Review of Income and Wealth
Series 50, Number 2, June 2004

DESIGNING ANALYTICAL DATA FRAMEWORKS

BY JAN W. VAN TONGEREN*

Development Research Institute (IVO) of Tilburg University, Netherlands

The article is about design of analytical data frameworks that make optimal use of micro data available to serve the needs of the policy analyses and create a maximum number of data checks at the macro or meso level. As the SNA is one of the most extensive analytical data frameworks, the notions developed in this paper will be applied mainly to data segments of the SNA and in particular to a much used data segment of the SNA, i.e. the Supply and Use Table (SUT). The larger part of the paper focuses on defining a simplified SUT and using a so-called "system of classifications and correspondences," that is based on international SNA and classification standards but taking into account specific features of the economy of a country, data availability and the type of analysis served by the framework. The simplified SNA should be comprehensive, but should avoid unnecessary cross-classifications of data and details.

INTRODUCTION

Designing data frameworks, even within the limited context of the SNA, is a much-neglected topic in which very few authors have expressed interest. Ruggles and Ruggles were among the few who did see the importance of this topic in relation to the SNA (Richard and Nancy D. Ruggles, 1975; Nancy D. and Richard Ruggles, 1983, 1986, 1992).

The objective of the present paper is to develop guidelines for the design of analytical data frameworks, which have the following features:

- Available data obtained from micro units are incorporated in the framework as aggregate data sets for groupings of micro units, referred to here as meso groupings or sectors. The meso groupings or sectors selected are those that are the subject of government and other policies.
- When aggregating the micro data into data sets for meso groupings,
  - a maximum amount of analytical information should be obtained from the available micro data;
  - a maximum amount of data checks available in the micro data should be incorporated, so that, after adjustment of data in line with the consistency requirements of data checks, optimal data consistency and reliability is achieved.

When designing an analytical framework, a delicate balance should be kept between comprehensiveness in terms of analytical information and data checks

Note: The author wishes to thank his former colleagues at the UN Statistics Division (Ralph Becker, Ivo Havinga and Vu Viet), and his present colleagues at Tilburg University (Arthur Giesberts, Judith Lammers, Jan Magnus, Ruud Picavet) for the technical support and constructive comments received on earlier drafts of this paper.

*Correspondence to: Jan van Tongeren, Tilburg University, Warandelaan 2, P.O. Box 90153, 5000 LE Tilburg, The Netherlands (JanWvT@compuserve.com).

that are included, and simplification and transparency of the framework in which unnecessary details are avoided. The latter is emphasized in Section 2.3, when referring to a simplified SNA.

In the remainder of this paper, references will be mainly to the 1993 SNA, as it is the most comprehensive representative of framework development. If its most extensive core of Supply and Use Tables (SUT) and Integrated Economic Accounts (IEA) were implemented, it would be able to enforce data consistency and analysis based thereon for almost all economic statistics; and if extended to satellite accounting, it is possible to extend such data consistency to social and environmental statistics, which can be linked to economic data. Frameworks also exist in specialized fields of socio-economic statistics, including IMF frameworks for Government Finance Statistics (IMF, 2001), Balance of Payments (IMF, 1993) and Monetary and Financial Statistics (IMF, 2000), Labor Accounts (Hoffmann, 2000) and Informal Sector statistics (ILO, 2001), developed by the ILO or frameworks for Agricultural Accounts and Food Balances developed by FAO (FAO, 1996). For population projections there are frameworks defining the relation between population variables, such as birth, death, and migration and between the changes in those variables (United Nations, 1983a).

Tradition in national accounting, and possibly in other frameworks, has it that tables in which data are presented are determined by international and regional SNA standards. While this practice contributes to international comparability, it makes less than optimal use of the available data when incorporating checks in the data compilation and when revealing the analytical features of an economy. In the present paper it is suggested that the international standards of data presentation are supplemented by analytical frameworks that are specifically designed for local use, in support of effective local analysis and compilation of data.[1] When describing the design of the SNA framework for local use, the paper uses a formalized description of the SNA that differs in some respects from that of the 1993 SNA Manual, in which reference is made to terms such as data vectors, matrices, identities and indicator ratios, systems of classifications, attributes, etc., which are not part of the terminology used in the SNA Manual.[2] They are, however, essential in the formalized description of the System pursued here.

[1]The topic of this paper is part of a continuing research project at Tilburg University carried out by the author and others to formalize a number of aspects of SNA and other analytical frameworks that are determined mainly by tradition until now. The research focuses on three aspects, i.e. (i) the design of frameworks, (ii) the approach to compiling the framework data (see van Tongeren, 1986; Magnus, van Tongeren, and de Vos, 2000), and (iii) the relation between the assumptions used in compiling framework estimates and analytical indicators in which there is much international interest at present. The intended formalization of national accounts and other frameworks is needed in order to make more transparent the process of designing those frameworks and developing framework estimates. It furthermore is an essential requirement of the increasing computerization of the national accounts compilation process. The present paper, while focusing mainly on the design of frameworks, will sometimes make reference to compilation and indicator ratio issues, but will not deal with these issues in detail.

[2]The term "attribute" is used in databases to indicate the various dimensions of information items included in the database. The term is used in the present paper in relation to classifications and cross-classifications, and implies that the suggested approach to classifications is close to a database approach and/or may be adapted to it.

1. Analytical Data Frameworks

The term analytical data framework is being used frequently by many, but in a loose manner. As the term is used throughout this paper, there is a need to define it more precisely.

An analytical framework is a table and/or a group of tables, which include data, and with relations defined between the data of those tables or groups of tables. The relations between the data included in a framework may be used in analysis (this explains the adjective "analytical"), and some of the relations may also be used in the compilation when checking the consistency of data in the tables. The tables may be one-dimensional vectors, or two- or multi-dimensional matrices. The relations between the elements of the tables may be identities or indicator ratios. The only relation that does not qualify a table to be an analytical framework is the identity between the row and sum-totals, which always holds in every table. The relations of identities and indicator ratios defined between the data of the framework are not only used in the analysis of the data, but also define the consistency of the data included in the framework.

