Human Resource Social Insurance Data Remote Reporting System Based on Big Data Technology

Cai-ming Zhang(✉) and Bo Sun

School of Labor Relations and Human Resource, China University of Labor Relations, Beijing 100048, China

Abstract. In order to cooperate with the China Insurance Regulatory Commission's national economic census work, meet the various needs of the Insurance Regulatory Commission for insurance statistical information work, and increase the transmission speed of social insurance data. Therefore, this paper designs a remote reporting system for human resources and social insurance data based on big data technology. Upload the data file structure through the web method and manually enter the data report to optimize the C/S structure in the system hardware. Improved the system software module functions and operating steps. The statistical data collected from the business system, system, and personnel system is reported from low-level (prefecture-level) to high-level (head office level). The data is summarized and sorted at all levels. A file containing all off-site statistical data is generated. Finally, the actual application of the human resources and social insurance data reporting system in different places is investigated and analyzed. The experimental results confirm that the system has a high reporting rate and fully meets the research requirements.

Keywords: Big data technology · Human resources · Insurance data

1 Introduction

It has the function of economic security and social security, and can promote the development of insurance and financing. In recent years, with the rapid development of China's insurance industry as a whole, the situation of increasing spatial differentiation is more and more obvious, which actually affects the sustainable development of the insurance industry and the effective play of its function [1]. At the same time, the theoretical research on the insurance industry is constantly enriched, while the academic research on the insurance industry space is relatively backward. The theoretical framework of insurance spatial differentiation is still in the exploratory stage, which leads to a lack of in-depth understanding of the trend and degree of insurance industry differentiation, differentiation mechanism, influencing factors and change trend. The practice of policy design and spatial layout of insurance supervision lacks theoretical guidance. In order to cooperate with the China Insurance Regulatory Commission's national economic census work, meet the various needs of the Insurance Regulatory Commission for insurance

S. Liu and X. Ma (Eds.): ADHIP 2021, LNICST 416, pp. 551–561, 2022.
https://doi.org/10.1007/978-3-030-94551-0_43

statistical information work, and increase the transmission speed of social insurance data. Therefore, this paper designs a remote reporting system for human resources and social insurance data based on big data technology. Upload the data file structure through the web method and manually enter the data report to optimize the C/S structure in the system hardware. Improved the system software module functions and operating steps. The statistical data collected from the business system, system, and personnel system is reported from low-level (prefecture-level) to high-level (head office level). The data is summarized and sorted at all levels. A file containing all off-site statistical data is generated. Finally, the actual application of the human resources and social insurance data reporting system in different places is investigated and analyzed. The experimental results confirm that the system has a high reporting rate and fully meets the research requirements.

2 Human Resource Social Insurance Data Remote Reporting System

2.1 Hardware Structure of Social Insurance Data Remote Reporting System

The development and design of social insurance data remote reporting system is realized by combining Java servlet, JSP, HTML and JavaScript [2]. This model is page based, with each JSP file or servlet file linked to another. Clicking on the link will send a request to the server. The server will process the request and replace the browser's page. Each user's behavior will lead to an HTTP request. Data is provided by each business system of each central branch company, uploaded to the statistical information reporting system by means of file or input and ETL automatic data. After the summary, review and

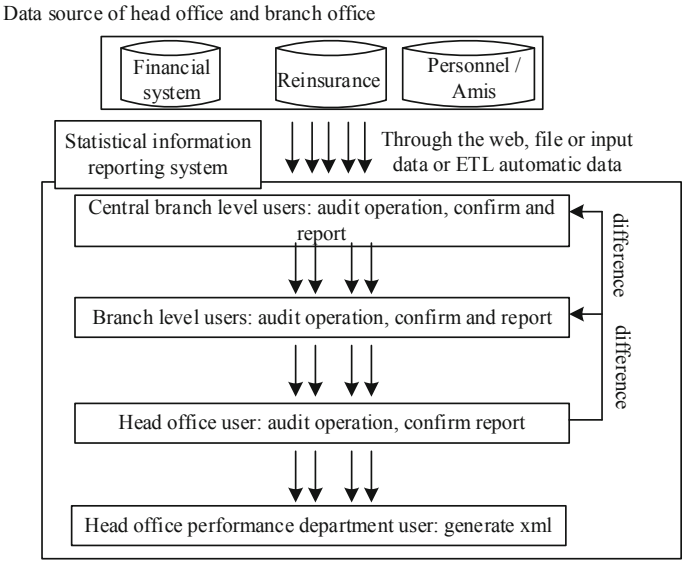

Fig. 1. Hardware configuration of social insurance data remote reporting system

confirmation of provincial and head office companies, it is submitted to the performance Department of the head office. The performance department uses the system to generate the overall data file of the whole company in XML format submitted to CIRC [3]. In order to test the effect of data processing, the hardware configuration of social insurance data remote reporting system is optimized, as shown in Fig. 1.

As shown in Fig. 1, server-side control technology simulates stateful connection based on stateless HTTP, and realizes state maintenance and event processing [4]. Write B/s program in the way of desktop program. Some desktop programs that used to be very complex can now be easily implemented in browsers, and even code migration is very easy [5]. The client does not need any additional plug-ins. Upload the data file through the web, input the data manually and report the data. Generate the data file in XML format submitted to CIRC. System management subsystem: used to maintain the data needed by the system operation. Further optimize the function of c/s system structure of social insurance data remote reporting system, as shown in Fig. 2.

