# FEASIBILITY STUDY

## ON AN OPTIMAL OPTION OF SOFTWARE AND HARDWARE FOR AUTOMATED PERSONIFIED DATA INPUT FROM THE PFU PRIMARY RECORDS USING OPTICAL SCANNERS OF DIFFERENT PERFORMANCE

**Prepared by USAID/PADCO under Social Sector Restructuring Program**

Kyiv 1999

# CONTENTS

INTRODUCTION ..... 3

1. ANALYSIS OF DIFFERENT TYPES OF AUTOMATED TEXT INPUT AND RECOGNITION SYSTEMS USING OPTICAL SCANNERS ..... 3
   1.1. Typical configuration of an automated input and recognition system ..... 3
   1.2. Low-performance SPOV-OS system (SPOV-OS-L, up to 1,000 pages/dayshift) ..... 4
   1.3. Low to medium-performance SPOV-OS systems (SPOV-OS-LM1 and SPOV-OS-LM2) ..... 4
      1.3.1 SPOV-OS-LM1, up to 2,400 pages/dayshift ..... 4
      1.3.1 SPOV-OS-LM2, up to 3,600 pages/dayshift ..... 4
   1.4. Medium-performance SPOV-OS system (SPOV-OS-M, up to 10,000 documents/shift) ..... 6
   1.5. High-performance SPOV-OS system (SPOV-OS-H, up to 15,000 documents/shift) ..... 6
2. ANALYSIS OF NEEDS OF THE SPOV-OS SYSTEMS OF DIFFERENT PERFORMANCE AROUND UKRAINE ..... 7

CONCLUSIONS ..... 10

# INTRODUCTION

According to the current legislation the Pension Fund of Ukraine (PFU) is responsible for keeping personified records of contributions to PFU from all working citizens of Ukraine (who are simultaneously the insured persons within the scheme of compulsory pension insurance).

One of the aspects in keeping personified records of the insured is collection and accumulation of information on individuals.

It is clear that, in order to complete this task, enterprises should be maximally encouraged to transfer the insured-related data to PFU on electronic data carriers. It is equally clear, however, that in many cases such information will be transferred on paper, especially in rural regions.

To speed up the process of paper-born data input, systems can be used that are based on automation of text input using optical scanners and its recognition using special software.

The following job was done in the study as presented in this report:

- performance analysis of different types of automated text input and recognition systems;
- needs analysis of different system performances;
- estimation of funds needed for purchase of the hardware/software systems for automated documents input in PFU personified record keeping.

Another report, "Choosing a character recognition software to speed up the input of personified data on contributions to the Pension Fund of Ukraine," presents an analysis of available stylized character recognition programs and reasons that the FineReader 4.0 Rukopis ["Manuscript"] of the ABBYY company should be used for this purpose. The analysis was done by making use of the Russian experience in implementation of similar systems carried out by ABBYY, whereof relevant information had been obtained through the courtesy of this company.

## 1. ANALYSIS OF DIFFERENT TYPES OF AUTOMATED TEXT INPUT AND RECOGNITION SYSTEMS USING OPTICAL SCANNERS

### 1.1. Typical configuration of an automated input and recognition system

The character recognition programs used for large-scale recognition of documents entered from optical scanners are a component of the automated text input and recognition systems or, using the terminology of PFU, *the program-and-technical complexes of means for automated input of primary reporting documents for personified records using optical scanners* (SPOV-OS systems). Such systems are usually composed of an optical scanner, several locally networked computers, and appropriate software. Scanners of a different pages-per-minute speed may be used in different systems, and as for the software, it includes an adapted version of FineReader 4.0 Rukopis, a Windows-type operating system and special software for the formation of recognition output and its transfer to further steps in the information handling process.

The performance of any given SPOV-OS system is defined not only by the number of pages that may be entered by a scanner per minute but also by the general number of documents that may be processed using this system per shift.

Reviewed below are possible configurations of SPOV-OS systems having different performance as recommended by ABBYY.

### 1.2. Low-performance SPOV-OS system (SPOV-OS-L, up to 1,000 pages/shift)

| Description | Model | Quantity | Unit price, $ | Amount, $ |
|---|---|---|---|---|
| Scan and recognition station | Pentium 200 with 4Gb hard disk | 1 | 1,600 | 1,600 |
| Operator station | Pentium 200 with 1Gb hard disk | 1 | 1,400 | 1,500 |
| High-speed scanner | Fujitsu ScanPartner 600C (8 pages/min) | 1 | 2,500 | 2,500 |
| Recognition software | FineReader 4.0 Rukopis | 3 | 1,200 | 3,600 |
| Other software and peripherals | Windows, network, etc. | | 500 | 500 |
| | | **Total** | | **9,700** |

The above system may be run by one operator. During the peak load hours, the scan station may need attendance of one more operator.

