

**eNAQ: A Prototype
for an electronic Version
of the UN National Accounts Questionnaire**

Patrick Müller, Jörn Eichler, Hans-J. Lenz
Horbach GmbH, Berlin, inubit AG, Berlin, Freie Universität, Berlin

Executive Summary

The project is the result of a two-step, two-parties system development effort. Requirement engineering was mostly done by the first author at the UN Statistics Division, New York, while design and implementation was performed at the Free University, Berlin, Dept. of Production, Information Systems and Operations Research. The overall objective is the development of an electronic National Accounts Questionnaire¹ (eNAQ) for the United Nations, Statistics Division. This paper focuses on the analysis of the present NAQ system and the requirement engineering for the new eNAQ. In a satellite paper further details about the system design and implementation of eNAQ will be presented.

The platform used for the prototype eNAQ is an IBM pseries 620, on top of which a DB/2 database engine under AIX is used. Note, that this paper only summarises the main features of the current prototype of an electronic National Accounts Questionnaire to be used by the United Nations. The authors hope to go on with stepwise improvements, and are sure to find the UN support in maturing eNAQ in such a way that its quality, performance and acceptance by the UN staff and the UN member countries will be far beyond any expectations.

This paper consists of four main parts. Chapter 2 describes the current system at the UN, and its challenges and limitations. Chapter 3 presents the gross design and a flavour of implementation of the prototype. The last part (Chapter 4) presents the major benefits of the new system.

¹ C.f. United Nations, UN National Accounts Questionnaire 1993 SNA.

1. Introduction

While working with a paper-based questionnaire (NAQ) at the UN, the members of the UN staff have to perform manually the complete and tedious processing (workflow) of each NAQ including inputting data, imputation and its validation. The workflow stops when the final figures are disseminated to the various user groups. The data provided is generally checked for completeness and validated. Inconsistencies in national data sets, which happen quite often, have to be detected and corrected. Footnotes, describing inevitable inconsistencies, have to be annotated by the UN staff since the UN member countries seldom provide footnotes themselves. Especially, the validation process is extremely time and resources consuming.

The purpose of eNAQ is to allow the UN to accelerate and to simplify their work processes in an integrated, multi-functional framework, and to produce a much higher data quality than achieved before. Data entry as well as data validation can be outsourced to as many countries as possible. To cope with such demands the electronic validation tool eNAQ was developed. The new system incorporates a precise and understandable set of rules and regulations of SNA methodology and expert's know-how of UN members of the staff.

Research on the System of National Accounts (SNA 1993)¹ was conducted in collaboration with the International Monetary Fund (IMF), the World Bank (WB) and the United Nations (UN). According to the needs and ideas of the UN staff and approved suggestions of the first author, certain additions and simplifications to the current system were settled. These activities led to a new national accounts questionnaire (NAQ_new), which is not fully compliant with the old one. Therefore NAQ_new cannot be simply rolled out in a one-to-one manner. The new, the old or both NAQs will need to be adapted and customised prior to the start of eNAQ.

¹ C.f. Commission of the European Communities, International Monetary Fund, Organisation for Economic Cooperation and Development, United Nations, World Bank: System of National Accounts 1993, 1993.

2. System Analysis of the UN Statistics Division System

In the following chapter a short description of the current situation, its challenges and limitations at the UN Statistics Division, respectively, the Official Data Sub-Section of the National Accounts Section will be presented.

2.1 Mission of the UN Statistics Division

The task of the Official Data Sub-Section - in the following the shortcut UN is used - of the UN Statistic Division, New York, is the compilation of the National Accounts data of all UN member countries, whenever those data become available at the UN headquarters. The data form the basis of the Annual Yearbook „National Account Statistics: Main Aggregates and Detailed Tables¹“ Further tasks of the Statistics Division are to support the member countries compiling the national accounts data according to the 1993 SNA standard, and to make National Accounts data accessible and available for authorities, scientists and the public.

2.2. Workflow

In order to get insight into the limitations of the eNAQ, the current workflow at the UN needs to be analysed. The questionnaire form is annually sent to the respective countries by regular mail. It is already pre-filled with the available data of the last years NAQ as a kind of default values. The countries are asked to complete, respectively, review and update the data of the current period as well as the data of the previous years, which are not included in the questionnaire. The countries are expected to send back the filled-in forms to the UN in time. There the data are entered from the questionnaires into the UN - NAQ database, cf. Fig. 1. Note, that data are usual manually validated, because the existing validation tool unfortunately provides results of limited use only. In case of physical incompleteness of data of a questionnaire or of the questionnaire itself, inconsistencies of main aggregates or missing data the related country must be addressed once again. Evidently, this heavily slows down the NA workflow, and may cause further errors of data capture. The Yearbook is produced from the released official data. Countries, which don't provide data, even if they

¹ United Nations, National Account Statistics: Main Aggregates and Detailed Tables.

are useful for publication, may be excluded from the Yearbook. A statistical office assistant checks it, and a professional clerk edits the final review.