An analytical data framework differs from a database. The latter, supported by several computer software programs, only organizes the data, using various attributes that are defined for each data element. A database, however, does not necessarily include relations between the data, as is the case in analytical data frameworks.

The relation between meso aggregation, data checks, and analytical information is reflected in the illustrative scheme of Table 1, which is based on the 1993 SNA. The data of meso groupings or sectors shown in the columns refer to the resident sectors that are distinguished in the SNA and that interact with each other in the economy of a country; they are the Non-Financial Corporations (NFC), Financial Corporations (FC), Government (GOV), Households (HHs) and HH subsectors (such as urban and rural HHs, or HHs with different composition and income or other characteristics) and Non-Profit Institutions (NPI).[3] The information items in the columns of the resident sectors are aggregated into those of the National Economy. The column for the Rest of the World (ROW) includes counterpart items such as exports and imports that constitute economic and other cross-frontier impacts of actions of resident sectors outside the economy and impacts from outside the economy on resident sectors. The sector columns are augmented with a column for the grouping of natural resource units (forests, fish stocks, mineral reserves, quality of air, etc.) that is used in economic-environmental satellite accounts analysis and brings together the information items on impacts of the other sectors on the natural environment.

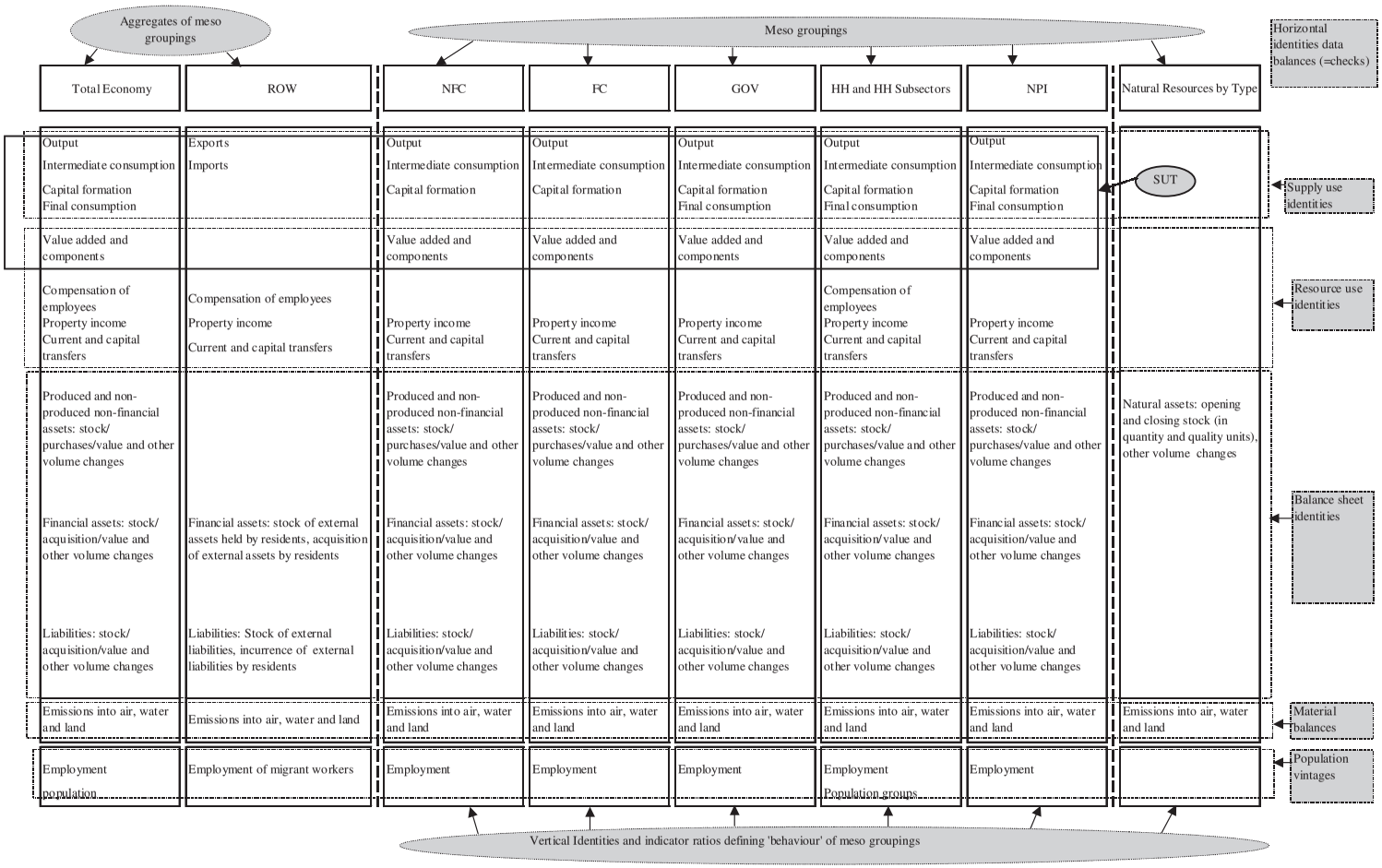
In the column of each sector are included data boxes for economic, social and environmental information items that describe the actions and impacts thereof for each sector.[4] The upper panels of economic data items refer to those of the

[3]It was thought to be more appropriate to use in the generic scheme of Table 1 the more general term of NPIs (Non-Profit Institutions) rather than the more specific SNA concept of NPISH (Non-Profit Institutions Serving Households).

[4]The 1993 SNA uses for the majority of the information items the terms "transactions, other flows, and stocks" (see 1993 SNA sections, B, C and D of Chapter III). As these terms, however, are too narrow for information items that relate to economic-environmental and socio-economic accounting, it has been preferred here to use the more general term "information item."

TABLE 1

ANALYTICAL FRAMEWORK OF COMPREHENSIVE SNA, INCLUDING ECONOMIC-ENVIRONMENTAL AND SOCIO-ECONOMIC SATELLITE ACCOUNTS



economic core of the SNA, while environmental and social data refer to information items in environmental-economic (United Nations, 2003) and socio-economic satellite accounts (National Statistical Office, Republic of Korea, 2000) and Social Accounting Matrices (see Thorbecke (2000) for the latest overview of the work on SAMs). The information items included in each meso sector column are obtained by aggregating the corresponding information of micro units. Illustrative examples of information items are included in the presentation of the scheme for each sector.

