there are many ways to measure it even within a particular survey situation. Alternative question wording and sequencing, page or screen layouts, fonts and formats, navigation aids, and computer hardware can have an effect on how well a concept is measured. In interviewer-administered surveys, interviewer experience and the quality of training materials, trainers, reference manuals and on-line help, maps, and supervisory and other types of feedback all have a role in determining the quality of the statistical product. Each of these and similar data collection design factors influences nonsampling error—be it bias or variance—and therefore influences ACCURACY and RELEVANCY. We will not try to identify which quality facet applies for a particular situation in this section; instead we will discuss

- What types of metadata help those involved in the ongoing production process monitor or evaluate the quality of the statistical output;
- What types of metadata help those who design, validate, and develop improved methods or processes.

21. In the interest of space, we will limit our discussion to just one facet of the interview design process—question wording; moreover, we will limit the discussion to self reporting since proxy reporting adds another dimension to the problem. The purpose of a question is to elicit usable, accurate data from a respondent. Most methodology researchers agree that in answering a question, the respondent completes four mental tasks: comprehension, retrieval or estimation, judgment, and verbalization. Thus, each question must be constructed to facilitate these mental processes across a wide variety of people with different levels of knowledge, intelligence, and experience. While entire books have been written on developing and evaluating alternative question wording (e.g., Sudman and Bradburn, 1982), no one has directly addressed the metadata needs of the survey methodologist performing these tasks.

22. Continuing our example related to the measurement of industry and occupation in the Current Population Survey, let's look at the five question sequence used prior to 1994 (attachment 1) as compared to that in the new computer-administered instrument introduced in 1994 (attachment 2). Answering open-ended questions like "What kind of business or industry is this?" is especially difficult for respondents who, in all likelihood, have no idea of how this information is to be used or that industry and occupation classification systems even exist. Context plays an important role in the comprehension of "What kind of" questions. Respondent's retrieval strategies will depend on what they perceive to be the intent of the questions. Alternative wordings like "do you do" versus "were you doing" or "usual" versus "most important" can lead to different judgments by the respondents about the relevance of the retrieved information to the question being asked. Respondents' abilities to verbalize and communicate the information they perceive to be important are crucial, since the inclusion of words like "managing" or the ordering of words as in "assistant teacher" versus "teacher assistant" can affect major occupational group classification. Self-perception also plays a role in a respondent's choice of words, e.g., bookkeeper versus accountant.

23. Some of the methods for testing alternative questions are: cognitive laboratory interviews using think-aloud and/or probing techniques; standard interviews with respondent debriefing where the debriefing could include direct probing, vignettes, or alternative questions; reinterviews; behavior coding; and record checks. Each method has its strengths and weaknesses, and each method results in metadata about different aspects of the respondents' tasks. For the CPS industry and occupation questions, research efforts were directed at evaluating the adequacy, accuracy, and consistency of responses using measures of agreement/disagreement between resulting codes, rates of referral (coders to expert coders or supervisors), and coder confidence (Gaertner, Cantor, Gay, Shank 1989). Multiple coders independently coded the same response, responses to a reinterview were compared to the original, and respondents' descriptions were compared to employers'. The basic assumption is that vague or ambiguous questions lead to incomplete, ambiguous, or variable responses that do not result in consistent codes over coders, respondents, or time.

24. Probing questions asked as part of a laboratory debriefing interview can be used to investigate potential problems seen by the survey designer, e.g., are respondents thinking about their specific work site or the company in general when they answer the question "What kind of business or industry is this?" But, not all problems with question wording are readily foreseeable, and considerable research has gone into developing methodologies which consider "verbal

reports as data.", (Ericson and Simon, 1993). Many of these methods are more qualitative than quantitative, leading to metadata which is not easily summarized or structured. Moreover, a person's facial expressions, body language and disfluent utterances can contribute to an observer's understanding of what has been said. In other words, video- and audio-taped interviews, with or without transcriptions or behavior coding, can provide crucial metadata to the survey designer trying to understand how well a particular question works.

25. In the end, research results indicated that respondents possess little information for placing the open-ended questions in the context of our classification systems. Thus, while the question wording and ordering changes implemented in 1994 marginally improved codability and coding consistency, only major questionnaire changes which address the basic concepts of the classification system have the potential for making significant improvements—an avenue of research that was impractical at the time since the basis of the new industry and occupational classification systems were only in their formative stages.

