Re-engineering of eNAQ - An electronic National Accounts Questionnaire Based on 1993 & 2008 SNA

Jens Chorus – Hans – J. Lenz[1]

1. Introduction

The prototype "eNAQ" - an electronic National Accounts Questionnaire - was developed for the electronic capturing and administration of National Accounts Questionnaires (NAQ) at the United Nations National Account Section (UN NAS).

In 2005 ENAQ was first implemented by Eichler (2004)) based on a requirement analysis and design specification of Müller (2003)). Recently, a complete re-engineering was undertaken by the first author, paying special regard to the additional requirements derived from the pending switch to the 2008 SNA version.

The task of eNAQ is to support all users (experts and administrators) at national and international organizations to efficiently enter data in a user friendly way by a personalized user interface (GUI). The systems novelty is to allow users to enter data worldwide, and to assist the NA experts in validating the data at the point of entry. eNAQ is not meant to pre-describe, produce or even dominate how to produce statistical annual reports on any national level, nor to compute automatically statistical data. eNAQ is a software tool supporting data entry backed up by an integrated validation rules checker. The rule system is based on internationally agreed rule sets, yet, allowing for national particularities of national accounts methodologies.

This paper only summarizes main features of the prototype, and the revised requirements according to the 2008 System of National Accounts (SNA). The features presented are based on the NAQ and the system currently used at the United

[1] Institute of Statistics and Econometrics, Freie Universität Berlin

Acknowledgment: The authors express their gratitude to Jörn Eichler, Darmstadt, for his contributions to clarify and improve parts of this paper as far as the implementation of eNAQ is concerned. They are very thankful to Patrick Müller, Washington, who always helped us to keep track of 2008 SNA new directions and changes. We feel free but are not authorized to call him "The Third Man".

Nations. The prototype and requirements analysis include visionary elements and features, which are not necessary part of current mission and strategy of the UN National Accounts Section. Major parts of the 2003 specifications where implemented with the new system at the United Nations, other essential parts were implemented later via satellite systems. The core part of eNAQ, the local data entry and validation by the country reporting to the United Nations, was never implemented. A lot of the work with respect to data entry and validation at National Statistical Offices (NSOs) and the UN NAS is still done manually and could be efficiently supported by eNAQ.

The envisioned benefits from using eNAQ are a simplified data transmission between NSOs and the UN NAS that don't require double manual data entry or editing for the UN. Data can be easily transferred with eNAQ via Internet, and, where Internet is not available, import and export functions guarantee a flawless data transfer. Only one single data entry effort needs to be done by the country office providing the data. In real time an automatic electronic validation is performed. Data submissions and entry using other sources (paper publications, pdf-files, or individualized excel tables) that are time consuming, costly and vulnerable to mistakes should become the exception. The main goal is the acceleration and simplification of workflows, reducing data handling cost while improving data quality.

2. National Accounts Compilation

Some figures are displayed to illustrate the complexity of the annual UN NAS report. The UN NAQ for the 1993 SNA consists of 31 tables containing in total 1046 items. The current rule set for the 1993 SNA contains of 814 rules that need to be validated. Note that additionally, revised country data beyond the coverage of the standard NAQ years, have to be updated and validated in the UN NAS database when it becomes available. Let us assemble some facts about national accounts compilation. NA seems to play a role in economics and national statistics similar to the role mathematics plays in science: National accounts and, of course, input-output analysis is the "Queen". NA delivers information on main statistics as economic indicators for Government, Parliament, Press, and Science. It is an ensemble of flows, stocks and values per units (VpUs). It assumes that the fundamental balance equations of an economic equilibrium are fulfilled: Supply ≡ Demand (EX-Post Identities). The main economic indicator system includes many indicators of which the Gross Domestic Product (GDP) is not the only but the most prominent one. For instance, it helps to determine per capita GDP using the population size PS and GDP.

The current world wide standardized and fairly harmonized system of national accounts (SNA) - with respect to definitions, measurements and quality - is a

distributed, loosely federated information system. This is true in the sense that national data is collected locally, autonomously on heterogeneous hard- and software platforms owned by very divergent countries, and validated using a national set of validation rules of type "edits" or "if…then" rules. Furthermore, there are supra-national rule sets for each pair of countries used for
validation on the international level. A very simple, nearly trivial example of a rule is the "Production in a Sector" - Rule:

$$Intermediate\ Consumption_{Sector} + Value\ Added_{Sector} = Output_{Sector}$$

NA computing is in its core 'Estimation', a term precisely coined in Mathematical Statistics. Missing units like companies or citizens, missing values ("null values" termed by IT specialists), outliers, non-linearity of relationships like sales=quantity*price/unit, non-Gaussian distributions, or national economic policy, the human behavior in general, or even strikes of labor-unions or economic sectors influence its timeliness, quality (precision, certainty) and cost of computing. Moreover, think of the existence and size of black
markets around the world.

Finally, it is worthwhile mentioning that various Excel-like tables which represent the annual NA reports are mandatory and necessary for any input-output-analysis, both on the national as well as international level. Take for instance Germany as a country, the EU (the European Union) as one out of many political or economical groups or even the world (=ALL) representing the world's economy status per toto at the UNO, IMF or World Bank.

3. The System of National Accounts

The System of National Accounts is a set of standard recommendations regarding definitions, concepts, classifications and foremost, accounting rules for the measurement and compilation of economic activities. Harmonization is a major problem in NA. For example, there is a lot of skepticism within the European Union regarding what is referred to as an over-standardization. The argument is that some national differences are vital to European diversity, as expressed by the second author in a meeting of EuroStat a couple of years ago, in the term the "Standardized European Tomato". A well designed "European Federated
Statistical Information System", as firstly proposed by Lenz and Shoshani (1996) could provide a viable solution for the compilation of NA data. While the technical prerequisites in form of software, hardware and middleware have been available for more than 15 years, political power to implement such a solution remains questionable.

Reasons for divergent statistical systems are as follows:

- Divergent efforts in analysis of statistical data
- Lacking autonomy of statistical institutions
- Budget restrictions
- Varying intellectual potentials and traditions in the classic fields of economics and statistics
- Divergent measures of services and financial assets
- Heterogeneous handling of insurance premiums
- Income from shares and other equities
- Different statistical definitions and approaches
- Measuring economic activities varies across membership countries
- Size of black markets
- Degree of illiteracy
- Degree of mobility of people and assets
- Data capturing facilities.

A further discussing of the problems of harmonization of statistical systems is out of the scope of this paper, and would drive us far away from transmitting the main ideas about an electronic questionnaire. Instead, below we will elaborate the current standard of NA computing, illustrating that it is viewed mathematically, rather then statistically, while using simple arithmetic operations. Let us remark, that many economic and IT experts underestimate the role and influence of applied statistics in conjunction with modern IT technology for improving data quality while reducing cost of data production. As stated by B. Efron, Stanford University: *"Those who ignore Statistics are condemned to reinvent it".*

The SNA offers three different approaches for the computation of the main economic indicator Gross Domestic Product (GDP). While various approaches to compute indicators are quite common for economists and other "soft domain scientists", civil engineers using precise measurements will stumble over this fact, as their metering is usually dependent upon one and only one measurement technique. The three approaches for computing NA are:

- Production Approach
    - Aggregation of Production data
- Expenditure Approach
    - Aggregation of Demand data
- Income Approach
    - Aggregation of Income data.