Requirements of analysis and policy use determine the selection and detail of sectors and the selection of the information items to be included for each. Indicator ratios between values of the information items in each column, such as input-output coefficients, labor productivity measures, tax ratios, etc., represent the instruments of analysis embedded in the scheme; and the values of some of these indicator ratios may also be used as data checks, when verifying the mutual consistency of the values of the information items. Other data checks are horizontal identities that may be defined between information items of different sectors. On the right-hand side of the table are included five types of horizontal balances that should hold in a comprehensive SNA, including (i) balances of supply and use of products of the SUT, (ii) resource–use identities in the IEA, that apply to receipts and disbursements of property income, current an capital transfers, etc. between sectors, and (iii) balance sheet identities between stocks of assets of sectors and financial claims on those assets by other sectors. (iv) The material balances of environmental accounting show how emissions "caused" by sectors have an impact on (or are "borne by") the natural resources in the last column, and (v) the identities of population vintages show how population groups move in the course of time from schools to employment, unemployment and retirement, etc. When more detail is included on the information items of each sector, more horizontal data checks and also more indicator ratios become available.

2. Adapting and Further Detailing the SNA-SUT with the Help of Classifications and Correspondences

An aggregate presentation of the economy and its social and environmental ramifications is useful for quantification and analysis at the macro level of concepts, such as production, investment and consumption, or impacts on the environment and population and employment groups. It is, however, too macro-oriented to identify and measure data structures describing behavior and other characteristics of groups of agents that are the target of economic, social and environmental policies, such as producers, households and groups of individuals including those employed, social institutions (schools, hospitals), natural resources, etc. Maintaining the overall macro structure, while incorporating more detail on meso sectors and data for each sector, is what is often called maintaining the requirement of micro-macro links.

The detail of meso sectors is implemented through the use of classifications. As many data sets are joined in a framework, many classifications are needed and several data sets can be classified according to different criteria, using binary or even multiple cross-classifications of the same data. The cross-classifications are an important instrument of analysis, as they establish relations between the data

details from different points of view, which is what is most useful when doing studies in support of policy analysis.

How classifications and correspondences will be used in the further detailing of an analytical framework will be illustrated in the remaining part of this article with help of a subset of the comprehensive SNA in Table 1, i.e. the SUT (Supply and Use Table). It is an SNA segment identified in the upper part of Table 1 and the variables included are located within the solid lines marked there. The SUT segment was selected because it combines several features that are relevant to analytical frameworks: i.e. many analytical relations are defined, and most SNA classifications interact in this part of the System. Furthermore, the application of framework features to the SUT is relevant as there is extensive use of this type of information in I-O types of analysis, and the SUT is also the starting point for several types of satellite accounts, including socio-economic, environmental-economic, tourism, export satellites, etc.

2.1. *International Classifications and Correspondences*[5]

Adaptation of classifications and cross-classifications should start from what is internationally available. The extent to which international classifications and correspondences may be used in specific country settings, however, depends on the specific features of an economy, the analytical objectives of the accounts, and also on data availability. In order to use the international classifications and correspondences in the most optimal manner for detailing analytical frameworks in general and the SUT in particular, the user is provided below with an overview of what classification detail is internationally available at the most detailed level and what type of correspondences are defined between them.[6] This information is generally not available from the internal classification references referred to above.

- The most detailed classification is the HS with 5224 categories at the most detailed level, and restricted to goods only.[7] The CPC (version 1.0) has fewer categories at the detailed level, covering both goods and services (CPC version 1.0: 2101 and CPC 1.1: 2098), and the COICOP has only 157 categories, but is restricted to HH final consumption of goods and services only. The number of categories of the international ISIC is in between that

[5]The paper makes reference to several classifications: i.e. ISIC (International Standard Industrial Classification), CPC (Central Product Classification), HS (Harmonized System), SITC (Standard International Trade Classification), EBOPS (Extended Balance of Payments of Services classification), BEC (classification by Broad Economic Categories), COICOP (Classification of Individual Consumption by Purpose), COFOG (Classification of Functions of Government), COPNI (Classification of the Purposes of Non-Profit Institutions Serving Households), and COPP (Classification of the Outlays of Producers According to Purpose). All classifications and correspondences can be accessed through the UNSD WEB site (http://unstats.un.org/unsd/class/default.htm). The Extended Balance of Payments of Services (EBOPS) classification and its correspondence with the CPC is published in the Manual on Statistics of International Trade in Services, published by UN, EU, IMF, OECD, UNCTAD, WTO, 2002, UN doc nr. ST/ESA/STAT/SER.M/86.

[6]The symbols for correspondences used in the text are 1:1 referring to ONE-ONE correspondences, 1:M for ONE-MANY correspondences, M:1 for MANY-ONE correspondences, and N:M for MANY-MANY correspondences.

[7]This paper will focus on the use of the HS in classifying foreign trade data, even though some countries still might use the SITC. The HS is more detailed than the SITC; only M:1 relations characterize the correspondence between the two.

of the CPC and COICOP, i.e. 292 for ISIC version 3 and 298 for ISIC version 3.1. The EBOPS breakdown of exports and imports of services includes 51 detailed categories.