Data for scientists and other institutions are provided upon request. Therefore a whole NAQ table, which may be regarded as a superset of the really needed data, is selected from the database. It is worthwhile noting that no more dedicated selection is possible. The resulting Access database has to be edited in further steps by SQL follow-up queries. Note, that a user-friendly GUI [Graphical User Interfaces] don't exist, which meet the needs of the scientists for data extraction, transformation and down-loading, i.e. erasing all unwanted data, possibly creation of Excel spreadsheets if requested etc.

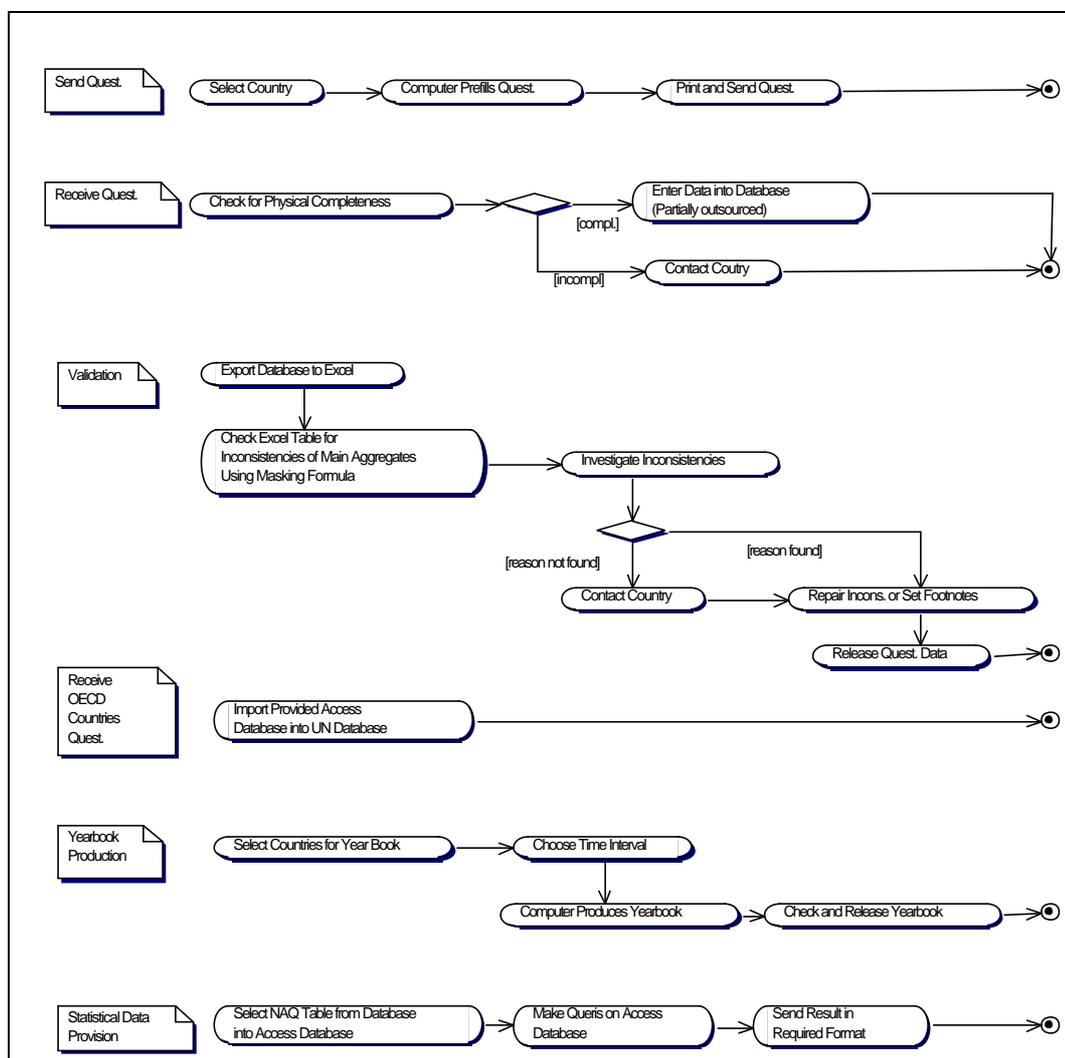


Fig. 1: Use Cases for various UN NAQ activities

2.3 Challenges and Limitations

The entries of the tables of the NAQ have to allow for statistical discrepancies and an specific input of each cell on the fly. Thus it is impossible to pre-compute any value of the questionnaire. Ex-post checking of data validity is therefore mandatory.

There are several weak points of the present NAQ system. The first pitfall is that media changes frequently happen. The countries have to fill in a paper questionnaire or, alternatively, they can type in the data in a blank NAQ Excel spreadsheet. When the UN NA staff receives a questionnaire, it has to be entered into the UN database. When the data is entered, the national accounts have to be manually validated, because the existing validation tool produces only an unreadable report. A recently developed tool provides a more convenient validation report. For instance, it shows at least the name of the transaction / flow corresponding to the UN internal code as well as deviations in absolute or percentage values.