### 1.3. Low to medium-performance SPOV-OS systems (SPOV-OS-LM1 and SPOV-OS-LM2)

#### *1.3.1 SPOV-OS-LM1, up to 2,400 pages/shift*

| Description | Model | Quantity | Unit price, $ | Amount, $ |
|---|---|---|---|---|
| Scan station | Pentium 200 with 2Gb hard disk | 1 | 1,300 | 1,300 |
| Recognition station | Pentium 200 with 2Gb hard disk | 1 | 1,300 | 1,300 |
| Operator station/file server | Pentium 200 with 1Gb hard disk | 1 | 1,500 | 1,500 |
| High-speed scanner | Fujitsu M3093 (20 pages/min) | 1 | 4,000 | 4,000 |
| Recognition software | FineReader 4.0 Rukopis | 3 | 1,450 | 4,350 |
| Other software and peripherals | Windows, network, etc. | | 500 | 500 |
| | | **Total** | | **12,950** |

The system is rated to being run by two operators: one for the scan station, and another one for the operator station (verification station).

#### *1.3.2 SPOV-OS-LM2, up to 3,600 pages/shift*

| Description | Model | Quantity | Unit price, $ | Amount, $ |
|---|---|---|---|---|
| Scan station | Pentium 200 with 2Gb hard disk | 1 | 1,300 | 1,300 |
| Recognition station | Pentium 200 with 2Gb hard disk | 1 | 1,300 | 1,300 |
| Operator station/file server | Pentium 200 with 1Gb hard disk | 2 | 1,500 | 3,000 |
| High-speed scanner | Fujitsu M3093 (20 pages/min) | 1 | 4,000 | 4,000 |
| Recognition software | FineReader 4.0 Rukopis | 4 | 1,210 | 4,840 |
| Other software and peripherals | Windows, network, etc. | | 500 | 500 |
| | | **Total** | | **14,940** |

The system is rated to being run by three operators: one for the scan station, and two for the operator (verification) and recognition stations.

## 1.4. Medium-performance SPOV-OS system (SPOV-OS-M, up to 10,000 documents/shift)

*Scanner Kodak 7500 - 1 unit*

| Pages/min as per specification | Pages/min actually | Max. doc./shift | Price*, $ |
|---|---|---|---|
| 80 | 50 | 10,000 | 55,000+9,000+6,000 |

* Scanner + scanner-specific software + one-year guarantee.

*Scan station - 1 unit*

| | Processor | RAM, Mb | Monitor | Operating system |
|---|---|---|---|---|
| Minimum requirements | Pentium 133 | 32 | irrelevant | Windows 95 |
| Recommended | Pentium II | 32 | irrelevant | Windows NT 4.0 |

*Recognition server - 3 units*

| | Processor | RAM, Mb | Monitor | Operating system |
|---|---|---|---|---|
| Minimum requirements | Pentium II 266 | 64 | irrelevant | Windows NT 4.0 |
| Recommended | Pentium II 350 | 64 | irrelevant | Windows NT 4.0 |

*Verification (visual control) stations - 13 units*

| | Processor | RAM, Mb | Monitor | Operating system |
|---|---|---|---|---|
| Minimum requirements | Pentium 133 | 32 | 14" | Windows 95/NT 4.0 |
| Recommended | Pentium II | 64 | 17" | Windows 95/NT 4.0 |

*File server - 1 unit*

| | Processor | RAM, Mb | Monitor | Operating system |
|---|---|---|---|---|
| Minimum requirements | Pentium 133 | 32 | irrelevant | Windows NT 4.0 |
| Recommended | Pentium II | 64 | irrelevant | Windows NT 4.0 |

- disk space 6 Gb+.

*Export server - 1 unit*

| | Processor | RAM, Mb | Monitor | Operating system |
|---|---|---|---|---|
| Minimum requirements | Pentium II 266 | 64 | irrelevant | Windows NT |
| Recommended | Pentium II 350 | 64 | irrelevant | Windows NT |

- necessary for data export in a format meeting requirements of the Addendum #14 to the Russia's Tax Inspection Instruction #35 of June 29 1995;
- all computers must be equipped with netcards (at least 10 Mbaud).

The number of attending personnel for one shift: 17 (two for the scanner, one system administrator, 13 operators and one shift supervisor).