- The HS-CPC correspondence is mainly an M : 1 correspondence, with many categories of the HS corresponding to fewer categories of goods in the CPC (5224–1151). The remaining categories of the CPC are services that are not linked to the HS, but to the EBOPS. The EBOPS-CPC correspondence of services (51–543) has mainly 1 : M relations, but there are several instances (education, health, insurance, sea transport services), where there is an M : 1 EBOPS-CPC correspondence. It should also be noted that the number of CPC service categories corresponding to the EBOPS classification are fewer than the number of services categories in the CPC (950), as not all CPC service categories are considered relevant for exports or imports.
- The CPC-ISIC correspondence, which is based on CPC version 1.0 and ISIC version 3, is also mainly M : 1 (2037–290), as there are much fewer ISIC than CPC categories at the most detailed level of each classification.[8]
- The CPC-COICOP correspondence includes three types of categories: i.e. (i) categories of COICOP that are linked to the CPC mainly in a 1 : M relation (116–819); (ii) goods and services not destined for final consumption (1281 CPC categories) that are only included in the CPC and not in the COICOP; and (iii) COICOP categories (41) of individual (as distinct from collective) consumption of government and NPIs that are not linked to the CPC.
- The BEC correspondence table with the HS not only makes a distinction within the HS between intermediate consumption, HH final consumption, and gross fixed capital formation, but also includes further detail within each of these three SNA categories, which permits a more precise correspondence with the HS and CPC. At the most detailed level it includes 19 categories, which refer among others to "food and beverages, distinguishing between primary, processed (products) and (products) for industrial and household use," "consumer goods n.e.c., durable, semi-durable and non-durable," "industrial supplies, primary and processed," "transport equipment, passenger motor cars, other, parts and accessories," etc. The HS has mostly M : 1 correspondences with the three BEC categories: of the HS categories 2850 correspond to intermediate consumption, 1246 to HH final consumption, and 954 to capital formation. A fewer number of HS categories, however, are linked to more than one category of the BEC (N : M relation), including 16 mixed categories of intermediate and HH final consumption, such as fuels, 136 mixed categories of HH final consumption and gross fixed capital formation, such as cars and other transport equipment, and 19 mixed categories of all three expenditure categories, covering non-specified goods (n.e.c.). At present there is no correspondence defined between the EBOPS classification and the BEC.

[8]Of the 292 categories of ISIC 3, only 290 have CPC 1.0 correspondences and these account for 2037 out of the 2098 categories of the ISIC.

2.2. *Adapting International Classifications and Correspondences to a National "System of Classifications"*

When further detailing the basic scheme, as presented in Table 1, one might incorporate all international classifications introduced in the previous section, as well as all cross-tabulations between those classifications. In practice, it is not possible to include all cross-classifications, because of data restrictions and lack of analytical interest. One reason, which relates both to compilation and analysis, is that multiple dimension matrices are difficult to handle because of their size. For instance, if output were cross-classified by ISIC, CPC and BEC, using, say, 20 categories of ISIC, 50 categories of CPC, and the three categories of the BEC, this would result in a multidimensional matrix with 3000 ($50 \times 20 \times 3$) elements. Similarly, the intermediate consumption matrix of the SUT would have 1000 categories, if only ISIC and CPC are used. The large number of information items resulting from this would be difficult to handle in analysis, and also surveys and other data sources may not support the large data detail required by such comprehensive output and input matrices. Furthermore, many elements may be either zero or very small and not significant enough in value for use in analysis.[9]

Simplification of the SUT matrices of cross-classified data should therefore be a principal objective in compilation and analysis. Simplification, however, should not be the only objective. It should go hand in hand with developing the SUT framework into an optimal instrument in the compilation and analysis of data. There are several ways of accomplishing this in specific country settings, and classifications and their correspondences play a major role in it. In the sections below are discussed criteria for the coordination of classifications; they may be taken into account when adapting international classifications and correspondences in a coordinated manner and arrive at a "system of classifications" for use in countries. In Section 2.3 the criteria are brought together and reflected in what will be called a simplified SUT framework.

The criteria are aimed at coordinating groups of classifications: ISIC-Sector (Section 2.2.1), CPC-HS-COICOP (2.2.2), ISIC-CPC (2.2.3), BEC-HS-CPC (2.2.4). If, on the basis of one of the criteria, one or more classifications are adapted to local circumstances, adaptations of other classifications may be needed as well, given the interlocking nature of the various classification criteria.

2.2.1. ISIC-Sector Correspondences

Two features of the organization of production in the country are brought out through the classification of data at the meso level of the SUT. The first one is the institutional integration of production units (establishments) and the second is the technical integration of production processes. The institutional classification—reflected in NFC, FC, GOV, etc. distinctions of sectors—was already

[9]These types of matrices are referred in the literature as "sparse" matrices (see Young, 1971). It is argued in that literature that calculations based on inverses and other manipulations of large sparse matrices lead to inaccurate results, due to the incorporations of the large number of zeros and other small values, and that use of approximations of the matrices, which avoid the large number of zero elements, may actually lead to improved results. National accountants may take these considerations into account when designing and compiling the SUT.

embedded in Table 1. The technical integration of production processes is represented at the meso level through the traditional use of the ISIC or industry breakdown of GDP and other production related variables. The correspondence between the two classifications of production related variables shows how the institutional organization of production relates to its technical organization. This correspondence was incorporated in the 1993 SNA in a separate table, the Cross Classification by Industries and Institutional Sectors of Production Account Items (CCIS), and thus given for the first time an analytical meaning and quantification in national accounting. Thus can be quantified the percent of production in manufacturing that takes place in small (HH) establishments, large corporations (NFC) and in the government (GOV) sector. Also within production by NFCs can be distinguished the part that takes place in public, foreign controlled and other private NFCs.

The objective of coordinating the ISIC and sector classifications effectively in an analytical SUT framework, in which production is analyzed in a particular country setting, is to aggregate or disaggregate sectors and ISIC categories in such a manner that the link between the institutional and technical organization of production at the micro level of the establishment is well reflected at the meso level of the SUT. Thus, the detail of sector and ISIC classifications in the CCIS at the meso level should bring out that selected establishments are predominantly organized institutionally, for instance as small (HH) or large (NFC or FC), form part of the government sector (GOV) or a KEY sector of the economy. If the characteristics of these establishments are not well represented in the ISIC and/or sector breakdown, the particular organization of production will not be brought out at the meso level of the data.

A simple and transparent CCIS matrix framework should include ideally only 1 : 1 and M : 1 correspondences between ISIC and sector categories, i.e. each ISIC category corresponds to only one sector category. In this case, the sector classification of data can be derived from the ISIC detail of the SUT framework by simple aggregation. This supports the integration of the SUT data with the other segment of the SNA, i.e. the IEA. From an analytic point of view, it has the further advantage that clear links can be established between the technical analysis of the SUT and more institutional analyses of the economy that do not only include analysis of production, but also fiscal, monetary and financial analyses.