26. On the other hand, the questionnaire research reports and other documentation should, if properly preserved, provide valuable metadata to the next set of investigators attempting to improve the instrument for measuring industry and occupation. But, is it as good as it could be? No. It is clear from reading over the available documentation from 1989-91 on this topic that much has been lost in terms of relevant metadata. Rationale for the questionnaire design, in particular, is poorly captured, and the inability to see instead of just reading about a problem makes it less salient. Both of these problems could be addressed using information technology research today, where they might not have been a decade ago. Thus, it is important to think about how metadata for the survey designer could be improved through advances in information technology.

IV. CODING

27. In the previous section, we mentioned the four stage thought process that (cooperative) respondents inevitably go through in order to answer a survey question, and we noted that nonsampling error can be contributed at any of these stages. In addition, nonsampling error can be introduced to the measurement process when respondents provide open-ended verbal reports that must than be classified by coders[2]. Assigning an occupation description to an occupational category, to pick one type of open response, can involve complex judgments, requiring coders to integrate information from various sources. It is relatively easy for things to go wrong, so some measure of the quality of the coding process should be included with the data.

28. It often makes sense to think of category membership in terms of degree rather than "all or none." A lamp, for example, could be a piece of furniture though not as typical as a chair, and it could also be an appliance or a lighting fixture or part of a home security system. This is also true of occupation descriptions. For example, someone who "fixes broken equipment" could be a repairer or a building engineer. The problem is that coding requires an all or none judgment – the description is assigned to only one category and it is considered to be as good an exemplar as any other description assigned that code. One consequence is that coders may develop somewhat arbitrary rules to definitively assign ambiguous cases to a single category. These rules will lead to codes that are reliable (ACCURACY), but not necessarily valid (RELEVANCY). Different coders can assign the same code to the same description, but the code may differ from the "true code", operationalized as expert judgment, employers' description, etc.

[2] In this section, we are primarily concerned with clerical – as opposed to batch automated or interactive – coding, as described in section 2.

29. Couper and Conrad (2000) observed a rule used by occupation coders that could produce metadata that is high quality in the ACCURACY sense but low in the RELEVANCY sense. The rule is for classifying descriptions that mention two job titles such as "nurses assistant and cook". These descriptions include not only job titles but duties, in this case "cooking and special diets personal care." In that study, coders provided verbal reports as they made their decisions. In their reports for these "two-job title cases," several coders referred to a rule directing them to choose an occupation code corresponding to the first duty listed; in this case, that leads to a code for cook instead of nurse. Couper and Conrad propose that, because such rules are arbitrary, they may lead to codes that could be wrong on the basis of the content of the description, though if they are followed consistently will lead to high agreement between coders.

30. Similarly, Hak and Berndt (1996) observed that a group of research coders developed rules based on superficial features of the descriptions for cases that could not be classified on the basis of explicit, theoretically based rules. By scrutinizing the conversations among these coders (who were working in a sociological rather than occupational domain), Hak and Berndt documented the creation of such rules. For example one coder said to the others "if the word 'being able to' occurs in the answer, we can almost always choose [the code] VALUE …" The authors argued that coding on the basis of such non-theoretical characteristics potentially undermines coding accuracy.

31. If coders systematically assign a code on the basis of a rule, this may improve intercoder reliability overall but, if they do not apply the rule universally, this can add correlated coder variance to the overall variance of the estimates for that code. This is generally interpreted as decreased coding accuracy (RELEVANCY). Martin, Bushnell, Campanelli and Thomas (1995) found more correlated coder variance among a small set of traditional occupation coders working in an office setting than among a larger set of interviewers who coded the descriptions they collected after each interview. Apparently, it is harder for a large, dispersed group to develop and maintain coding conventions than it is for one which is small and concentrated.

32. The quality of coding can also be compromised by certain characteristics of the descriptions. On intuitive grounds, longer descriptions could clarify what is ambiguous or vague in shorter descriptions (RELEVANCY). One measure would be greater intercoder agreement for long than short descriptions. However, Couper and Conrad (1996) found the opposite trend. They asked all respondents the basic occupation questions from the Current Population Survey (CPS) and also asked half of the respondents an additional job title question. Intercoder agreement for the basic CPS questions was 86.4% but was 82.1% when job title was also collected.