The cost of the system: $92,000, including scanner at $70,000 and 17 FineReader Rukopis licenses at $1,299 each.

## 1.5. High-performance SPOV-OS system (SPOV-OS-H, up to 15,000 documents/shift)

*Scanner Kodak 9500 - 1 unit*

| Pages/min as per specification | Pages/min actually | Max. doc./shift | Price* |
|---|---|---|---|
| 160 | 80 | 15,000 | 90,000+10,000+8,000 |

* Scanner + scanner-specific software + one-year guarantee.

*Scan station - 1 unit*

| | Processor | RAM, Mb | Monitor | Operating system |
|---|---|---|---|---|
| Minimum requirements | Pentium 133 | 32 | irrelevant | Windows 95 |
| Recommended | Pentium II | 64 | irrelevant | Windows NT 4.0 |

*Recognition server - 4 units*

| | Processor | RAM, Mb | Monitor | Operating system |
|---|---|---|---|---|
| Minimum requirements | Pentium II 266 | 64 | irrelevant | Windows NT 4.0 |
| Recommended | Pentium II 350 | 64 | irrelevant | Windows NT 4.0 |

*Verification (visual control) stations - 20 units*

| | Processor | RAM, Mb | Monitor | Operating system |
|---|---|---|---|---|
| Minimum requirements | Pentium 133 | 32 | 14" | Windows 95/NT 4.0 |
| Recommended | Pentium II | 64 | 17" | Windows 95/NT 4.0 |

*File server - 1 unit*

| | Processor | RAM, Mb | Monitor | Operating system |
|---|---|---|---|---|
| Minimum requirements | Pentium 133 | 32 | irrelevant | Windows NT 4.0 |
| Recommended | Pentium II | 64 | irrelevant | Windows NT 4.0 |

- disk space 8 Gb+.

*Export server - 1 unit*

| | Processor | RAM, Mb | Monitor | Operating system |
|---|---|---|---|---|
| Minimum requirements | Pentium II 266 | 64 | irrelevant | Windows NT 4.0 |
| Recommended | Pentium II 350 | 64 | irrelevant | Windows NT 4.0 |

- necessary for data export in a format meeting requirements of the Addendum #14 to the Russia's Tax Inspection Instruction #35 of June 29 1995;
- all computers must be equipped with netcards (at least 10 Mbaud).

The number of attending personnel for one shift: 24 (two for the scanner, one system administrator, 20 operators and one shift supervisor).

The cost of the system: $140,000, including scanner at $108,000 and 25 FineReader Rukopis licenses at $1,299 each.

The quantity of verification stations specified in the tables above is the upper bound. The specifics of verification operator work allows involving in it persons not professionally trained in keyboarding. Therefore, the number of verification staff may be less than specified in the comments under the tables, while involving additional manpower only in cases of peak loads.

## 2. ANALYSIS OF NEEDS OF THE SPOV-OS SYSTEMS OF DIFFERENT PERFORMANCE AROUND UKRAINE

To choose an acceptable configuration of SPOV-OS for any given application conditions it is necessary to estimate its required performance, i.e. the number of documents the system is expected to enter per shift.

The Table 1 shows the numbers of Ukrainian districts against employed population brackets.

As can be seen from the Table 1, the employed population distribution among the districts is very uneven, overwhelming majority of the districts (531 of 735) having 5,000 to 30,000 employed population. It is obvious that these are predominantly rural districts having

**Table 1. The distribution of employed population among Ukrainian districts**

| | Employed population brackets, ×1,000 | Number of districts |
|---|---|---|
| 1 | 5-10 | 109 |
| 2 | 11-20 | 291 |
| 3 | 21-30 | 131 |
| 4 | 31-40 | 64 |
| 5 | 41-50 | 48 |
| 6 | 51-60 | 31 |
| 7 | 61-70 | 27 |
| 8 | 71-80 | 9 |
| 9 | 81-90 | 9 |
| 10 | 91-100 | 9 |
| 11 | 101-110 | 4 |
| 12 | 111-120 | 2 |
| 13 | 121-130 | 1 |

virtually no computers. This implies the necessity of paper-born data input.

The following variants of SPOV-OS allocation in Ukraine are possible:
1. In each district.
2. In inter-district centers for several small districts.
3. In oblast centers.
4. In regional centers (four to seven for entire Ukraine).

In case of the first variant, employers will submit their statements on paper carriers to a PFU district office where these are immediately entered using SPOV-OS. According to the second version, the employers' statements will not be entered in the district offices but, duly packed and labeled, will be forwarded to an inter-district center for the data input. The third version provides for forwarding the packed statements for this purpose to the respective oblast center. And finally, the statements will be sent to the nearest regional center to be processed there if the fourth option is chosen.