In order to arrive at the ideal case as closely as possible, both the ISIC and sector breakdowns would need to be adapted to local circumstances. Thus, if for instance there were two types of agriculture in the country, i.e. plantation and subsistence agriculture, this should be brought out in a national adaptation of the ISIC breakdown, so that the first one would be allocated to the NFC sector and the second one to the HH sector. Similarly, the introduction of a KEY sector in the sector breakdowns would make it possible in an oil producing county to distinguish between the key sector of oil production and production units producing other minerals, which are NFCs, but do not belong to the key sector. Different sectors also could be distinguished on the basis of compilation criteria. Thus, a distinction may be made between industries on which direct survey information is available and that may be allocated to what may be called a "registered" NFC sector, and industries for which estimates are made on the basis of assumptions

about employment-output ratios and that may be allocated to a "non-registered" NFC sector.

2.2.2. Use of the CPC as the Core Product Classification

The CPC is the main product classification of the SUT. To establish it as the core classification requires that it be possible either to convert the other product classifications to the CPC through an M : 1 correspondence or to convert the CPC into a more summary classification through an 1 : M correspondence between that classification and the CPC. The first type is represented by the M : 1 correspondence between the HS classification of exports and imports and the CPC used to classify output, and the second type by the 1 : M correspondence between the COICOP classification of HH final consumption data obtained from HH surveys and the CPC classification of output.

If countries use the detail of the international HS and CPC classifications, the above condition should not present any difficulty in the actual practice of countries, as the present HS-CPC correspondence only includes M : 1 relations. There would also not be any difficulty, if the actual CPC detail of output at the country level were less than the number of categories included in the international version of CPC; in that case, international HS-CPC correspondences may simply be further aggregated. A problem may be encountered, however, if a country has less HS data detail than the 5219 HS categories that have correspondences with CPC 1.0.[10] As this may result in 1 : M and N : M HS-CPC correspondences, the HS detail of exports and imports and/or CPC detail of output and the correspondences between them may need to be adjusted.

This may be done after comparing the actual values of HS detail of imports and exports with the corresponding values of CPC detail of output. Based on this comparison, quantitatively small categories in either one or both of the two classifications may be aggregated, or alternatively larger categories disaggregated, so that actual 1 : M and N : M HS-CPC correspondences are converted into a 1 : 1 correspondence between the two classifications. If the CPC detail is reduced in line with HS export and import data availability, one should take into account the impact that such reduction has on the CPC-COICOP correspondence dealt with in the next paragraph, and on the CPC-ISIC correspondence discussed in the next section.

The conditions set out above are also satisfied in the present international COICOP-CPC correspondence, which links output to HH final consumption. The international COICOP has much fewer categories (116) than the final consumption categories that can be identified in the international CPC (819), there is an M : 1 correspondence between the CPC and COICOP. If the detail of the CPC is adjusted at the national level, because some product categories are eliminated as they are not relevant nationally, detail has been added for national purposes, or CPC categories have been aggregated because data detail is not available—the

[10]This may occur when countries are still using the SITC or a previous version of the HS, i.e. the BTN (Brussel's Trade Nomenclature), to classify foreign trade. Both classifications have less detail than the present HS. The SITC has 3122 categories at the most detailed level, compared with the 5224 categories of the HS.

CPC-COICOP correspondence should be also adjusted so that the M:1 correspondence is maintained.

Another consistency requirement that should be met relates to the BEC-CPC correspondence. If the BEC is extended to goods and services of the CPC, as is suggested in Section 2.2.4 below, it would identify within the CPC all goods and services that correspond to HH final consumption. This scope, of course, would need to be made compatible with the scope of the 819 CPC categories of HH final consumption that are identified through the CPC-COICOP correspondence.

2.2.3. Coordinate ISIC-CPC Correspondences

With regard to the technical integration of production data, the objective is to coordinate ISIC and CPC detail in such a manner that the meso SUT reflects as well as possible the production structure at the micro level. This refers in particular to the distinction between principal and secondary products, which can be only made well at the micro level of the establishment. If aggregations of ISIC and CPC categories are not well coordinated, at the meso level a distinction between characteristic and non-characteristic products results, which does not match the micro distinction between principal and secondary products. In particular non-characteristic products at the meso level may identify not only secondary products, but also other non-characteristic products that are only included as a result of the aggregation of ISIC and CPC categories.

The distinction between principal and secondary products can be made both in a square and also in a rectangular ISIC-CPC output matrix. In a square output matrix the diagonal elements refer to principal products, and the non-diagonal elements should only include secondary products that are identified as such at the micro level. In a rectangular matrix the secondary products identified at the micro level should show up at the meso level as the entries with minor values in each industry column. A more accurate distinction between principal and secondary products can be made at the macro level by introducing more detail in the CPC breakdown and thus making the output matrix more rectangular.

The percentage shares of non-characteristic products in total output at four different levels of aggregation of ISIC and CPC categories may illustrate the impact of ISIC and CPC aggregation.[11] At the most detailed level of both (13 ISIC and 23 CPC categories), the share of non-characteristic products is 9.3 percent. If one assumes that the output table at this level of aggregation is a fair representation of the distinction of principal and secondary products at the micro level, the share of non-characteristic products only refers to secondary products, and characteristic products are principal products. When the number of CPC categories is aggregated to 10 only, while leaving the number of ISIC categories (13) unchanged, the share of non-characteristic products decreases to 6.2 percent, as some secondary products that were identified at the previous level of aggregation, are now included with the value of characteristic products. On the other hand, if the

[11]The example is based on the illustrative output data presented in table 15.1 of Chapter XV of the 1993 SNA.

number of ISIC categories is reduced from 13 to 8 (leaving the number of CPC categories at 23), the share of non-characteristic products increases dramatically to 44.1 percent. This is because products that were identified as principal products at the previous level of aggregation, are now included with non-characteristic products. If both ISIC and CPC are aggregated to fewer categories (8 ISIC and 10 CPC categories), the share of non-characteristic products is somewhere in between that of the two previous values of the shares, i.e. 15.1 percent. In all instances except the first one, the share of non-characteristic products differs from the share of secondary products, which was assumed to be 9.3 percent as measured in the most detailed output table. Whether the assumption is correct that the distinction between principal and secondary products is well reflected at the meso level of the first ISIC/CPC aggregation, of course, can only be verified after a detailed quantitative investigation of the scope of secondary products at the micro level.