Below, the GDP computation of each approach is listed introducing a number of intermediate variables. As all these examples equate to the same indicator, GDP, the reader can deduce that all these variables form an equation system of main economic indicators based on relations given by given by definitions. According to

our main topic and the space limitations of this contribution the main indicator system will remain "behind the scene"[2].

Utilizing the *Production Approach* the GDP can be compiled as follows:

+ Output value (at basic prices)
- Intermediate consumption
= Gross value added (GVA)
+ Taxes on production
- Subsidies on production
= Gross Domestic Product (GDP)

The GDP can also be compiled by the *Expenditure Approach*

+ Final consumption expenditures (households, government and NPISH)
+ Gross capital formation (corporations, households, government and NPISH)
+ Exports
- Imports
= Gross Domestic Product (GDP)

Finally, the GDP can be computed using the *Income Approach*

+ Compensation of employees (wages and salaries, employers social contributions)
+ Other taxes on production
- Other subsidies on production
+ Taxes on products
- Subsidies on producs
+ Operating surplus
+ Mixed income (unincorporated enterprises owned by households)
= Gross Domestic Product (GDP)

[2] The interested reader is referred to the Pubns (2009), Nations et al. (2008), Nations et al. (1993),
or Wikipedia

At this point the reader should remind himself that using these approaches for computing GDP will most likely lead to statistical discrepancies. Hence, any questionnaire must allow an entry of statistical discrepancies for the variables. It will be shown later how eNAQ handles those deviances.

4. System Analysis of the SNA and the UN NAS

There are various aspects concerning UN NAS activities, i.e. the present situation at the UN NAS, the variants of questionnaires, and drawbacks of the working process:

- Situation at the UN Statistics Division
    - Tasks
    - Workflow
    - Budget
- National Accounts Questionnaire
    - Questionnaire for the 1993 SNA
    - Future questionnaire for the 2008 SNA
- Weak points and bottlenecks in the workflow

Task at the UN NAS

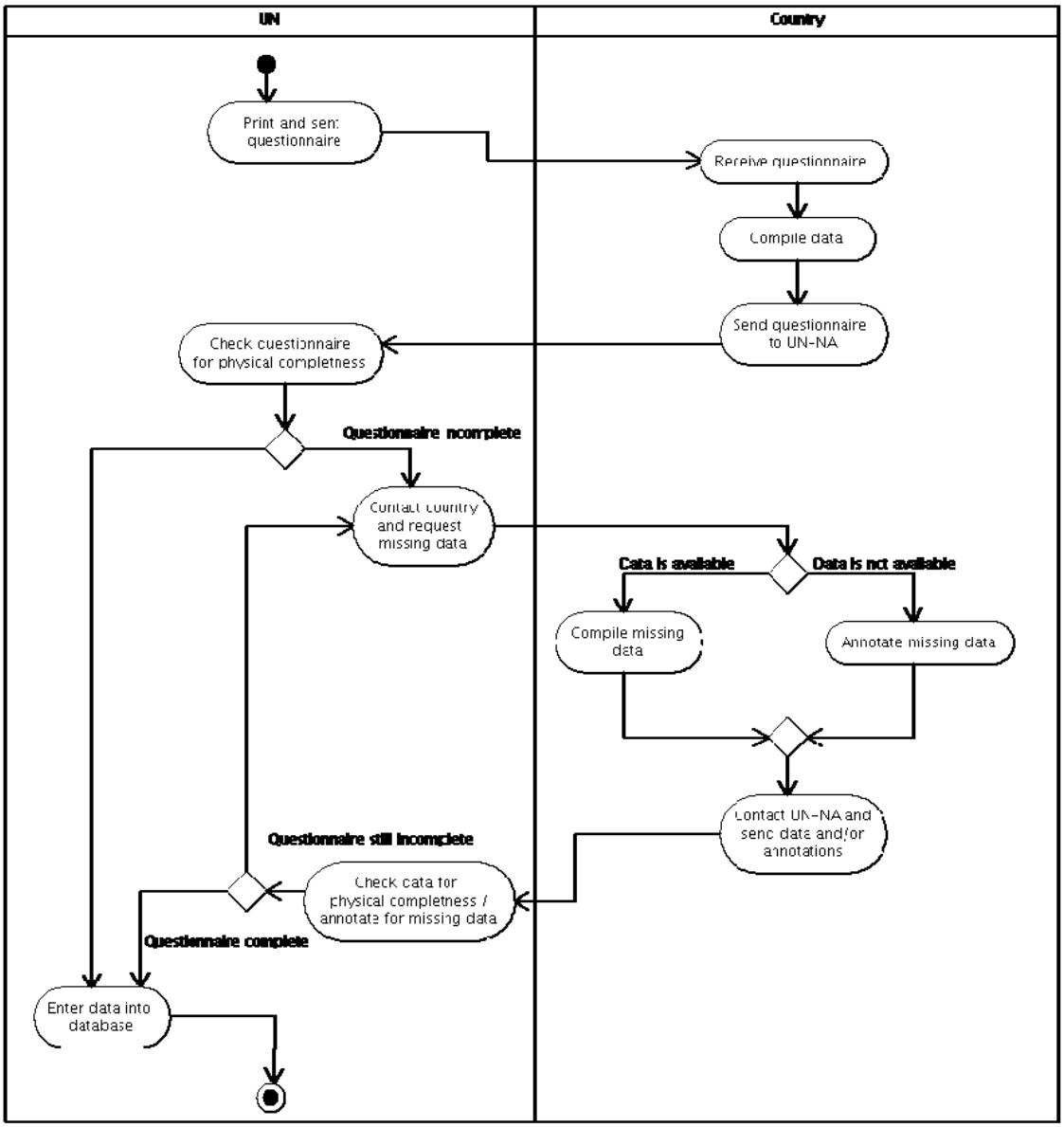
Tasks of the National Accounts Section include:

- Collection and compilation of data received from member countries of the UN
- Contribution to the statistical Yearbook published by the UN
- Publication of the NA Yearbook: National Accounts Statistics: Main Aggregates & Detailed Tables.
- Publication of National Accounts Statistics: Analysis of Main Aggregates.
- Provision of data to governments, the private sector, or Universities for scientific purposes etc.
- Provide support for the member countries in data compilation and adherence to the guidelines developed in the SNA