33. Couper and Conrad (2000) found additional evidence for this trend in a subsequent study of CPS occupation coding. For quality control purposes, two independent coders routinely classify a subset of the descriptions. In a sample of these "double-coded" cases, descriptions were shorter (8.94 words on average) when coders agreed than when they disagreed (9.76 words on average). In a follow-up experiment, where they could control the characteristics of the descriptions, the authors showed that longer descriptions only led to lower agreement when the descriptions included words that had previously been associated with high agreement (e.g., secretary, cashier, driver); when the descriptions included words that had previously given coders trouble (e.g., operator, director, clerk), the length of the descriptions did not affect coding agreement and there was a hint that longer descriptions increased agreement.

34. Coders seem to realize that more information is not necessarily better. Cantor and Esposito (1992) asked a set of CPS coders to listen to tape recorded interviews and to indicate at what points they would have liked the interviewer to probe for additional information. Presumably coders know where additional information could clarify the descriptions. As it turned out, the coders virtually never indicated that a probe would have been helpful. In the context of the findings about description length, this suggests that interviewers recognize that it is easier to code shorter descriptions as their colleagues would code them (ACCURACY) than is the case for longer descriptions.

V. CONCLUSION

35. Data quality metadata (i.e., information about the quality of a particular statistic) are a survey methodologist's life-blood. Design, development, and evaluation of survey processes are dependent on defining, collecting, and analyzing metadata. Understanding and tracking the quality of those processes produces essential metadata, and the quality of that metadata must be assessed just as the data are, using concepts such as accuracy and relevancy. Thus, the quality of that metadata is crucial, and it deserves as much attention as the quality of the data itself. On the other hand, one needs to avoid being too focused on summary statistical measures as the only type of data quality metadata. The outliers (e.g., the double job title responses, the questions eliciting disfluent utterances) help the methodologists decide when process improvements might be useful or have worked.

VI. REFERENCES

Abraham, K., and Prewitt, K. (2000). "Current Population Survey – Design and Methodology", Bureau of Labor Statistics and Census Bureau, Technical Paper #63.

Ambler, C. (2000). "Introducing the North American Industry Classification System", Presented at Government Statistics Section of American Statistical Association, Indianapolis, IN, August 2000.

Cantor, D. & Esposito, J. (1992). "Evaluating interviewer style for collecting industry and occupation information", *Proceedings of the Section on Survey Research Methods*, 661-666.

Chen, B., Creecy, R. H., and Appel, M. (1993). "On Error Control of Automated Industry and Occupation Coding", *Journal of Official Statistics*, Vol. 9, No. 4, 1993, pp. 729-745.

Couper, M. P. and Conrad, F. G. (October, 2000). "Classifying open-ended reports: Coding occupation in the Current Population Survey", Paper presented at 5th International Conference on Social Science Methodology, Cologne, Germany.

Couper, M. P. and Conrad, F. G. (July, 1996), "Collecting Data to Facilitate the Classification of Occupations Using a Skill-Based Approach", Paper presented at the Fourth International Social Science Methodology Conference, Essex, UK.

Dippo, C. and Gillman, D. (1999). "The Role of Metadata in Statistics", Invited Working Paper #2, UN/ECE Work Session on Statistical Metadata, Geneva, Switzerland, September, 1999.

Ericsson, A, & Simon, H. (1993). *Protocol Analysis: Verbal Reports as Data* (2nd edition). Cambridge, MA: MIT Press.

Gillman, D. W. and Appel, M. V., (1994). "Automated Coding Research at the Census Bureau", Statistical Research Report Series, No. RR94/04, Census Bureau, 1994.

Hak, T. and Bernts, T. (1996). "Coder training: Theoretical training or practical socialization?", *Qualitative Sociology*, 19, 479 -501.

Johnson, P. (2000). "Revising the Standard Occupational Classification", Presented at Government Statistics Section of American Statistical Association, Indianapolis, IN, August 2000.

Martin, J., Bushnell, D., Campanelli, P. and Thomas, R. (1995). "A comparison of interviewer and office coding of occupations", American Statistical Association: Proceedings of the Section on Survey Research Methods, 1122-1127.

Scopp, T., Priebe, J., and Earle, K. (2000). "Adaptation of the Standard Industry and Occupation Classification Systems For the 2000 Census", Presented at Government Statistics Section of American Statistical Association, Indianapolis, IN, August 2000.