2.2.4. Use the BEC to Trace Imports and Outputs to Uses

Once the CPC is recognized as the core classification of the SUT, further refinements could be introduced, when applying the international classification in countries the refinement may include:

- Distinctions, where possible, within the CPC categories between products that are imported and those that are produced in the country.
- Redefinition of the present BEC-SITC link to a BEC-CPC link and thus cover services as well as goods, and apply the BEC not only to imports, but also to output.
- Limitation of the number of CPC categories that corresponds to more than one BEC category.
- Extension of the BEC from the traditional distinction between intermediate consumption, HH final consumption and capital formation, to include—at the country level—also links with exports, GOV and NPI consumption.

If these refinements are introduced to a maximum extent, most products would have one origin—imports or output—and only one of the destinations mentioned. These constitute the five attributes (p, m, ci, c, k and x) that are indicated in the scheme in Table 2. To these are added the distinction between goods and services (g,s), as that is a distinction which is made between two foreign trade classifications, i.e. HS and EBOPS. These seven attributes are used in the center of the table to identify the attributes for reclassifying the data of statistical sources for the collection of data on supply (row) and use (column) of products. The headings of rows and columns indicate respectively the data sources, SNA items and type of classification categories that need to be reclassified in this manner. For instance in the second row on the supply side, foreign trade merchandise data on imports (m) from the customs register that are classified by HS, should be reclassified into three groups, i.e. imported (m) goods (g) destined for intermediate consumption (m/g/ci), capital formation (m/g/k) and HH final consumption (m/g/c). Similarly in the last column on the use side, data on HH final consumption (c) from the HH survey should be reclassified into four groups, i.e. HH final

TABLE 2

ATTRIBUTES OF PRODUCT FLOWS IN THE SUT

Use → / supply ↓		Data source	Economic survey, GOV registers	Customs register	BOP	Household survey
		SNA item(s)	Intermediate consumption, capital formation	Exports of merchandise (goods)	Exports of services	HH final consumption
Data source	SNA item(s)	Classification	CPC	HS	EBOPS	COICOP
Economic survey, GOV registers	Output	CPC	p/g/ci p/s/ci p/g/k p/s/k	p/g/x	p/s/x	p/g/c p/g/c
Customs register	Imports of merchandise (goods)	HS	m/g/ci m/g/k			m/g/c
BOP	Imports of services	EBOPS	m/s/ci m/s/k			m/s/c
Data attributes	p	Output	c	HH final consumption	g	Goods
	m	Imports	k	Gross fixed capital formation	s	Services
	ci	Intermediate consumption	x	Exports		

consumption goods domestically produced (p/g/c) and imported (m/g/c) and HH final consumption services domestically produced (p/s/c) and imported (m/s/c).

If all classification categories corresponding to the SNA items in the data sources indicated were used on both the supply and use side, it would result in a maximum number of data checks between supply and use, as the same data categories would be available on both sides.

In order to accomplish this, however, existing international classifications and correspondences, when used in countries, would need to be:

- Utilized in the maximum detail feasible. If less detail is used, regrouping of the data from the sources as indicated would be more difficult.
- Amended in line with products that are locally produced or used, and for which the origin and destination can be clearly identified. Thus HS and EBOPS would need to be adapted for local use, so that these classifications only include categories of goods and services that are actually imported or exported by the country, and COICOP would need to be amended for national use so that it includes only products that are locally consumed by HHs.

What is described above may be called an ideal situation, which in practice may be difficult to accomplish, given data restrictions. Already in the present BEC there are several categories of HS, which are either used for HH consumption or capital formation, HH consumption or intermediate consumption, or for either one of the three categories, i.e. HH consumption, intermediate consumption or capital formation. If the detail of the HS as used in a country were less than the detail available in the international HS, the number of product categories with combined distinctions may be increased further. On the other hand, if an effective adaptation of the HS to local circumstances were made, this problem may be mitigated. Similarly, it may difficult on the use side to distinguish between domestically produced products and imported products. But also in this case, the problem may be reduced by amending the classifications on the use side to locally produced or used products, for which it would be easier to determine which ones are imported and which ones are domestically produced.

2.3. *Classification Refinements Incorporated in a Simplified SUT Framework Within the Extended SNA*

The framework design and the classification refinements discussed above have been reflected in Table 3 in an operational format of the SNA-SUT. The SUT presented is comprehensive in the sense that it includes all classifications and cross-classifications that might be used to refine the detail of the SUT. On the other hand, it is referred to as a simplified SUT framework, as it includes various simplifications, particularly in the cross-classification of data. The SUT table has been designed for purposes of compilation and analysis of data, and may thus be used in the actual SNA compilation in countries.[12]

The features of the SUT that derive directly from Table 1 are reflected in the columns of the sectors that are identified in the left hand columns of Table 3, i.e.

[12]The author has developed simplified SNA frameworks for Guyana, Oman, Nicaragua and St. Vincent, along the lines of Table 3.