Another major challenge of the existing system is the time required at the UN for data entry and validation. Entering the data into the database takes 2-3 hours per questionnaire and 30-45 minutes per questionnaire for validation. Due to limited expert's resources at the UN, the data entry process is more or less outsourced.

As experience shows, mistakes regularly occur during the completion of a questionnaire. Major sources of errors are:

- misunderstandings and wrong interpretations of a questionnaire,
- twisted numbers, or
- erroneous figures provided by single countries; this is caused by misuse of Excel-functions when the NAQ Excel-spreadsheet is filled in.

Often enough, the UN staffs discover mistakes or inconsistencies, which can only be corrected by the responsible countries. Footnotes, which may explain inconsistencies, are seldom provided by the countries and have to be added by UN NA experts.

Time-series analyses or specially tailored reports of data cannot be produced with the tool currently in use. Database queries only retrieve complete NAQ tables. The needed information has to be tediously extracted from an Access database in several steps. A graphical user interface is not available for this workflow of UN NA Group.

Our system analysis can be summarised as follows: The prerequisites of the use of the NAQ form are extensive with respect to the required intellectual level of understanding SNA processing by authorised people. The processing time

required for editing, completion and checking of a questionnaire form is unnecessarily large. Thus, the development of an electronic questionnaire is mandatory. The objectives of such an IT strategy and planning of an eNAQ are clear: Improved input, throughput and output, more user friendliness at the site of data suppliers and UN staff, and less time and money for data capturing, completion and checking. Reasons for inconsistencies within the national account tables may result from e.g. shortcuts which are used for the congregation of values e.g. using output/intermediate consumption ratios to drive the gross values added for the industries.

Some countries do not compile all requested data of the NAQ or they are unable to do so. Missing data may be single entries up to whole industries, whole accounts or sectors of the Sequence of Accounts or even whole tables.

3. Gross Design and Implementation Aspects

Based on the system analysis described above, a new electronic NAQ called "eNAQ" was developed. Its main goal is to overcome the described pitfalls and limitations. eNAQ is in a kind of β -release, and it is not fully tested. Final testing and improvements may be considered as a prerequisite for eNAQ to successfully address the issues mentioned above. However, it is a matter of fact that the new system will eliminate or fix errors caused by manual data entry. The new system also leads to an improved workflow, which will save time while simultaneously providing higher quality of requested data and reports. The most important feature of eNAQ is a real innovation: The new system will allow for direct data entry by the collaborating National Statistical Offices around the world. Last but not least, the user friendly, multi-functional and self-explanatory (by help functions) user-interface will strongly assist the National Office staff to edit data at their sites. Thus, the UN staff is unburdened, and outsourcing as before will be no longer of any concern.

The validity check is very important since it guarantees that data which are sent by the countries are either consistent, or that footnotes describing the reasons for discrepancies are provided. Validity checks result in a list, showing all violated validation rules and their corresponding footnotes. This schema enables the experts to correct violations, and footnotes can be double checked for their pertinence. To implement the validity check, a confirmed body of rules and regulations for the SNA as well as a body of rules for the deviations is compiled¹. As an example of the 715 rules in the eNAQ, the rule for Gross Value Added is chosen. It applies to all domestic sectors of the economy thus one

¹ C.f. Müller, Patrick, Annex to Analysis and Requirement Engineering of the UN National Accounts, 2003, Chapter IV.

finds “Domestic” in the sector column representing Non-Financial Corporations, Financial Corporations, Government & NPISH and Households. The table columns have to be read as equations. The left hand side is one line, specifying a single entry in the NAQ

Since all entries of the NAQ have to be entered the rules doesn’t compute any values, but check those entries for consistency. It is checked, whether the left hand side of the equation equals the right hand side, including inputted statistical discrepancies in the right hand side computation

On the right hand side, are several lines specifying several fields that have to be summed up as described in the +/- column. On the right hand side, sector and account can be omitted if they correspond to the left hand side.

							Accounting Rule				
Rule#	SNA Code	Name	Sector	Account	U/R	= +/-	SNA Code	Name	Sector	Account	U/R
1	B.1g	Gross value added in basic prices	Domestic	I	U	=	P.1	Output in basic prices			R
						-	P.2	Intermediate consumption in purchaser's prices			U
						+		Statistical discrepancy			R

The account number is used to define an account in the SNA and U/R stands for use and, respectively, resource. The rule could be read as:

Gross value added (use) (in account I, applying to all sectors) = output (resource) less intermediate consumption (use) plus statistical discrepancy (resource).

The system will allow the UN staff to double-check the national accounts data and their related footnotes and meta data as data about real data, which is provided by the countries. Based on this information the UN may decide to accept, reject or edit incoming National Accounts Questionnaires.