Workflow

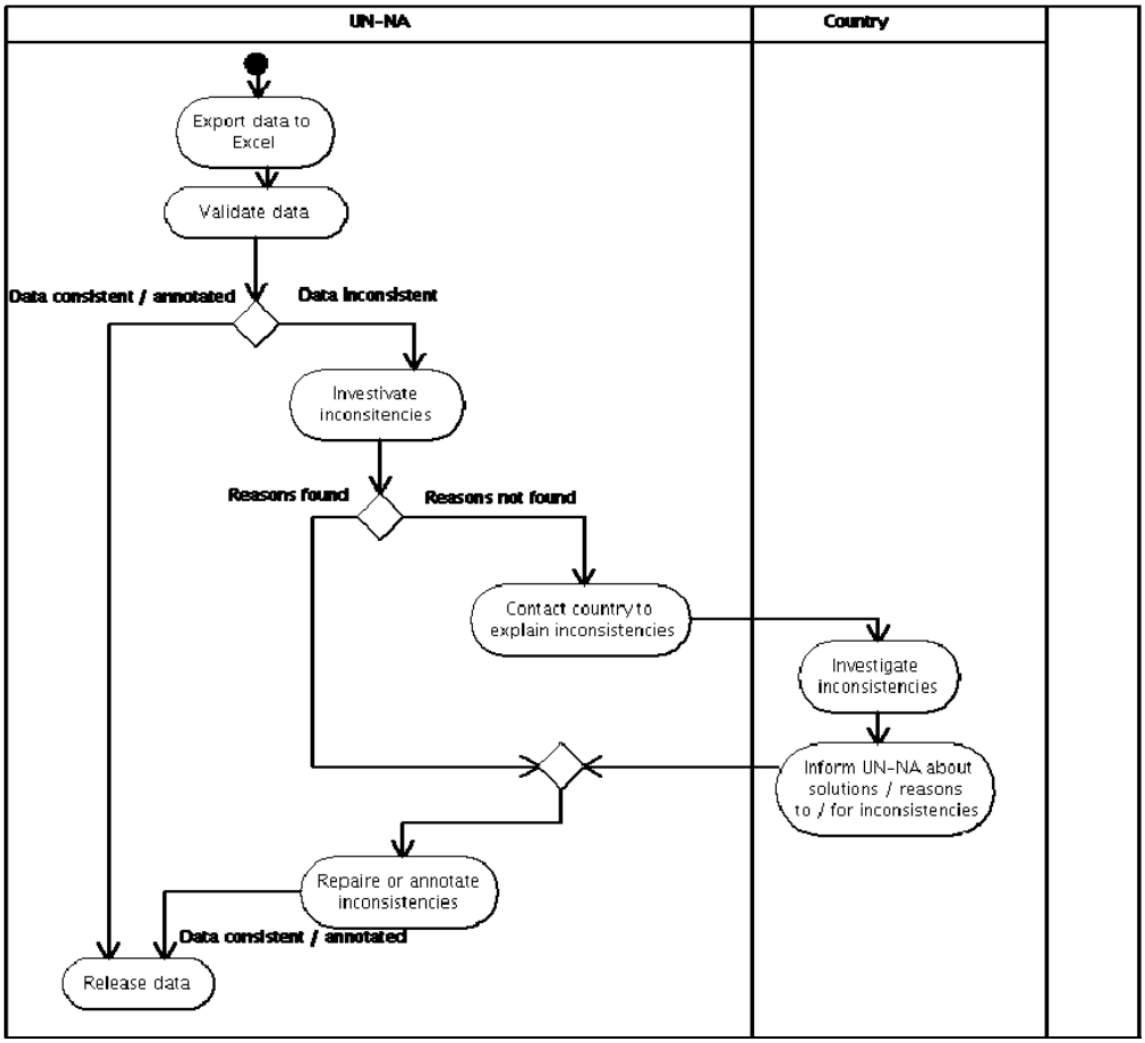
The illustrations 1 and 2 represent the data capturing process and its validation as currently done by the UN.

Illustration 1. Current Workflow "Enter data of a country's questionnaire"



The diagrams are self explaining, but let us stress the high number of data transfers between the UN NAS and breaks which require manual data input.

Illustration 2. Current Workflow "Validate and release country's data"



4.1. The National Accounts Questionnaire (NAQ)

A NAQ consists of the following elements:

Country Notes
The *country notes* provide information regarding the national institution that compiled the data, a general note regarding the kinds of tables provided, the method

of computation, special remarks concerning the countries' data and a specification of the method used for GDP computation, cf. Nations et al. (1993)).

Conceptual Table

The *conceptual table* is an addition to the NAQ and was used to collect information on the progress of SNA 1993 implementation.

National Accounts Data

The *national accounts data* is comprised of various tables within the categories briefly summarized below.

- Main Aggregates
  - In this part of the NAQ, three tables are requested that summarize information on gross domestic product by expenditure in current and constant prices (Tables 1.1 and 1.2), as well as information regarding the relations among the main aggregates: production, income, saving, and net lending aggregates (Table 1.3).
- Domestic Production by Industries
  - The tables 2.1 and 2.2 compile the GDP in current and constant prices summing the individual GDP value of industries as defined by the ISIC Rev.3.1[3]. Table 2.3 computes the contribution of these industries to various main aggregates, such as gross value added, gross capital formation, and others.
- Classification of Expenditure According to Purpose
  - Table 3.1 requests data on the government consumption expenditure, classified according to the 10 divisions of COFOG[4]. Table 3.2 requests similar information for individual household consumption expenditure, individual consumption expenditure for NPISH[5] and individual consumption expenditure of general government, classified according to COICOP[6].
- Integrated Economic Accounts
  - The integrated accounts assemble information for the institutional sector[7] accounts, the rest of the world accounts, and asset and liabilities accounts. They show principal economic relations as well as the main aggregates. In theory, they form an accounting framework in which the sum of resources and uses of the total economy should balance to zero with those of the rest of the world.

[3] International Standard Industrial Classification

[4] Classification of the Functions of Government

[5] Non Profit Institutions Serving Households

[6] Classification of Individual Consumption According to Purpose

[7] Sectors are the Total economy, Rest of the world, Non-financial corporations, Financial corporations, General government, Households, and Non-profit institutions serving households

- Cross Classification of Gross Value Added by Industries and Institutional Sectors
  - The table 5.1 requests the individual contributions of sectors to the gross value added by industries, as classified by the ISIC Rev.3.1.

4.2. Changes Introduced by the 2008 Revision of the SNA

To date the UN NAS has not yet published a revised version of the NAQ for the 2008 SNA. The revised NAQ will also reflect the new ISIC Rev. 4. The study of the new framework suggests that no new tables will be introduced. However, the tables themselves will be changed, introducing new items, removing and renaming items. This will affect the structure of the tables as well as the relations between their items[8], and will approximately double the numbers of 1046 items and 814 rules mentioned above.

For illustration purposes a data and a conceptual table is presented below. Note, that entries for twelve successive years are requested within the data table. This provides an opportunity for the countries to update already submitted data. However if data were revised for more than the years shown in the NAQ, this data should be provided to UNSD, too., in order to maintain the latest available data for all the countries in the UN National Accounts database.

Table 1. Gross domestic product by expenditures at current prices - NAQ: 1.1

Item Description	Item Code	S	1995	1996	...	2006
EXPENDITURES OF THE GROSS DOMESTIC PRODUCT						
P.3 Final consumption expenditure	10107					
P.3 Household final consumption expenditure	10108					
P.3 NPISHs final consumption expenditure	10109					
P.3 General government final consumption expenditure	10110					
P.31 Individual consumption expenditure	10111					
P.32 Collective consumption expenditure	10112					
P.5 Gross capital formation	10113					
P.51 Gross fixed capital formation	10114					
P.52 Changes in inventories	10115					
P.53 Acquisitions less disposals of valuables	10116					
P.6 Exports of goods and services	10117					
P.61 Exports of goods	10118					
P.62 Exports of services	10119					
P.7 Less: Imports of goods and services	10120					
P.71 Imports of goods	10121					
P.72 Imports of services	10122					
Plus: Statistical discrepancy	10123					

[8] The relations between the items are also referred to as *rules*. They will be elaborated below.