However, taking into account that PFU has planned storage of the primary reporting documents only in its district offices, the first and, probably, second variants of SPOV-OS allocation should be considered basic.

Pursuant to the current reporting rules, all employers' statements must be submitted to the PFU district units within the first quarter of the year following the reporting year. This means that all the statements should be entered within three months or, assuming 21 working day (21 shift) a month on average, during 63 shifts. In our further calculation we will proceed from one-shift operation of SPOV-OS systems at eight-hour shifts.

The individual data are currently written on one page. However, the form suitable for the automated input and recognition will occupy at least two pages. To estimate the necessary performance, let's assume that the number of pages of the documents to be entered is twice the number of employed population.

The estimates of the quantity of SPOV-OS systems having different performance for the different district brackets are presented in the Table 2. It was assumed that at least one SPOV-OS system of some or other performance should be installed in every district. For example, whereas the input of about 1,905 documents should be secured per shift in the districts of the sixth bracket, this may be achieved by using two SPOV-OS-L systems or one SPOV-OS-LM system or one SPOV-OS-M system (the latter will allow to process this number of documents within 12 shifts, i.e. the rest of the time the system is going to be idle).

**Table 2. The estimates of the quantity of SPOV-OS systems required for Ukraine**

| | Employed population brackets, ×1,000 | Number of document pages, ×1,000 | Number of documents per shift | Number of SPOV-OS systems* per district | | | Number of districts | Total Number of SPOV-OS systems* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | L | LM | M | | L | LM | M |
| 1 | 5-10 | 20 | 317.46 | 1 | 1 | 1 | 109 | 109 | 109 | 109 |
| 2 | 11-20 | 40 | 634.92 | 1 | 1 | 1 | 291 | 291 | 291 | 291 |
| 3 | 21-30 | 60 | 952.38 | 1 | 1 | 1 | 131 | 131 | 131 | 131 |
| 4 | 31-40 | 80 | 1269.84 | 1 | 1 | 1 | 64 | 64 | 64 | 64 |
| 5 | 41-50 | 100 | 1587.30 | 2 | 1 | 1 | 48 | 96 | 48 | 48 |
| 6 | 51-60 | 120 | 1904.76 | 2 | 1 | 1 | 31 | 62 | 31 | 31 |
| 7 | 61-70 | 140 | 2222.22 | 2 | 1 | 1 | 27 | 54 | 27 | 27 |
| 8 | 71-80 | 160 | 2539.68 | 3 | 1 | 1 | 9 | 27 | 9 | 9 |
| 9 | 81-90 | 180 | 2857.14 | 3 | 1 | 1 | 9 | 27 | 9 | 9 |
| 10 | 91-100 | 200 | 3174.60 | 3 | 2 | 1 | 9 | 27 | 18 | 9 |
| 11 | 101-110 | 220 | 3492.06 | 4 | 2 | 1 | 4 | 16 | 8 | 4 |
| 12 | 111-120 | 240 | 3809.52 | 4 | 2 | 1 | 2 | 8 | 4 | 2 |
| 13 | 121-130 | 260 | 4126.98 | 4 | 2 | 1 | 1 | 4 | 2 | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | **Overall** | **916** | **751** | **735** |

* The number of SPOV-OS systems has been calculated by comparing the necessary performance (documents processed per shift) with the respective system performance (L: 1,000 per shift; LM: 2,400 to 3,600 per shift; M: 10,000 per shift).

Hence, to install at least one SPOV-OS system in each district, Ukraine will need 916 SPOV-OS-L systems or 751 SPOV-OS-LM systems or 735 SPOV-OS-M systems that will cost $8,885,200 or $9,725,450 or $67,620,000, respectively.

As demonstrated by the Table 2, if the data input centers are located in each district center, the medium, and the more so high, performance SPOV-OS systems will considerably excess the requirements, even for the district with the largest employed population. Keeping also in mind their high price, one can come to a conclusion of inexpediency of their use given such an SPOV-OS allocation layout.

In the districts falling within the first and second brackets, even the SPOV-OS-L systems would have a performance higher than necessary for one such district. This allows installing SPOV-OS not in every district but using such a system of one of districts for the input of documents from other, close to this one, districts where the employed population is small (i.e. using it as an inter-district documents input center). In the districts where more than 1,000 documents must be entered per shift, several SPOV-OS-L or SPOV-OS-LM systems may be used, or a combination of these.