Manual data entry and validation at the UN will generally not be necessary anymore. The UN staff only has to revise the delivered NA questionnaires, especially, the footnotes, but the whole process will be much simpler and faster than before. Once the data of the questionnaire is validated and accepted, the system will allow exporting data by standardised interfaces. In order to achieve high reporting flexibility, the “3D-search principle”¹ is supported. That means that the needed data which are demanded may be compiled through the GUI by just clicking on icons, which represent the three fundamental dimensions: subjects (topics of national accounts data or meta data), space (country), and time. Of course, a query can be assembled along any sequence of the three basic

¹ c.f. c

dimensions. This special feature enables so-called ad-hoc queries. Moreover, the UN can create templates for standard reports and set tolerances for the deviations in a simple and straightforward way.

System Administration is to be supported by the responsible UN system administrator on a global level, and by local administrators ("responsible experts") on a local level in the various member countries.

The following sections will give a short overview to the design and implementation of the prototype eNAQ. The first part introduces the primary criteria that were derived from the system analysis. The architecture and implementation is presented in the second part. The third part illustrates the graphical user interface (GUI).

3.1 Main Design Criteria

The following design criteria are essential for eNAQ

- Multi-functional and intuitive GUI to support editing and validation of the data
- Flexible logically (software) and physically (hardware) independent data structure to allow the extension of the application with regard to upcoming SNA tables or validation rules without the any need to change the program code
- Interfaces to external applications to allow collaboration and integration with existing infrastructure and legacy applications
- Client-side installation without manual interventions
- Modular and scalable application architecture to allow the extension of the application with respect to further functionality
- Implementation has to be independent of the underlying hardware and operating system
- Avoidance of license fees for libraries or middle ware

3.2 Architecture and Overview on Design and Implementation

The architecture of eNAQ follows the classical three-tier architecture. Therefore the functionality is unified with respect to presentation, business logic and data access. Well-known design patterns have been used to structure the classes

within the tiers¹. This promises a better stability and transparency of the system. The following figure shows the main components and their relations.