Table 2. Conceptual Table 1993 SNA

	Yes	No	Partly	Not Applicable
Is government defence expenditure on fixed assets that can be used for civilian Purposes included in GCF?				
Is consumption of fixed capital included on all government fixed assets (airfields, roads, hospitals, docks, dams and breakwaters and other forms of construction)?				
Is all successful mineral exploration capitalised?				
Is all unsuccessful mineral exploration capitalised?				
Are purchases of computer software included in GCF?				
Is expenditure on software development on own account and for sale included in GCF and output?				
Is expenditure on entertainment, literary or artistic originals included in GCF and on Their development included in output?				
Is expenditure on valuables included in GCF?				
Is the natural growth of cultivated forests included in output and GCF?				
Is financial intermediation services indirectly measured (FISIM) allocated to users?				
1993 SNA extends the production boundary of households to include goods that are not made from primary goods? Are these goods included In output?				
1993 SNA extends the production boundary of households to include all goods that are produced by households whether for sale or not ? Are These goods included in output?				
Are volumes estimated using a chaining procedure at least on an annual Basis?				
Are un-funded social contributions (for sickness, unemployment, retirement etc) by enterprises imputed as compensation of employees and included as contributions to social insurance?				
Do non-life insurance estimates include premium supplements rather than being based just on premiums less claims?				
Do life insurance estimates include premium supplements rather than being based just on premiums less claims?				
Are reinvested earnings estimates included In the rest of the world Account?				
Are foreign workers? remittances excluded from GNI?				
Is output and value added measured at basic prices?				
Is output and value added measured at producers? prices?				
Is value added measured at factor cost?				
Is government final consumption expenditure broken down into individual and collective consumption?				

4.3. Weak Points and Bottlenecks in the Workflow

Currently the UN NAS provides Excel tables, but submissions by coutries in other formats, electronic as well as paper, are accepted, too. Not all countries have real time and on-line Internet connections, creating obstacles for the communication with national offices, especially for the dispatch of the Excel NAQs, and follow ups. Tough nearly all Excel NAQs are received via email and in the desired format, some countries provide their data in their own excel spreadsheets or formats like pdf. This situation causes three negative main effects: *(1)* diminished data quality due to media breaks, *(2)* much time spent on double data entries,

and *(3)* high cost of data manipulations.
Some of the problems are that:

- Multiple changes of media cause high costs and ample opportunity for errors.
- The validation process is tedious and time consuming, requiring 814 rules to be validated. Typical problems of the validation include avoidable errors like switched numbers, balance equations contradicting data and missing data.
- Ex-post validation of country data at the UN NAS, New York, causes a high need for communication in order to correct invalid data and is politically risky, since it is not necessarily the country submitting the data finally published.

5. Requirement Engineering for the eNAQ System

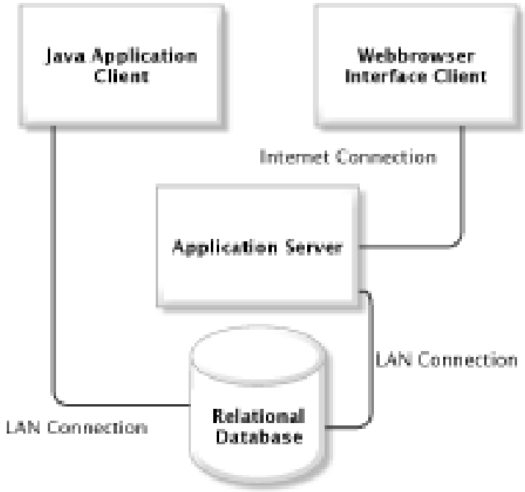
5.1. Objectives

The main goal for an eNAQ is to improve the workflow at the United Nations Statistics Division. This is to be achieved by a framework that supports direct data entry into the national electronic questionnaire. The system should assist the national offices by providing help functions, and deliver feedback about the validation of the entered data, thus strongly raising data quality. Of course, such a system must support multiple languages. Furthermore, it should support communication between the national offices and the United Nations Statistics Division (UNSD) about the data. When data is accepted by the UNSD, automatic data storage within the database is required. Stored and transmitted data has to be protected against unauthorized modifications, cf. Müller, Eichler and Lenz (2004)). For countries with no reliable Internet connection and for importing delivered data, import and export functions are required. Additionally, the generation of data for reports for scientific or public purposes and the production of a statistical yearbook are to be considered.
eNAQ's main function is to offer an electronic version of the *System of National Accounts Questionnaire* as a framework for the whole process of entering data sets, checking data sets for completeness, checking data with respect to the validation rule set, and generating reports. Note, that eNAQ is explicitly not meant to compute (new) variables in order to generate derived data.

eNAQ may be used by the national statistical offices of all UNO membership countries and UNSD. Their *user groups* include: UN staff members, country specialists, and system administrators. Data access for other users (scientific purposes etc.) requires an additional user group with different access rights, to be managed by the UNSD staff.

eNAQ has to support two different user modes. It should be usable as a local Java application client within a local area network connecting to a local database via JDBC[9] by a TCP/IP[10] connection. The second scenario