TABLE 3

SIMPLIFIED SUT

SECTORS/ISIC		NFC	FC	GOV	HH	NPI	NFC/HH	NFC	NFC/HH	NFC	NFC/HH	NFC/HH	NFC/HH	NFC/FC/HH	GOV	NFC/GOV/HH/NPI	Total Economy and ROW
							Agriculture, Forestry, Fishing (A,B)	Mining and Quarrying (C)	Manufacturing (D)	Electricity, Gas and Water Supply (E)	Construction (F)	Wholesale and Retail Trade; Repair ofMotor Vehicles, Motorcycles and Personal and Household Goods (G)	Hotels and Restaurants (H)	Transport, Storage, and Communication (I)	Financial Intermediation, Real Estate, Other Business Services, Private Household Services (JPK)	Public Administration and Defense; Compulsory Social Security (L)	Education, Health, Other Community, Social and Personal Services (MNO)	
Characteristic CPC							01–04	11–16	21–49	69	53–54	61, 62	63	64–68, 84	71–83, 85–87	91, 99	92–98	
Total output		1,753	102	440	1,269	40	89	35	1,680	164	244	123	68	100	478	168	455	3,604
Selected CPC categories	21–25								448									448
	26–31								422									422
	32–37, 41–42								796									796
	64													44				44
	66													31				31
	72–83, 85–87																98	98
	92–94																276	276
	95–98																79	79
Imports							37	60	283	1	1	79	0	20	13	0	5	499
Taxes less subsidies on products							2	0	89	5	17	0	3	5	8	0	4	133
Trade and transport margins							2	2	74	0	0	-68	0	-10	0	0	0	0
Supply at market prices							130	97	2,126	170	262	134	71	115	597	168	366	4,236
Intermediate consumption (directly estimated by industry)		899	29	252	694	9	47	17	1,000	127	114	64	36	60	132	118	168	1,883
Value added		854	73	188	575	31	42	18	680	37	130	59	32	40	346	50	287	1,721
Employment							2,058	292	28,682	3,300	5,024	5,000	2,078	2,032	3,700	8,000	9,203	69,369
Changes in inventories (directly estimated by industry)							1	-1	15	0	23	0	0	0	0	0	0	38
Gross fixed capital formation (directly estimated by industry)							11	6	104	13	9	13	7	39	144	12	18	376
Intermediate consumption (allocated by product)							88	96	992	123	40	61	0	78	309	1	95	1,883
HH final consumption					1,015		28	2	567	36	3	51	0	14	250	6	58	1,015
GOV final consumption				368			2	0	3	0	0	0	0	0	0	159	204	368
NPI final consumption						16	0	0	0	0	0	0	0	0	0	2	14	16
Gross fixed capital formation (allocated by product)							2	0	161	0	190	0	0	0	23	0	0	376
Changes in inventories (allocated by product)							1	-1	15	0	23	0	0	0	0	0	0	38
Exports							7	6	422	1	6	0	67	19	8	0	4	540
Use at market prices							128	103	2,160	160	262	112	67	111	590	168	375	4,236
Supply-use discrepancies							2	-6	-34	10	0	22	4	4	7	0	-9	0

Notes: Shaded rows refer to products (CPC) of the supply and use identity.
Non-shaded rows refer to components of the production accounts of sectors and industries (ISIC).
The non-shaded row of "Output" is included both in the production accounts of industries and in the supply and use identity.

NFC, FC, GOV, HH and NPI. The additional columns on the right hand side refer to the ISIC categories of industries that were discussed in Section 2.2.1, and that traditionally are used for the industry breakdown of GDP and other production related variables. The right hand column of the table is a combination of the columns for the National Economy and ROW in Table 1. Thus it includes all totals, which are aggregate data of resident sectors, and also some ROW data, particularly exports and imports. Most items are specified by sectors and their totals are included in the National Economy & ROW column: i.e. output, intermediate consumption, value added and final consumption by HH, GOV and NPI sectors. Some items are only available in totals and presented only in the National Economy & ROW column, i.e. gross fixed capital formation, changes in inventories, and taxes less subsidies on products. The supply-use identity holds, but only for the items in the National Economy & ROW column and not for the sectors individually.

The CCIS cross-classification of ISIC and sector categories is incorporated by indicating in the first row above the industry groupings, the sectors to which they correspond.[13] For instance, the ISIC category "Agriculture, forestry, fishing (A,B)" corresponds to the NFC and HH sectors, and the ISIC category "Public administration and defence; compulsory social security (L)" is allocated to the GOV sector. The CCIS correspondences only cover the items of output, intermediate consumption and value added, which are classified alternatively by sector on the left hand side and by ISIC categories on the right hand side of the table. The totals of the three categories are the same for both types of classifications (respectively 3,604, 1,883, 1,721). Other items, including employment, changes in inventories, gross fixed capital formation (directly estimated by industry), have been classified by ISIC categories only.

The CPC breakdown has been incorporated in the table in two ways. Firstly, CPC categories have been included as correspondences to ISIC categories at the heading of each ISIC column. Thus, CPC categories 01–04 correspond to ISIC category Agriculture, forestry, fishing (A,B), or CPC 53–54 corresponds to ISIC Construction (F). The ISIC-CPC correspondence is used to incorporate in the columns on the right had side of the table not only the production categories of output, intermediate consumption (directly estimated by industry), value added, employment, changes in inventories (directly estimated by industry) and gross fixed capital formation (directly estimated by industry), classified by ISIC categories, but also the supply and use categories of output, imports, intermediate consumption (allocated by product), HH final consumption, gross fixed capital formation (allocated by product), changes in inventories (allocated by product), and exports, which are classified by CPC categories. The right hand side of the table thus integrates the ISIC columns and CPC rows of the traditional SUT, using the assumption that there is a correspondence between ISIC and CPC categories. The distinction between the ISIC and CPC categories in the table has been made by shading the rows corresponding to the CPC product categories. This needs two qualifications. The first one is that non-shaded output is classified by both ISIC and CPC, and the other one is that there are three items of intermediate con-

[13]The ISIC-sector correspondences are based on those included in table 15.3 of the 1993 SNA.

sumption, gross fixed capital formation and changes in inventories, i.e. one for each directly estimated by industry (ISIC), and another one allocated by product (CPC).