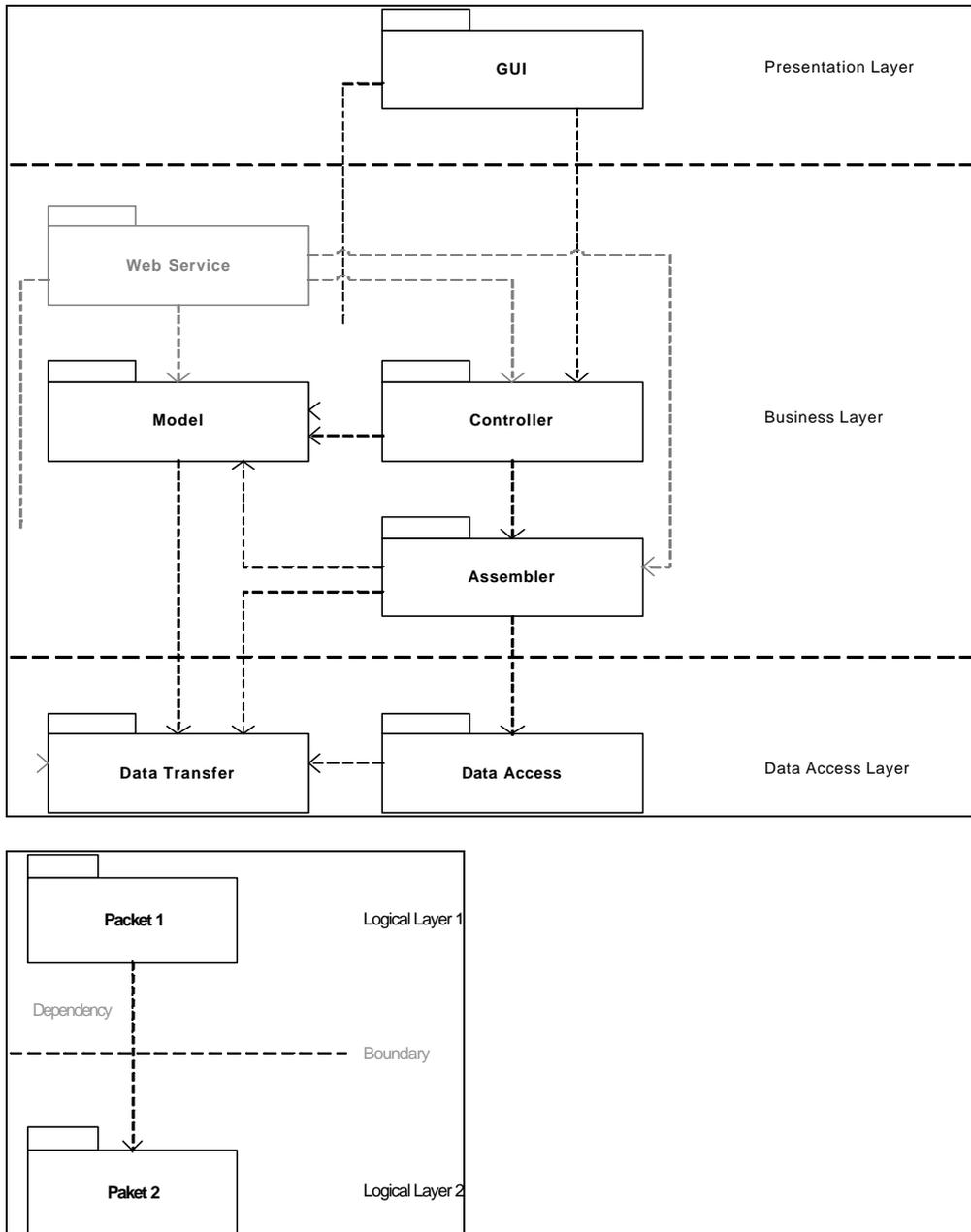


Fig. 2: Main Components and Layers of eNAQ

The implementation allows two installation strategies.

¹ Cf. Eichler, 2004, pp. 39ff, for details

The first variant allows that the application logic is distributed according to a rich client-/server-model: The major part of the application logic is executed on the client-side (rich client). Authentication, authorisation, multi-user access etc. is realised on the server.

The second variant focuses on a service oriented architecture (SOA). Therefore the major part of the application logic is executed on the server-side and offered as a web service. Only the graphical presentation is executed on the client side (thin client). A relational database is used, which needs a database management system as supplied by all main software manufacturers like Oracle, IBM, Microsoft, Sybase, Informix etc. Opposite to a HTML-based web application, the rich client approach allows a highly reactive user interface, which is robust against temporary breakdowns of the server connection. Nevertheless, the solution presented above easily allows developing a HTML-based web application using the implemented eNAQ solution as a back-end.

The communication between client and server can be realised as a web service as already mentioned. Therefore future client applications can use the functionality of the server application without communicating by a proprietary protocol. Furthermore it allows integration within automated business workflow. Because web services use common protocols (SOAP over TCP/IP), it is normally not necessary to adopt the existing infrastructure in order to run the application (e.g. a firewall set-up).

Multiple import and export formats are supported to allow a smooth transition from the actual infrastructure to the electronic questionnaire: a copy-and-paste approach may be used to import data from Microsoft Excel or Access files, XML-coded files may be read or written from the application to import or export data to other (statistical) applications.

eNAQ is implemented in pure Java. All libraries used within the application are available free of license fees. The access to the database is realised manufacturer independent (JDBC), the utilised SQL statements follow the SQL92 standard. Therefore there is no hardware or software manufacturer "lock-in". The client-side set-up can be realised using the Java "WebStart" technology, so there is normally no need for an intervention of the UN administration staff.

The schema of a questionnaire (tables, items etc.) and the validation rules are stored within the same relational database managing actual figures. It is possible to store several releases (or versions) of the same questionnaire (e.g. "SNA93") or different validation rule sets within the same database. Therefore extensions to the data structure of the NAQ or changes in the rule sets can be made without the need to change the application logic or to transform pre-existing data. The following figure presents a part of the conceptual data model concerning the structure of the questionnaires. Due to limits of space in this report, and the complexity of the data structure, reasoning on the details is skipped, cf. Eichler (2004).