Speizer, H. and Buckley, P. (1998). "Automated Coding of Survey Data", in Couper, M. *et al* (eds.) *Computer Assisted Survey Information Collection*, Wiley Series in Probability and Statistics, 1998, pp. 223-243.

Sudman, S. and Bradburn, N. M. (1982). *Asking Questions: A Practical Guide to Questionnaire Design.* San Francisco, CA: Jossey-Bass.

Attachment 1

Pre-1974 Questions Related to Industry and Occupation

23A. For whom did … work? (*Name of company, business, organization or other employer.)*

23B. What kind of business or industry is this? (*For example: TV and radio mfg., retail shoe store, State Labor Dept., farm.)*

23C. What kind of work was … doing? (*For example: electrical engineer, stock clerk, typist, farmer.)*

23D. What were …'s most important activities or duties at this job? *(For example: types, keeps account books, files, sells cars, operates printing press, finishes concrete.)*

23E. Was this person
An employee of a PRIVATE co., bus. Or individual for wages, salary or comm…
A FEDERAL government employee
A STATE government employee
A LOCAL government employee
Self-empl. In OWN bus., prof. Practice, or farm
Is the business incorporated? Yes
No
Working WITHOUT PAY in fam. Bus. Or farm
NEVER WORKED

Attachment 2

1994 Industry and Occupation-Related Questions

>IO1INT< (Now I have a few questions about the) (job/main job)
(at which (you/he/she) last worked/at which (you/he/she) worked (LAST WEEK/THE WEEK BEFORE LAST)/from which (you/he/she) (were/was) absent (LAST WEEK/THE WEEK BEFORE LAST)/from which (you/he/she) (are/is) on layoff.)
(Were/Was)/(LAST WEEK/THE WEEK BEFORE LAST), (were/was) (name/you) employed by government,
by a private company, a nonprofit organization, or
(was/were) (you/he/she) self employed(or working in the family business?)
<1> Government
<2> Private for profit company
<3> Nonprofit organization including tax exempt
and charitable organizations
<4> Self employed
(<5> Working in the family business)
===>_

>IO1GVT< Would that be the federal, state, or local government?
<1> Federal
<2> State
<3> Local (county, city, township)
===>_

>IO1INC< (Is/Was) this business incorporated?
<1> Yes
<2> No
===>_

>IO1WP< (Was/Were) (name/you) working for pay?
<1> Yes
<2> No
===>_

>IO1NMP< What is the name of the (company/nonprofit organization) for which
(you/he/she) (work/works/worked)(at (your/his/her) MAIN job?)
DO NOT READ TO RESPONDENT:
(name of company, business, organization or other employer)
===>_____

>IO1NMG< What is the name of the government agency for which (you/he/she)
(work/works/worked)(at (your/his/her) MAIN job?)
===>_____

>IO1NMB< (What is the name of (your/name's) business?/What is the name of the business for which) ((you/name) (work?/works?/worked?))
===>_____

>IO1IND< What kind of business or industry is this?
READ IF NECESSARY: What do they make or do where
(you/he/she) (work?/works?/worked?)
(**THIS CASE WAS REFERRED LAST MONTH.**)
(PLEASE PROVIDE MORE DETAILED INFORMATION FOR THE CODERS.)
===>_____

>IO1MFG< ASK IF NECESSARY:
Is this business or organization mainly manufacturing,
retail trade, wholesale trade, or something else?
<1> Manufacturing
<2> Retail trade
<3> Wholesale trade
<4> Something else
===>_

>IO1OCC< What kind of work (do/does/did) (name/you) do,
that is, what (is/was) (your/his/her) occupation?
(For example: plumber, typist, farmer)
(**THIS CASE WAS REFERRED LAST MONTH.**)
(PLEASE PROVIDE MORE DETAILED INFORMATION FOR THE CODERS.)
===>_____

>IO1DT< What (are/were) (your/his/her) usual activities or duties
at this job?
(For example: typing, keeping account books, filing,
selling cars, operating printing press, laying brick.)
(**THIS CASE WAS REFERRED LAST MONTH.**)
(PLEASE PROVIDE MORE DETAILED INFORMATION FOR THE CODERS.)
===>_____
===>_____