In the table, as in practice, the assumed ISIC-CPC correspondence only holds approximately, and this is reflected in the statistical discrepancy at the bottom of the table between supply and use. In order to mitigate the impact of this assumption on the accuracy of the SUT, further refinements have been introduced in a second appearance of CPC categories in the rows below total output, reflecting the classification coordination criteria developed in Section 2.2.3. In each industry column have been presented among others the secondary product categories that are included in the output of one industry, but characteristically belong to another industry. An example is in the last column of "education, etc. (ISIC:MNO/CPC: 92–98)," where a figure of 98 appears as secondary output corresponding to CPC 72–83, 85–87, which characteristically corresponds to ISIC category "Financial intermediation, real estate, other business services, private household services (JPK)." With this CPC refinement in the rows, supply of CPC category 72–83, 85–87 has been augmented with secondary output produced elsewhere (597 = 478 + 98 + 13 + 8), and supply of CPC category 92–98 was correspondingly reduced (388 = 455 – 98 + 5 + 4). By introducing these refinements in the CPC rows related to secondary output the statistical discrepancy at the bottom of the table is reduced. The advantage of this approach is that a comprehensive output table is not needed, and only a selection of quantitatively important secondary outputs is introduced as rows below total output. This reduces the amount of work of national accountants on quantitatively unimportant secondary output categories, improves the reliability of the quantitatively important ones and avoids the large number of zero (0) and very small data elements presently included in large output matrices.

The CPC rows have not only been used to identify secondary output, but also to identify products that would help to trace better the correspondence between supply and use of products as discussed in Section 2.2.4. Thus, for instance in the column of "Manufacturing (D)" have been distinguished three product categories: 21–25 (Food and tobacco), 26–31 (Textile, leather and wood products), 32–37, 41–42 (Low-tech manufacturing products for industrial use).[14] The first group of products is predominantly destined for HH final consumption, the second group for exports and the third group for intermediate consumption, according to a BEC adapted to the country in question. By making this additional CPC distinction it is thus possible to allocate from the supply side output to its corresponding use category and compare those estimates with the direct estimates of each destination category. Further CPC detail could also be incorporated in the case of imports, when the total value of imports recorded in each column does not correspond in its entirety to the CPC category indicated in the heading of the column. Additional ISIC detail could furthermore be incorporated in additional rows below "Intermediate consumption (allocated by product)," if for selected ISIC categories more CPC detail on intermediate inputs is available from surveys and other data sources.

[14]The category "Low-tech manufacturing products for industrial use" referring to categories 41–42 is not an aggregate CPC category, but has been created solely for the purpose of this paper.

3. Conclusion

The simplified SUT framework including the use of classifications and cross-classifications to introduce further detail is flexible enough to incorporate various analytical options. Thus, one should keep in mind that not all details suggested above are to be implemented in practice when designing the SUT framework. This conclusion applies not only to the limited segment of the SUT, but also to the larger SNA framework and to analytical frameworks in general. Additional details require additional data, and this may sometimes be too costly and/or too time consuming, as compared to the analytical benefits obtained. Therefore the actual format of a framework should be decided in close cooperation between national accountants and policy analysts, who can only determine together what would be the most optimal scheme from a statistical and analytical point of view. Thus, in the case of the SUT, they may need to decide whether the ISIC breakdown be used solely without CPC detail, in which case it should be accepted that only limited data checks and analytical indicator ratios are available. Or alternatively, more detailed CPC breakdowns may be introduced, which increases the data requirements of the framework, but also increases the number of data checks and the analytical potential of the framework. Incorporating cross-classifications between ISIC and sectors would further enhance the analytical potential of the framework, as it would allow joining GDP growth analysis based on ISIC and CPC data of the SUT with fiscal, monetary and financial analyses based on the IEA.

References

Commission of the European Communities–Eurostat, International Monetary Fund, Organization for Economic Cooperation and Development, United Nations, World Bank, *System of National Accounts, 1993* (United Nations publication sales no. E.94.XVII.4).

FAO, *System of Economic Accounts for Food and Agriculture*, 1996 (ISBN 9251038074, Job Number W0010/E).

Hoffmann, E., "Developing Labour Account Estimates: Issues and Approaches," in *Household Accounting: Experiences in the Use of Concepts and their Compilation. Vol. 2: Household Satellite Extensions, Handbook of National Accounting,* United Nations, 2000.

ILO, *Compendium of Official Statistics on Employment in the Informal Sector*, by Ralf Hussmanns and Brigitte du Jeu, STAT Working Paper No. 1, Geneva, 2001.

IMF Statistics Department, *Balance of Payments Manual*, Washington, D. C., September 1993.

———, *Monetary and Financial Statistics Manual*, Washington, D. C., September 2000.

———, *Government Finance Statistics Manual*, Washington, D. C., 2001.

Korea (Republic of) National Statistical Office (Kyung Hee Kim, Eun Pyo Hong, Eun Soo Lee), UNSD (Jan W. van Tongeren and Berndt Becker), UNDP, *Human Resource Accounts for Korea: A Study of Education and Employment*, Seoul, October 2000.

Magnus, Jan, CentER Tilburg University, Jan W. van Tongeren, UNSD, Aart F. de Vos, Free University, Amsterdam, National Accounts Estimation Using Indicator Ratios, *Review of Income and Wealth*, 6(3), September 2000.

Ruggles, Richard and Nancy, D. Ruggles, "The Role of Micro Data in the National Economic and Social Accounts," *Review of Income and Wealth*, 21(2), June 1975.

Ruggles, Nancy, D. and Richard Ruggles, "The Treatment of Pensions and Insurance in the National Accounts," *Journal of the International Association for Research in Income and Wealth*, 29(4), December 1983.

———, "The Integration of Macro and Micro Data for the Household Sector," *Review of Income and Wealth*, 32(3), September 1986.

———, "Household and Enterprise Saving and Capital Formation in the United States, A Market Transactions View," *Review of Income and Wealth*, 38(2), June 1992.

Thorbecke, Erik, Cornell University, "The Use of Social Accounting Matrices in Modeling," paper presented at the International Association for Research in Income and Wealth 26th General Conference, Cracow, Poland, 27 August to 2 September 2000.

United Nations, *Manual X, Indirect Techniques for Demographic Estimation*, 1983 (sales no. E81.XIII.7).

———, *Integrated Environmental and Economic Accounting,* 2004.

Van Tongeren, Jan, "Development of an Algorithm for the Compilation of National Accounts and Related Systems of Statistics," *Review of Income and Wealth*, 22, March 1986.

Young, David M., *Iterative Solutions of Large Linear Systems*, Academic Press, London and New York, 1971.