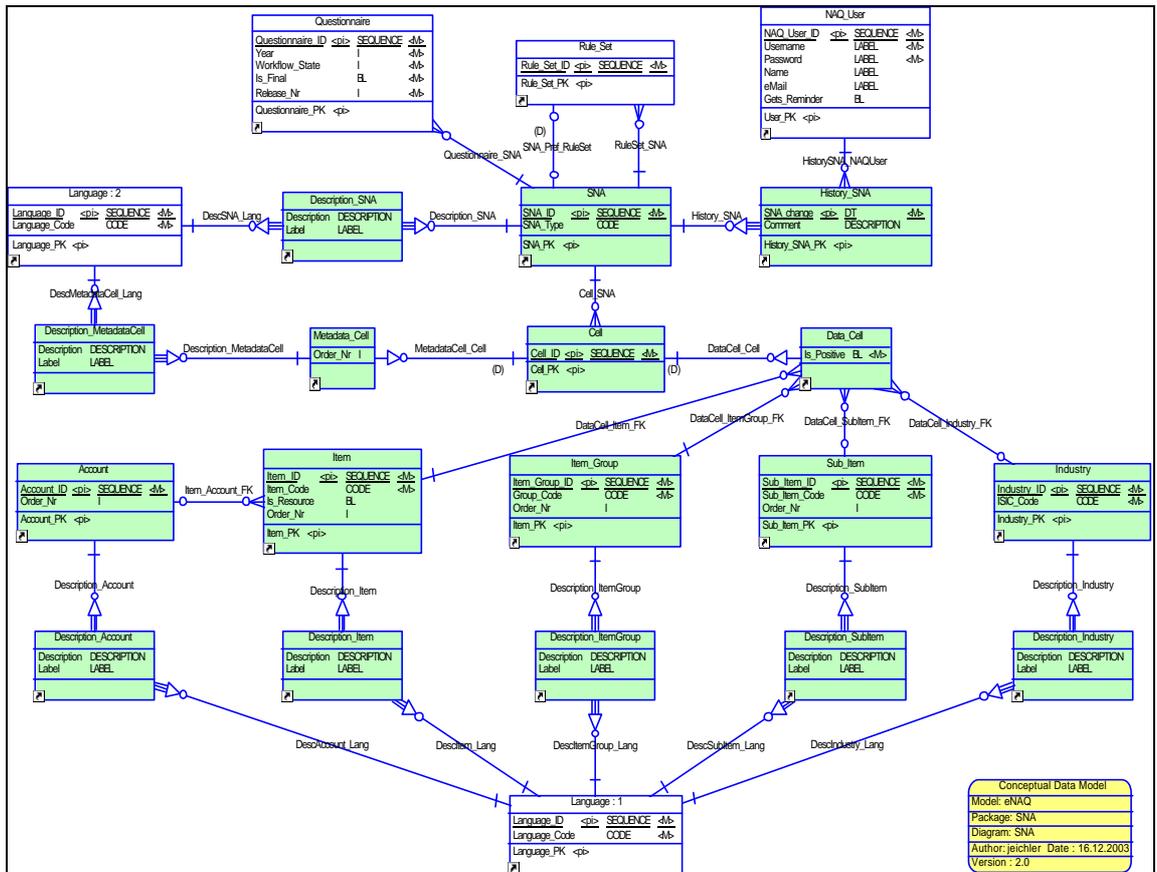


Fig. 3: Conceptual Data Sub-Model of eNAQ

The user interface allows for two main views: a powerful editor to enter and edit data values [1] and a validation report generator to show the results of the real-time validation [2].¹ The results of validation are also shown in the edit mode by highlighting the data values in different colours and their validation rule status [3]. Additionally, the user interface provides a property panel [4], which shows background information concerning the selected entry. The following figure represents a screen shot of the editor, the validation report and the property panel visible. The property panel shows some information about the selected validation rule.

¹ See the following figure for the numbers used.