Illustration 3. Three tier architecture of eNAQ



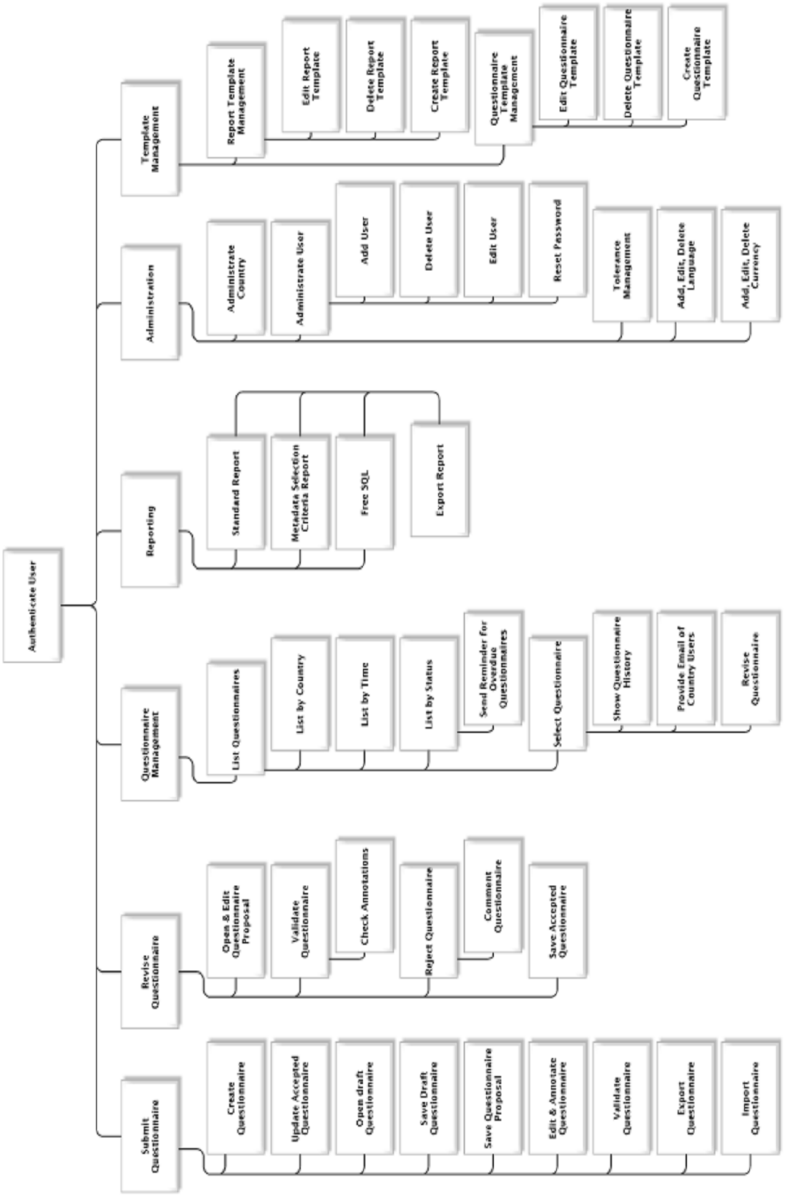
uses an Internet connection. The client connects to the server using SOAP[11], while on the server site eNAQ is run as a web service, locally accessing the database via JDBC.This results in a three-tier client-server architecture (displayed in illustration 3) where the presentation layer (top tier) provides the user interface. The interface can be implemented using either a browser interface or an application program written for example in Java, cf. Eichler (2004). The middle tier is an application server which provides security checks, identity verification and may implement part of the business logic[12]. The bottom tier is a database server running a relational Database Management System (DBMS). During the implementation specialregard should be paid to the software that will be necessary to use eNAQ. Countries using the client must not be forced to buy expensive software licenses from international vendors. In order to allow countries to set up their own database system and application servers, embedded commercial licenses must be minimized. The user interface should be similar to the paper version of NAQ, so experienced users feel familiar with the all tables and intuitively know where to enter data.

[9] Java Database Connectivity.
[10] Transmission Control Protocol - Internet Protocol.
[11] Simple Object Access Protocol
[12] Business logic may also be implemented by the client or the database.

5.2. Functions

The function tree (displayed in illustration 4) provides an overview of the required functions that can be deduced from the objectives specified above.

Illustration 4. Function tree



5.3. Validation Rules and Meta Data

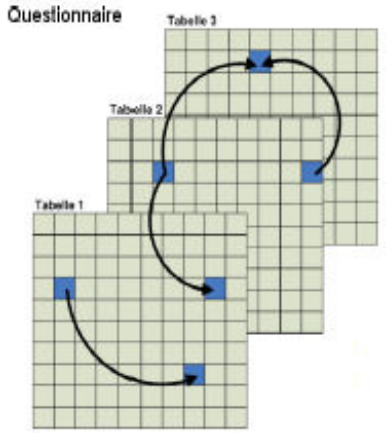
Validation rules are of a great importance for any system. How validation rules look like and how do they work? Remember that data tables contain data items. Items are a synonym for variables or economic indicators which are the basics of SNA. Hence, relations among items are defined as rules, according to economic accounting principles. Some rules define relations between items within only one table; other rules define relations between items of various tables. Table 3 illustrates a rule defining the relations within table 1.1 GDP *by expenditure in current prices*. It contains all the information a user should recognize when he validates (aggregated) data of a NA questionnaire. The table is to be read as an equation, where the elements on the right hand side must equate the element on the left hand side (the "caller"of the rule). SNA codes identify economic variables and the item codes uniquely identify any item within the NAQ. A description provides interpretation of a given variable. Special regard should be paid to the entry *Statistical Discrepancy*, which designates how much the sum of values of the right hand side may deviate from the left hand side.

Table 3. SNA validation rule of eNAQ

Rule Group				**101**	**Accounting Rule**			
Rule #	SNA Code	Item Code	Name	,	+/-	SNA Code	Item Code	Name
1010001	B.1*g	1012400	GDP by expenditure	=		P.3	1010700	Final consumption expenditure
					+	P.5	1011300	Gross capital formation
					+	P.6	1011700	Exports of goods and services
					-	P.7	1012000	Imports of goods and services
Applies to SNA				1993 ; 2008	+		1012300	Statistical discrepancy

Now remind yourself that the various tables are stored as spreadsheets with relations among their items defined in rules. One can imagine a complex rule set based on arithmetic relationships, cf. illustration 5. Keep in mind the complexity of the data structure with 31 tables, 1046 items and 814 rules, for the 1993 SNA, only.

