

UNITED NATIONS SECRETARIAT
Statistics Division

ESA/STAT/AC.91/18
7 November 2003

English only

Expert Group Meeting to
Review the United Nations Demographic Yearbook System
10-14 November 2003
New York

The eForm Revolution: Implications for the
United Nations *Demographic Yearbook*

By

Edwin St. Catherine
Central Statistical Office
St. Lucia

A. Introduction

1. Electronic forms have increased in significance over the past year. With the growing adoption of Extensible Markup Language (XML) as the common language for data exchange, the announcement by Microsoft of the impending release in October 2003 of Infopath as part of the Office 2003 System and the growth experienced by Adobe of its Server workflow structure for handling PDF and Client side interactive PDF (Portable Document Format) forms, the spotlight has been shining on eforms as the most cost effective method of data collection for public service and commercial reasons in recent times. This paper will describe the process of eforms data collection, the main advantages of its use over traditional methods and the major ways in which eforms data gathering solutions are being implemented.

2. The use of eforms as a mechanism for data collection involves the following broad processes:

- i) Forms conversion and design: This involves the generation of fillable PDFs or flexible infopath forms from existing static electronic to paper based forms format. A number of form designers and conversion tools have in recent times been developed by several software vendors, Microsoft, Adobe, Cardiff, Filenet, pureedge amongst others to generate the interactive PDF and to a lesser extent the flexible infopath form formats. These form designers include point and click design tools for adding basic validation logic and the possibility of building quite complex Javascript logic directly into the PDF.
- ii) System connectivity: Database lookups and automated field fillings can be designed using the new form designers. This is especially important for establishment surveys or where national, regional and international agencies collect information from satellite offices or firms in the case of an establishment survey. In the case of the *Demographic Yearbook* the requirement is that data on the immediate past year be inputted and data already contained in the database of the United Nations Statistics Division be used to pre-fill the form for verification by the respondent who would then proceed to correct the information where it is deemed to be incorrect or otherwise maintain it. Connectivity to a backend database, for example, MS SQL server via the client side form is therefore an important property of eforms and ensures accuracy of data by constantly having the user review and update past information.
- iii) Process automation platform: PDF, infopath and what we may be more familiar with - word files and excel spreadsheets - are client side formats which do not have a server infrastructure built around them. In other words, the process of form transfer from the country and submission to the United Nations Statistics Division is basically via the use of email. By this method, the process of doing form validation, submission tracking, review and approval of submissions is ad hoc and non-standard. A server infrastructure built around PDF, infopath and other World Wide Web Consortium (W3C) compliant Xform formats has the distinct advantage of allowing users submitting information to do so in a structured environment, and in a repeatable way.

Many organizations have a portal site run by a web server usually IIS or Apache Tomcat. A form server running on this web server can be configured to serve up to several users via secured logins or emailing of web references to the user, public facing forms for structured submission of data to the backend database system (MS SQL Server or Oracle).

- iv) Web based online forms repository: Because the form server is online all the benefits inherent in a web-based application apply. These include round the clock availability, offline filling and submitting of forms at an appropriate time using the same server infrastructure described earlier, secure access with user group permissions and online authentication, and integrated search, amongst others.

B. Description of an eform infrastructure

3. An eform infrastructure has a data process environment consisting of two entirely different sets of users. There are the internal and remote users within the United Nations Statistics Division who would access the system from any web desktop (a computer running internet explorer and having access to the form server web site) on the intranet or the Internet. These users would design, publish, route, approve and track forms and attached documents. There are also the external participants who access and complete the form online which are designated for anonymous access; forms can be routed to these external participants via URL links embedded in emails, the main issue being that these users have access rights limited to form submission.

4. An eform infrastructure would also fit into the existing network environment and would come with all the features necessary to allow this to happen. An eform infrastructure would be installed against a backend database like MS SQL or Oracle. It would leverage existing network login security protocols through the use of lightweight directory access protocol (LDAP) and single sign-on technology. It would be possible to complete event notification and routing using the existing email system. This type of notification can, for example, be triggered by the approval of a particular country's data submission by a member of staff of the specific United Nations agency.

C. Advantages of using an eform infrastructure

- 5. These advantages stem from three distinct phases in the eform implementation, namely:
 - i) The design and publication of the form: The key advantage in this area is the possibility of having one instance of a public facing form that any country, regional office, satellite office of a national statistical organization that a respondent can complete. This can be compared with at worst printing and mailing these forms to each of the respective respondents. Many national statistical offices are still heavily dependant on printed forms as a means for data gathering (from establishments of the Balance of Payments and National Accounts data principally, but other forms of

establishment data as well). The implementation of public facing eforms can significantly improve the efficiency with which this activity can be accomplished. This is even more so with the rapid improvement in offline fillable PDFs occurring within Adobe.

- ii) Data collection and form validation: Many international statistical organizations and to a lesser extent national statistical organization now collect data by emailing of predominantly word or excel spreadsheets. While this approach has the effect of saving printing cost, it has significant problems stemming from the lack of validation rules and the design logic inherent in W3C certified Xform specific forms format. The lack of these basic tenets of XML-based eform design results in erroneous and invalid entries, changes to the structure of the original document submitted due to default or direct format changes by the respondent, the risk of the introduction of a computer virus since the form is entirely exposed to the respondents internal computing systems and a variety of other similar problems. Public facing eforms has distinct advantages in this respect.
- iii) Process automation: these advantages stem from the server infrastructure surrounding the form and its adherence to rules laid out in the workflow design inherent in the form and the structures imposed by the form server which it is tied to.

D. The Implementation of eform data gathering solutions

6. Eforms infrastructure and management is now an integral part of most enterprise-wide content management solutions available from Cardiff, Filenet, Adobe and Microsoft with the marrying of infopath with Electronic Content Management (ECM) software, etc. These solutions are increasingly solving problems in a commercial environment with document management and other supporting software infrastructure and internal business process activity. They are therefore, cost effective in relation to the problem at hand, which is, the effective gathering of data through forms driven processes. A complete eform centric cost effective server solution is the most efficient approach to implementation of this solution.