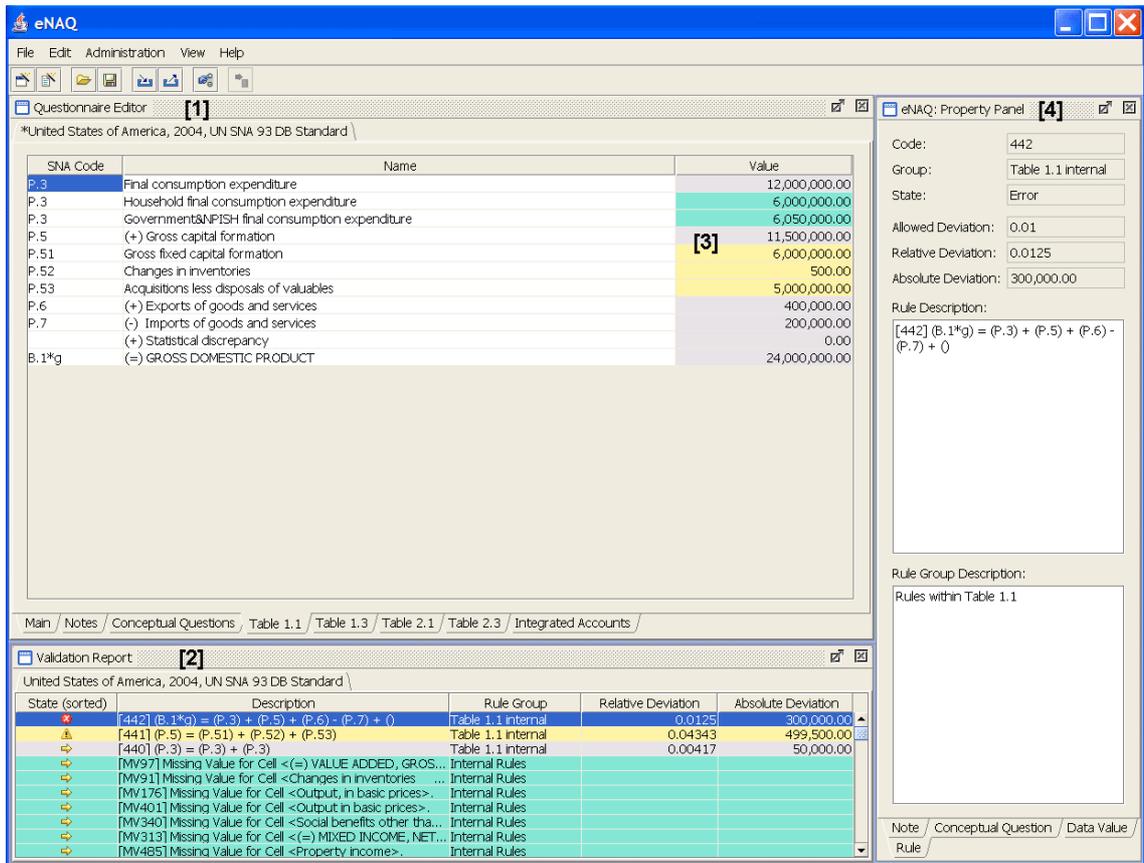


Fig.4: GUI of questionnaire editor of eNAQ

The navigation within the application makes use of the file metaphor. Therefore even untrained users in the member countries can very quickly learn to run the application. It is easy to see, that one has to choose “File/New ...” within the menu to generate a new questionnaire, to save it with “File/Save ...” and to edit an existing questionnaire with “File/Open ...”. Furthermore an online help and a user manual are available.

The revision of questionnaires simplifies a lot through a filterable validation view and its tight integration with the editor. As an example, a UN staff member can simply filter out all entries in the validation report that do not result in an error or a validation rule breach (see following figure). Afterwards the selection of an entry in the validation report shows details to the broken rule in the property panel and highlights the corresponding values in the editor.

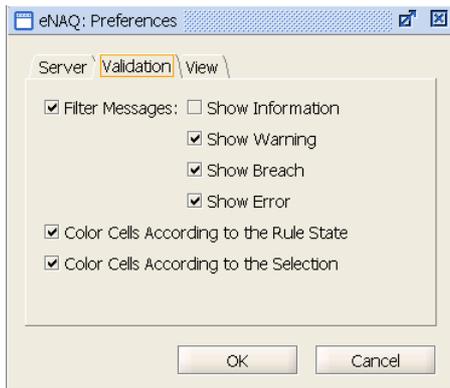


Fig. 5: Dialog Form “Preferences / Validation”

The interface is kept very simple and provides only access to data and functionality needed for the user group to which the current user belongs. For example, not authorised users cannot select the menu item to create a report, questionnaires of other countries are not visible for the staff in the member countries. The following figure shows the main navigation paths of eNAQ.

