# **EPOS** Project Report

Final report to the Swedish Council for Scientific Information and Documentation, Stockholm

by Mats Lindqvist, Christer Bryntesson, Roland Hjerppe, Jan Hultgren, Bo Johansson, Rolf Larsson and Göran Thorén.

**JUNE 1973** 

Project EPOS Royal Institute of Technology Library S 100 44 Stockholm, Sweden.



157 232

# TABLE OF CONTENTS

PROJECT REPORT	
Background and objectives	1
Project scope	3
Project organization	4
EPOS COMMUNICATION LANGUAGE	
EPOS in the SDI-system	6
Input transactions	7
Sub-profiling	15
Input via GUTS	16
SYSTEM DESCRIPTION	
System overview	18
File organization	19
Data structure	25
Main functioning	29
Utility programs and macros	34
SYSTEM PERFORMANCE	36
REFERENCES	39
GLOSSARY	40
APPENDIX: EXTENDED STUDIES	

A Method Defining a Limited Set of Character-strings with Maximal Coverage of a Sample of Text

#### PROJECT REPORT

# BACKGBOUND AND OBJECTIVES

The first phase of the system development was done before project EPOS was launched and the work was carried out by <u>Rolf Larsson</u>, <u>Christer</u> Bryntesson, and Claes Rickeby<sup>\*</sup>.

Based on the results from this pre-study project EPOS was initialized in June 1972 to develop a system that would improve man's possibilities to manipulate large text files; especially query maintenance in SDIsystems.

The "core" of the project staff has been the analyst/programmers Christer Bryntesson and <u>Bo Johansson</u>, and the following people have all contributed to the development of EPOS to a smaller or larger part depending on their other commitments: <u>Roland Hjerppe</u>, RIT Library <u>Rune Isaksson</u>, EMDC (Karolinska Institutet) Rolf Larsson, SINFDOK Göran Thorén, SINFDOK

<u>Mats Lindqvist</u>, RIT Library, has been the principal investigator for the project.

During the latter half of the project Mr. Jan Hultgren has been attached to the project as a consultant in linguistics and mathematical statistics. EPOS was developed within a research project sponsored by the Swedish Council for Scientific Information and Documentation (SINFDOK). The project description was "the development of a terminal oriented system for the manipulation of text; especially profile maintenance in SDIsystems".

The computer-based documentation systems in Sweden have grown considerably since 1967 and are today sizeable even in an international perspective.

It has become quite obvious that the bottleneck in the continuing expansion is the query-formulation and follow-up ("profiling"). The most valuable asset of the documentation services is the "know how" of profiling which is materialized in the query master file. The profiling function in the SDI-system can be described as a very expensive craft where akin processes are performed mostly without a planned experience feed-back.

The economy of the whole enterprise could be improved if the accumulated experience of the documentalists could be fed back into the query formulation process. By making the complete profile easily accessible and served by a programming system that will allow copying of profiles and parts thereof to an, in principle, unlimited extent, a more "concept oriented" query base could be achieved which by the creation of "subprofiles" would give more efficient profiling as well as better search economy.

Special attention was drawn to the problem of compatibility; it was felt that practical restrictions, such as machine availability to the SDI-services in Sweden, dominated the decision space and the development was to be made on an IBM 360/75.

## PROJECT SCOPE

As a result of the pre-EPOS study it was decided that the system should be transaction oriented and should use a command language with a fairly general assignment statement as its central feature. The last decision was made in order to keep the query or profile handling concise and yet clear.

The primary function of the EPOS system was to serve as query or profile editor for an operational SDI-system, VIRA, which was being run on a university computer (IEM 360/75) in Stockholm. This installation was also chosen for the software development of EPOS, as any other choice would have implied additional probably manual interface routines between the search system and the profile handler. This choice also decided what terminal type to use; in spite of its manifold drawbacks a typewriterterminal was chosen. The software development was re-assessed in the early fall 1973 in light of the new software services that were being offered from the computing center. The GUTS time sharing system, with its conversational RJE, was economically a very attractive alternative to a dedicated interactive system. Functionally the GUTS alternative would give editing facilities to the profile-manipulator and GUTS' procedures would make the system fairly easy to handle for a non-professional user.

The major drawback with GUTS-EPOS was the response time but after negotiations with the computer center file-access was given 7 hrs each day. This meant that response time could be brought down to a couple of minutes, as compared with half-hours in the SETUP-queue if the file were to be mounted upon demand.

After this development it was decided that EPOS should be adopted to the GUTS conversational environment instead of having the project develop a special software package for interactive manipulation of files.

The decision to develop GUTS-EPOS also made it possible to devote more attention to "extended studies" within project EPOS. It was felt, although EPOS was progressing, that more and deeper background studies were needed to improve on similar systems in the future. One field of study that in many ways was related to the problems relevant to EPOS was "data compaction" and a study of string coverage in available reference data bases was conducted as a preliminary step to compression of textual data.

## PROJECT ORGANIZATION

The question of how a research project like EPOS should be organized and managed was discussed within the project as part of the continuous "project assessment". This assessment had the form of weekly meetings or "talks" and had a co-ordinating function in addition.

Based on the experiences from EPOS the desirability of a semi-formal intraproject training program with seminars, meetings and possibly conferences was clearly seen. The small attempts made in project EPOS in this respect had a very positive effect on the project as a whole: communication within the project was more effective and job partitioning and co-ordination was improved.

Project work in the field of "information retrieval" is in some ways different from system development in the ADP-field in general. The focus on very large information collections, that sometimes are not well structured, puts proportionately very much importance on methods for information handling and the solution to methodological problems often have a fundamental impact on the gross system. This means that in projects like EPOS responsibility should not be partitioned by the "programmer - analyst" distinction, but rather by sub-system boundaries and within that sub-system responsibility should encompass " overall system design" as well as "coding".Responsibility divided by sub-systems also gives better possibilities of effective integration of the sub-systems.

As for the external organization the main question is the connection between the project and the operational environment for the system. For project EPOS the ties with the SDI-service at the RIT library have been strong and intimate to both advantage and disadvantage to the parties.

The close connection has made it possible to effectuate relevant optimizations of the SDI-system and valuable experiences about the applicability of different methods for information retreival have been made. The size and quality of the SDI-service have been instrumental in assuring validity and generality of these experiences.

The disadvantage that has been most obvious is the existence of external dis-economies for the project due to unclear responsibilities for the implementation of the system. With the relatively small staff that has been working on EPOS it was not possible to take care of implementation-emergency-problems without disturbing the progress of the project. Such disturbancies were increasingly common as the system approached completion.

Although far from unaware of the problems of implementation at the beginning of project EPOS we now consider this problem to be fundamental in the project definition.