Thus:

- in the districts falling within the brackets 1 through 6, one or two SPOV-OS-L systems may be used and inter-district centers organized in some of the districts;
- in the districts falling within the brackets 7 through 13, two to four SPOV-OS-L systems or one or two SPOV-OS-LM systems may be used.

The Table 3 provides estimates of the quantity of SPOV-OS systems (L or LM) for Ukraine, given arrangement of the inter-district centers, and their costs.

**Table 3. Quantity of SPOV-OS systems**

| | Employed population brackets, ×1,000 | Number of documents per shift | Number of districts | Number of SPOV-OS systems adjusted for inter-district centers | |
|---|---|---|---|---|---|
| | | | | L | LM |
| 1 | 5-10 | 317.46 | 109 | 34 | 15 |
| 2 | 11-20 | 634.92 | 291 | 185 | 77 |
| 3 | 21-30 | 952.38 | 131 | 125 | 52 |
| 4 | 31-40 | 1269.84 | 64 | 81 | 34 |
| 5 | 41-50 | 1587.30 | 48 | 76 | 32 |
| 6 | 51-60 | 1904.76 | 31 | 59 | 25 |
| 7 | 61-70 | 2222.22 | 27 | 60 | 25 |
| 8 | 71-80 | 2539.68 | 9 | 23 | 10 |
| 9 | 81-90 | 2857.14 | 9 | 26 | 11 |
| 10 | 91-100 | 3174.60 | 9 | 29 | 12 |
| 11 | 101-110 | 3492.06 | 4 | 14 | 6 |
| 12 | 111-120 | 3809.52 | 2 | 8 | 3 |
| 13 | 121-130 | 4126.98 | 1 | 4 | 2 |
| | Overall | | 735 | 724 | 304 |
| | **Costs** | | | **$7,022,800** | **$3,936,800** |

When comparing the data in Tables 2 and 3, it can be seen that having the inter-district documents input centers allows a significant decrease in the SPOV-OS costs. However, in case of the inter-district centers, the costs of delivering the statements to and returning them from the centers, now hardly yielding to estimation, should be also taken into account.

An administrative reform is expected to start in Ukraine in the near future, being aimed to reduce the number of districts and oblasts. This may have a palpable impact on the estimation of the systems quantity and overall value.

Also to be considered is the computerization level in enterprises in different Ukrainian districts. In the industrialized districts, up to 90 percent of statements may be submitted on electronic carriers, and 10 percent on paper carriers, while in the rural districts it is vice versa. Therefore, to make a substantiated choice, the computerization level in the accounting departments of enterprises in different districts of Ukraine should be analyzed, and the number of paper forms to be entered in each district should be made more precise.

The above (and, maybe, also other) factors may significantly influence the choice of some or other SPOV-OS allocation variant, calling for a multi-criterial assessment. Since there is no complete information on all of the mentioned factors at the moment, the present estimation of the value of systems having different performance and the costs of the Ukraine-wide introduction of a corresponding variant has been solely based on the prices for the hardware and software.

**Note:** All of the above tables specify approximate prices for the hardware and software. In particular, the prices may vary depending on the order volume. In case of purchase of several systems, each of them will cost less than one bought separately. In case of purchase of one high-performance system, its components and accessories may turn cheaper than those of a less productive system. For example, having got a general license for use of FineReader 4.0 Rukopis, the costs may be reduced to a great extent.

The scanner prices may also significantly vary at the same level of performance, inter alia depending on the scanner manufacturer.

It should be noted that the prices herein have been quoted for Russia. The features of Ukrainian legislation, in particular customs legislation, may lead to a change in the quotations.

The scanners brand names specified in the tables are best but may be substituted with others when necessary.

## CONCLUSIONS

1. For the current concept of allocation of the documents input and storage centers (in every district), the SPOV-OS systems of a low or low-to-medium performance will be sufficient for processing individual data on the insured.
2. In case of organization of inter-district documents input centers, the SPOV-OS costs may be significantly reduced. In this case, entire Ukraine will need 724 SPOV-OS-L systems or 304 SPOV-OS-LM systems at $7,022,800 or $3,936,800, respectively.
3. In case of using SPOV-OS for the input of not only the individual data but also other primary reporting documents, a new calculation of the SPOV-OS systems quantity and value should be done.
4. In cases of a change in the concept of the documents input and storage centers allocation (e.g. in oblast or regional centers), an administrative reform in Ukraine (i.e. merger of districts and oblasts), use of cheaper computers and scanners (or use of earlier bought computers), obtaining a general license for FineReader 4.0 Rukopis, and some other factors, the costs of SPOV-OS purchase may be significantly reduced.