Illustration 5. Multi relationship between items and tables

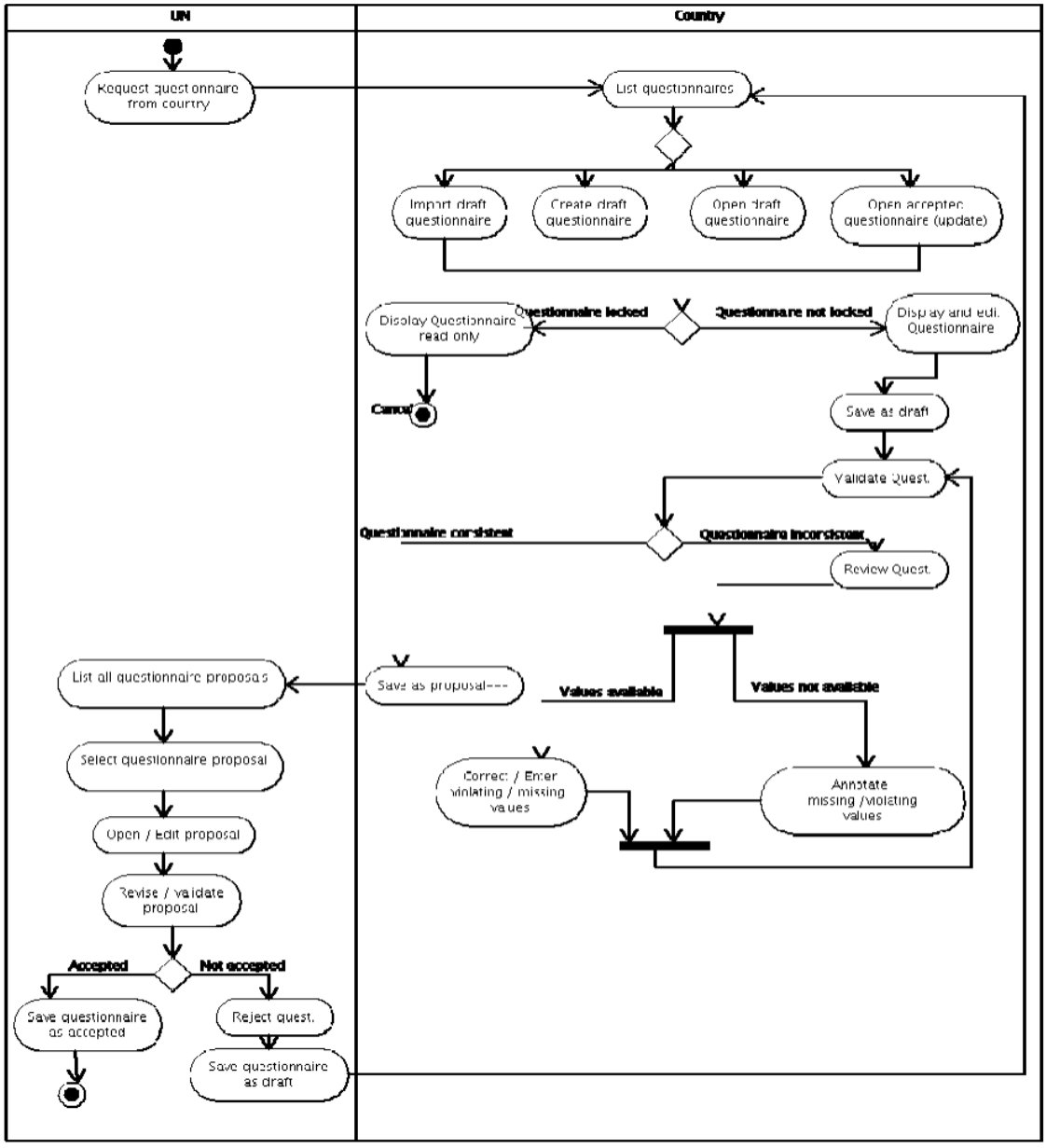


Note, that not only variables have to be provided by the system, but also meta data about the variables themselves. Such meta data deliver information for the user like "To which table does a certain item belongs?", "What is its semantic?", and "What is its relation to all ancestral variables?". Moreover, remember that this information has to be provided for multiple languages. Clearly, the complexity of data structure stresses the need for a sound database design. But before going a step ahead, let us take a short look at the streamlined workflow of eNAQ.

5.4. Workflow required

The diagram below represents the required workflow. It encapsulates the same functionality as the two diagrams displayed above. Note that the major improvement lies in shifting the validation process to the countries and forcing country users to validate data before submitting it to the UNSD. Comparing only the length of the graph, it becomes clear that the new workflow should be more efficient.

Illustration 6. Data compilation and validation of eNAQ



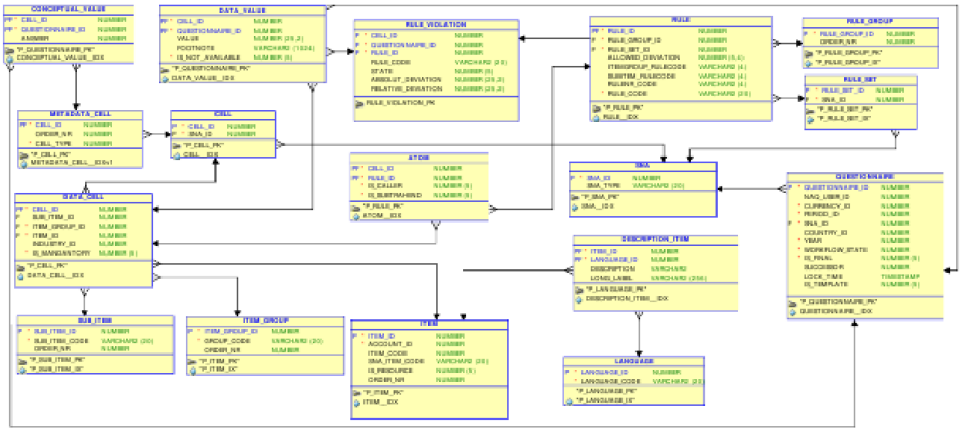
6. System Design

6.1. Conceptual Database Schemata

In this section some highlights of the conceptual database design are provided. It is based upon the relational model. Note, that this presentation is far from comprehensive, concentrating on the elements essential to the display of questionnaires, the storage of data entered within questionnaires and the validation process. Within the relation QUESTIONNAIRE information regarding a specific questionnaire is stored, such as the country, currency, year, and the SNA version. The relation is linked with the relation SNA, which provides connection to a specific rule set for any version of the SNA as well as a connection to the relation CELL.
Cells are used to store all elements of a questionnaire. Using the relations DATA_CELL and METADATA_CELL provides a necessary distinction between conceptual data and economic variables. The relation DATA_CELL is used to construct tables containing the economic variables. Tables are represented in the relation ITEM_GROUP, table items are stored in the relation ITEM, and the sub items of table 2.3 are stored in the relation SUBITEM. Note, that the relation ITEM has a link to the relation DESCRIPTION_ITEM. This relation contains meta data for every single item, such as a label or a description which can be used, for example, by the help system to provide context information regarding single variables. Using a composite primary key consisting of the ITEM_ID and the LANGUAGE_ID this provides the necessary basis for storing meta data in multiple languages[13].

[13] Note that this construct is used in other relations as well, but for reasons of clearness they are removed.

llustration 7. Questionnaire and Validation Rules (simplified) of eNAQ



A composite primary key is used in the relations DATA_VALUE and CONCEPTUAL VALUE, too. It is composed of the CELL_ID and the QUESTIONNAIRE_ID. In these relations specific values entered by users for table items, and respectively the questions in the conceptual table as well as the country notes, sources, special remarks and explanations regarding the chosen method of GDP computation, are assigned to a specific questionnaire.

Concerning the validation process, the relation ATOM is used to assign further attributes to each combination of a data cell and a rule applying to this cell. This relation specifies whether the value of a specific cell is either positive or negative in the context of a given rule, and also whether it is found on the left hand side, or the right hand side of the rule equation, as illustrated in table 3. The relation RULE in turn, belongs to a rule set, which is assigned to a specific version of the SNA, as can be seen in the relation RULE_SET.

Furthermore, on each update of the relation DATA_VALUE a combination of stored procedures and triggers uses this information to validate the new values against the stored rule information. Rules may be satisfied, violated but with a feasible statistical discrepancy, violated, with an infeasible statistical discrepancy, or simply refer to missing values. Depending upon the single case, information is stored within the relation RULE VIOLATION. Further triggers ensure that questionnaires can't be sent to the UN NAS for revision until at least all rules are satisfied inside the statistical discrepancy or violations have been annotated. This greatly facilitates client development as it implements essential parts
of the business logic inside the database, thus decoupling business and application logic, allowing for the development of light clients that only display views of the database.

The database stores much more information, regarding the various countries, currencies, the different users and user groups within the system as well as templates, but their presentation is outside the scope of this paper. The same is true for *external views* and the *internal schema*.