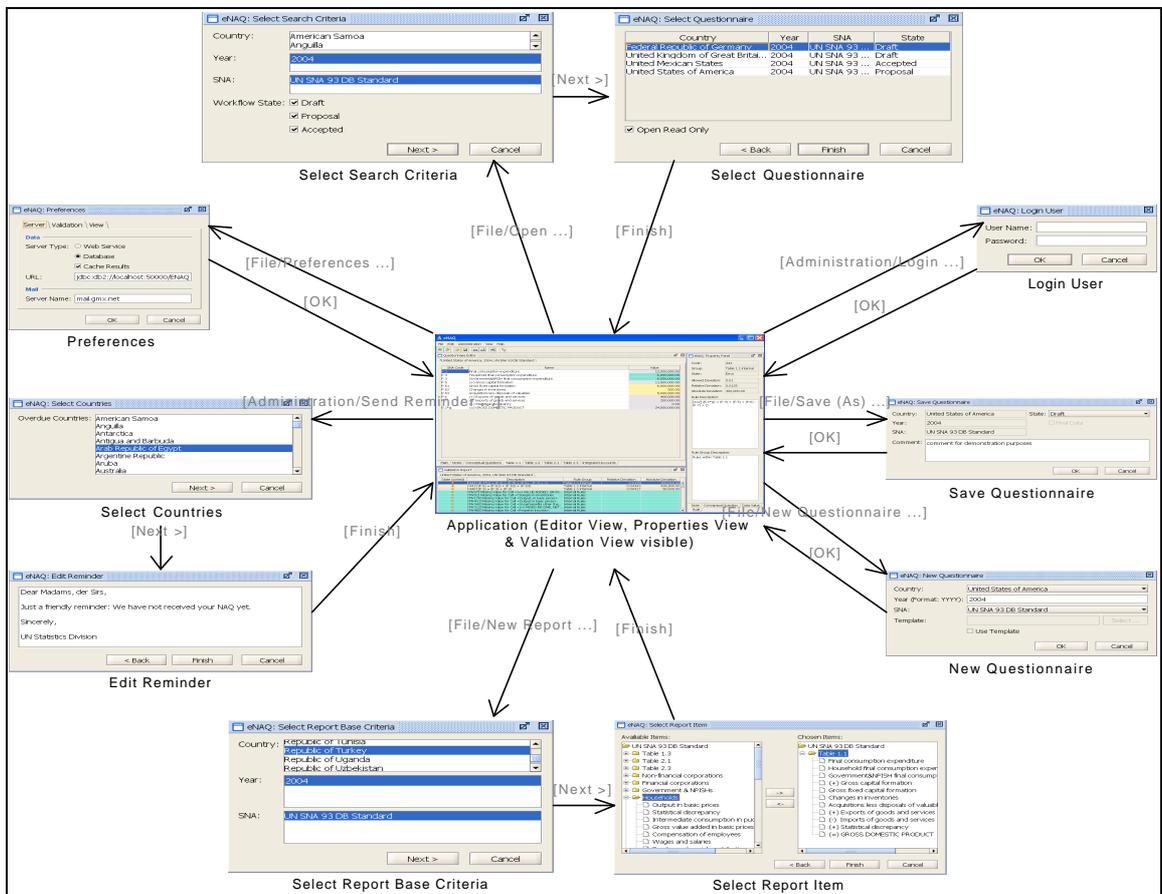


Fig. 6: Navigation in eNAQ

4. Major Benefits from the New System "eNAQ"

In this chapter the major benefits of the utility of an electronic version of the National Accounts Questionnaire – eNAQ – are presented. The benefits can be classified into three categories: higher data quality, savings of resources and increased user-friendliness of the system.

- staff is assured, that data, which passed the validation process, meet the requirements of the UN NA.
- data are either consistent and any deviations in between allowed limits, or a footnote is provided in the opposite case.
- eNAQ is time and money saving since there is no need to manually enter data
- needless to check again the data for validity on low (national) level – except for exceptional cases.
- need only to double-check the footnotes which are provided for inconsistent values and their respective values
- need to finally decide whether to accept / reject a supplied questionnaire.

The actual timesaving for the UN cannot be measured, because the data entry is partly outsourced from the UN to other service providers. However, the fee for outsourcing can be saved.

Further benefits are:

- simple handling of the system through an user-friendly user interface, especially, simplified data entry at National Offices with very limited resources.
- on-line definitions of SNA terms, which can be made available by the context sensitive help system
- users, who are familiar with SNA, but not with the former UN system, will be able to handle the new system or, at least, understand the figures without the background knowledge of the specific UN terminology with its confounding code numbers.
- report functions for NAQ data (up to a wanted level) for the UN and for researchers who need to deal with specific SNA data.

5. Perspectives

Obviously, the prototype eNAQ is to be customised to the current UN definitions, rules and requirements. Firstly, all missing tables, i.e. “Government final consumption expenditure by function” (Table 301) and “Individual consumption expenditure of households, NPISH and general government” (Table 302), the “Cross classification of gross value added by industries and institutional sectors” (table 501), and the missing NPISH sector, have to be included in the questionnaire. Secondly, the system should allow the entry of constant prices for the tables “gross domestic product by expenditure” (Table 102) and “value added by industries” (Table 202). Also, all new introduced, added and removed items in the electronic version of the NAQ have to be adjusted to the existing national accounts questionnaire or vice versa.

To provide a more sophisticated tool for a validation of National Accounts data, one can go far beyond simply checking rules and testing whether deviations are in between allowed limits or not. If there is high belief that that the data of a country is completely consistent, a help function can be provided by the system to input, transform or load data nearly automatically, e.g. the overtaking of already entered values into equivalent entry fields of given table schemas. Furthermore, the eNAQ system could be extended in such a way, that the validation process helps the user not only to find discrepancies but reports about reasoning about causality. Thus the system may help the national statistics offices to identify, analyse and solve discrepancies.

The eNAQ system could also be extended in order to serve as an analytical tool for registered scientist.

Researchers might be interested to merge NA data with other indicators like interest rates or the number of unemployed people as well as a total number of citizens per country. Such expansion can be handled with moderate efforts. By the way, an input field for the exchange rate to a basic currency is already included in the prototype. Note that the official data section does currently not compare figures based on a basic currency.

References

Eichler, Jörn (2004). Entwurf und Implementierung eines Systems zur Validierung von Daten des System of National Accounts der Vereinten Nationen, MA Thesis, Institute for Applied Computer Science, Free University of Berlin.

H.-J. Lenz :A rigorous Treatment of Microdata, Macrodata and Metadata, COMPSTAT Proceedings, W.Grossmann et al. (eds.), Physica, Wien, 1994

Müller, Patrick, (2003). Analysis and Requirement Engineering of the UN National Accounts, MA Thesis, Institute for Applied Computer Science and Institute of Statistics and Econometrics, Free University of Berlin.

Müller, Patrick, (2003). Annex to Analysis and Requirement Engineering of the UN National Accounts.

United Nations, Publisher (1993). System of National Accounts 1993. New York: United Nations. ST/ESA/STAT/SER.F/2/Rev.4

United Nations, Publisher (2000e). National Accounts Statistics: Main Aggregates and Detailed Tables, 1996 – 1997. New York: United Nations. ST/ESA/STAT/SER.X/25, Part 1

United Nations, Publisher (2001). UN National Accounts Questionnaire 1993 SNA, Part 1, Questionnaire Tables. New York: United Nations.

United Nations, Publisher (2002a). UN National Accounts Questionnaire 1993 SNA, Part 2, Supporting Booklet. New York: United Nations.