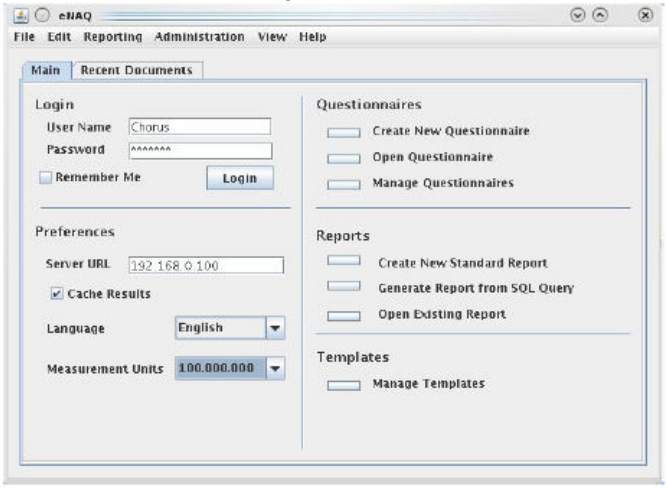
6.2. Functionalities embedded into the GUI

Finally, we have a look at the GUI. Its design includes calls of functions for editing and revising questionnaires, a help function, functions for questionnaire management, user, country and currency administration, functions for the generation of statistical reports as well as functions for using questionnaire and report templates. Due to the space limitations we will concentrate on presenting the main windows of the eNAQ client, as well as the functions for editing questionnaires and questionnaire management, as they are considered core functionalities. For better oversight of all available functions, the reader is referred to the function tree as displayed in illustration 4.

6.3. The Main Window

The main window, see illustration 8, provides an overview of the core functions to be invoked by eNAQ's GUI. It presents a login field, the required information for connecting to the server, as well as the selected language and the chosen measurement units. Furthermore, it provides an overview of core functions available, such as creating, opening and managing questionnaires as well as reports and templates.

Illustration 8. eNAQ Main Window, Start Screen

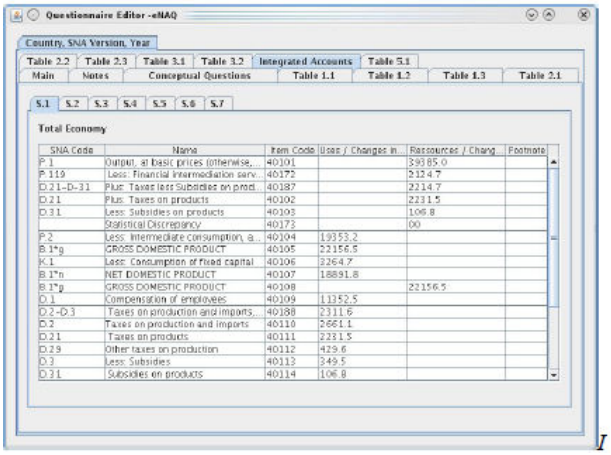


All these functions, as well as additional functions are also available through the menus below the title bar. The *Recent Documents* tab provides easy access to the last opened questionnaires and reports.

6.4. The Questionnaire Editor

Selecting the country and year, activates the questionnaire editor (illustration 9). In the top region of the window, the available tables are displayed, including the country notes, and the conceptual table.

llustration 9. Questionnaire Editor - Table 4.1



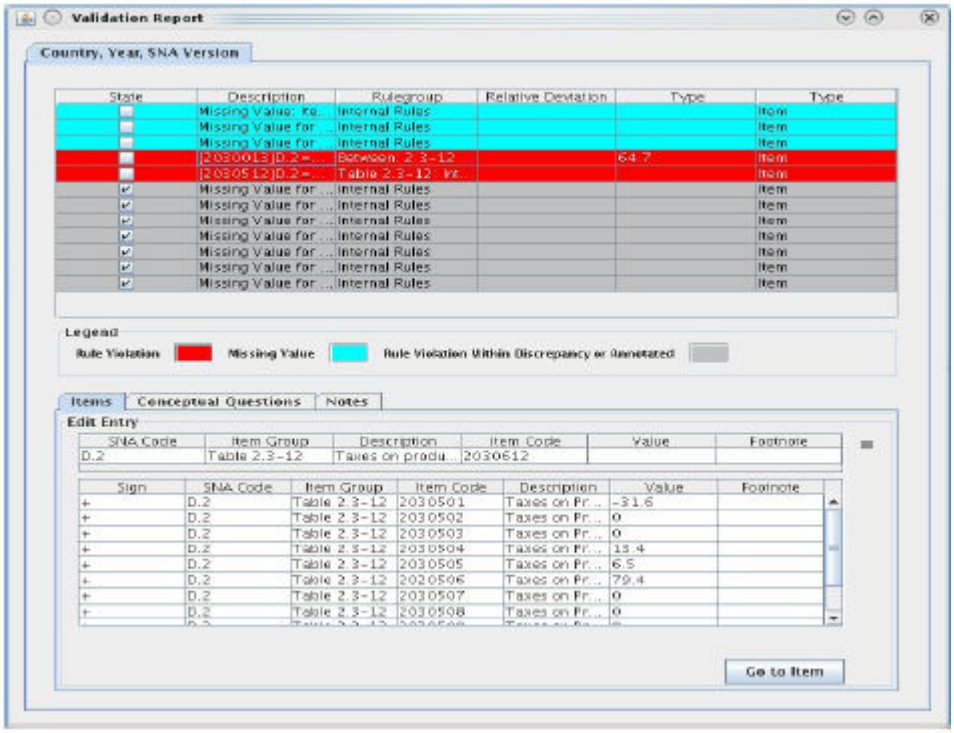
The main tab contains general information regarding the opened questionnaire, such as a questionnaire history, and the questionnaire status. Directly below more tabs are used to display a selection of sub tables for the selected category. For example, the integrated accounts tables contain seven tables, one for each sector. The table itself contains the SNA codes, a short description of the corresponding items, the item codes and a column for values to be entered by the user. An additional column is available for entering footnotes to annotate discrepancies or missing values.

6.5. The Validation Report

According to the requirement engineering any country expert is expected to validate his questionnaire before he can send it to the UN NAS. The validation function is invoked by a menu click. The validation form is displayed in illustration 10. The top table provides an overview of all violated rules. The coloring indicates different degrees of violations. Red is assigned to rule violations beyond the statistical discrepancy allowed. Cyan indicates fields with missing data. Light Grey is used for

rules that are violated, but within the statistical discrepancy or violations that have been annotated with footnotes. The checkboxes within the state column are automatically checked when a rule violation is resolved to the degree that a questionnaire can be saved as a proposal. Violations that are completely resolved are removed from the table. The description provides a short explanation regarding the cause of the violation. If the cause is a missing value, the item code is provided so that the user can identify the missing item within the questionnaire. If rules are breached the rule formula is provided in SNA Codes. The table also provides information about the relative and the absolute deviation, as well as the type of the violating entry. Types are items, conceptual questions and notes. The table can be sorted by any column.

llustration 10. Validation Report, based on Eichler (2004)



The validation report provides another table within the panel *Data Entry*. It contains tabs for the different types of violations (violations regarding relations between the various items, missing entries in the conceptual table or missing country notes). The tabs are automatically selected and filled with the corresponding data according to the type of the row selected in the top table. For items, on the top of the item tab the

item causing the violation[14] and its properties are displayed, including a field for its value and a footnote, in case a discrepancy cannot be resolved. If the violation concerns an unanswered question within the conceptual table, the value field is displayed as a combo box, containing the allowed answers. The equality operator indicates the rule to the user. The bottom table displays the SNA codes, item codes, the item group, a description containing the item names and their values. Thus, all elements designated by the violated rule are displayed with their properties, allowing for their identification within the NAQ and their interpretation regarding economic significance within the SNA. Users can directly edit or annotate these values, either by correcting or annotating the item calling the rule, or by correcting or annotating the rule elements. For missing values the bottom table remains empty and only the missing item and its properties are displayed.

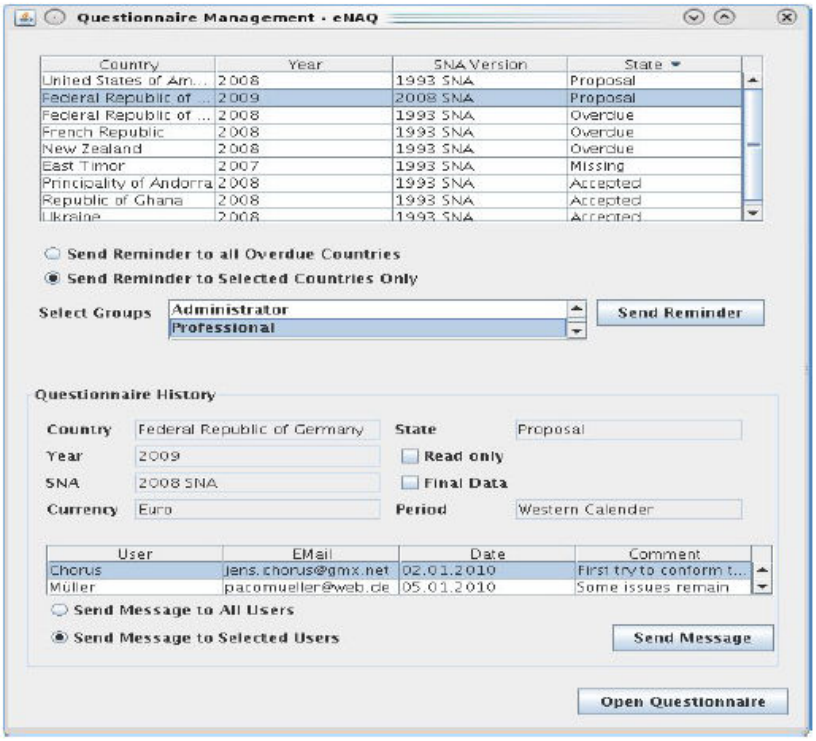
Note, that entries within the questionnaire editor are colored according to the violations detected within the validation report. This provides additional oversight, and allows users to use the questionnaire editor to correct values should they choose to do so. Entries for which violations have been detected, now also allow for editing footnotes.

6.6. Questionnaire Management

This function provides an overview of all questionnaires within the database, allows for contacting the

[14] In terms of the database design this item is the caller of the rule.

Illustration 11. Questionnaire Management, based on Eichler (2004)



countries of selected questionnaires, displays the questionnaire history of a selection, and provides an alternative path to access a questionnaire for further editing or revision. The top table of this form (illustration 11) can be used to sort all questionnaires within the database by country, year, SNA version and status. The value set (domain) for the status column is *Accepted, Proposal, Overdue and Missing*. The missing property was introduced to allow UN employees to flag countries for which no questionnaire is expected to be submitted given a fixed year[15]. The radio buttons *Send Reminder to all Overdue Countries* and *Send Reminder to Selected Countries only* allow users to choose the corresponding action for the *Send Reminder* button. A message is sent to all users selected in the corresponding list field. Selecting multiple user groups is allowed. The table also allows for the selection of multiple rows.

[15] For example a severe breakdown of government structures due to war, natural catastrophes or other events typically cause data insufficiency or lack of participation.

The *Questionnaire History* provides detailed information for any single questionnaire selected in the top table. If none or multiple tables are selected this should be deactivated. Users can contact other users who edited the questionnaire, using the radio buttons and the corresponding *Send Message* button. They can also directly edit the selected questionnaire by using the *Open Questionnaire* button. This action will activate the questionnaire editor.

7. Conclusions

Let us summarize the major improvements that can be achieved by implementing eNAQ at the UN NAS and the various NSIs around the world. First of all, the workflows are simplified and streamlined. Shifting the validation process to the countries themselves which submit the data, an increase in data quality will arise in most cases while simultaneously decreasing data transfers and communications between the UN NAS and the various NSIs around the world.

Once a questionnaire has been filled in, no double manual data entry is required at any stage of the workflow. The support facilities for annotations of missing values or data which contradict rules together with the integrated communication support, the intuitive GUI and the context sensitive help function guarantee a high level of user friendliness for all parties.

All components needed for the implementation of the presented design can be implemented using all major operating systems and do not require the purchase of expensive licenses. Even a cost reduction is achieved in comparison to the software currently required, as the tables provided by the UN NAS require Microsoft Excel for usage. However, this is negligible relative to cost reduction for data production and management gained by an improved workflow and no further need of a manual data validation process.

The system design allows an easy integration of new rules or rule sets, as will be required when 2008 SNA will be rolled out. Furthermore, statistical discrepancies can not only be set for each desired indicator, but even for indicators in the context of a specific rule. Once implemented in the UN NAS database, a rule change could be propagated to all countries with only minor efforts.

eNAQ can easily be adapted to fit other requirements such as scientific, government, and press appliances. As all data is stored in a relational database, derived data can easily be generated using simple SQL aggregation functions like *min, max, sum* and *count,* and utilizing meta information from the database.

The solid start-of-the-art design and the chosen technologies provide a good basis for future developments. Relational database management systems are available from many vendors and the GUI can be implemented, choosing from a

wide range of options. This ensures that no vendor lock in can occur, and thus guarantees future independency.

eNAQ promises to open the door for a cost reduced computing of NA questionnaires at an improved data quality level. The authors hope that this approach will close the gap between an advanced NA methodology due to the 2008 SNA and modern software engineering based either on web services and browsers, or application programs backed up by a user friendly GUI.

References

Chorus, J. 2010: Reengineering of eNAQ - electronic National Accounts Questionnaire - 1993 & 2008 SNA - Requirements and Design Specification. Msc Thesis, Freie Universität Berlin, Berlin.

Eichler, J. 2004: Entwurf und Implementierung eines Systems zur Validierung von Daten des 'System of National Accounts' der Vereinten Nationen. MSc thesis, Freie Universität Berlin, Berlin.

Lenz, H. J. - Eichler, J. - Müller, P. 2008: A Prototype for an electronic version of an UN National Accounts Questionnaire - or A Pledge for using Advanced IT and Applied Statistics. Revista Romana de Statistica, 11, 7–16. p.

Müller, P. 2003: Analysis and Requirement Engineering of the UN National Accounts. Diplomarbeit, Freie Universität Berlin, Berlin.

Müller, P. - Eichler, J. - Lenz, H. J. 2004: eNAQ: A Prototype for an electronic Version of the UN National Accounts Questionnaire, Technical Report, Freie Universität Berlin, Berlin.

Pubns, United Nations 2009: System of national accounts 1993.

URL: http://unstats.un.org/unsd/sna1993/toctop.asp

United Nations et al. 1993: Handbook of National Accounting - Links between Business and National Accounting. United Nations.

United Nations et al. (2008): System of National Accounts 2008 -Pre-edited version of Volume 1. United Nations Pubns.