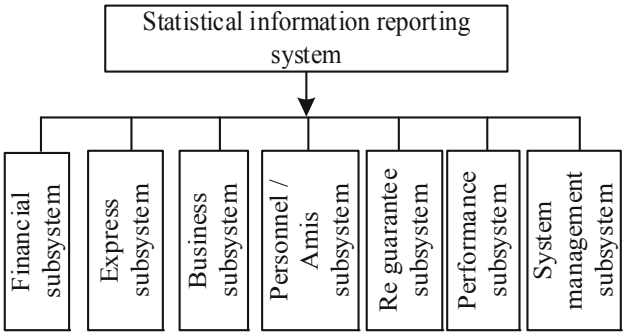

Fig. 2. C/S architecture of social insurance data remote reporting system

For the C/S architecture can not meet the needs of the development of distributed enterprise application system, developers constantly put forward new software architecture to improve the difficulties in the development and maintenance process [6]. In the current enterprise and application system development field, there are two camps of software architecture, they can provide a powerful and complete development platform. These two platforms are known as. Net architecture and J2EE architecture.

2.2 Software Flow of Insurance Data Remote Reporting System

From the business point of view, the information reporting system can be divided into the following subsystems.

Business subsystem: including data upload, data modification, account verification, report generation, summary, data audit and other functions.

Personnel and Amis subsystem: including data input, data modification, account verification, report generation, summary, data audit and other functions.

Reinsurance subsystem: including data upload, data modification, account verification, data reporting, single account query, data prompt, report generation, summary and report audit, excel upload, data modification, account verification and report generation.

Performance subsystem: including data upload, data modification, account verification, export XML file and download functions.

System management: including role management, organization management, system user management, function module management and data source configuration. The subsystem reporting process is shown in Fig. 3.

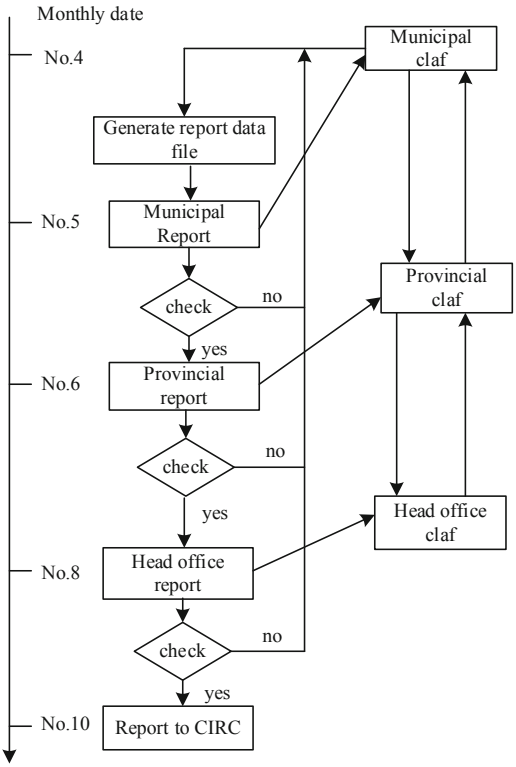

Fig. 3. System software operation process optimization

Municipal users generate data submission documents through core business systems and systems, and then upload data submission files to the information submission system [7]. The system will upload data for analysis and generate city level reports, and then the user will verify the report items. Submit data to provincial branch after verification is successful. The report of prefecture and city level is in a paragraph.

After the provincial users log in, the data of their subordinate prefectures and cities are collected and sorted, and the provincial reports are generated. After the report item is verified successfully, it shall be submitted to the head office.

After the head office user logs in, the data of all its branches will be summarized and sorted, and the head office report will be generated. After the report items are verified successfully, the XML file containing all the submitted subjects will be generated. This module is the main module for each city branch and business data upload [8]. Support

to upload the TXT text of account information separated by "," to the server. At the same time, check the uploaded account information. The system administrator and the business role can configure the header code of the uploaded file to determine the name of the uploaded file.

Table 1. Function point description of upload module

Function point	Describe
File name check	Check whether the upload file name matches the configuration name of
Subject information check	Check the contents of the uploaded file, including whether the account ID belongs to the current role. Whether the report period is correct. Whether the account data meets the specification, such as whether the account data should not appear in English, etc.
File transfer	Upload files from client to server
File deletion	Delete unnecessary files
File data storage	Put the data of the file uploaded to the server into the database, and delete the file on the server after the database is successfully put into the database

Based on the data in Table 1, the check file name method is further called to check the file name. The naming rule of upload file name is: GF +6 digits of institution +4 or 5 digits of system name +1 digit of report period (M/Q/h/y, etc.) +6 digits of report period + generation date (8 digits). TXT file [9]. After passing, the file data should be checked, and the file data should be put into storage after checking the data items. So as to effectively ensure the effect of data processing, system efficiency.

2.3 Realization of Human Resource Social Insurance Data Reporting in Different Places

Through the detailed investigation and analysis of the social security fund data acquisition system, the functional requirements of the product are described in detail. This section describes the functional requirements of each module and its subordinate function points from the perspective of business function according to the order of function module structure [10]. Data conversion is one of the main modules of social security fund data acquisition system. Its main subordinate functions include conversion scheme management, setting account correspondence, data conversion and data deletion. Two disequilibrium measure indexes are introduced. Based on the research results of regional financial differences, the Gini coefficient formula (1) reflecting the differences in the history and current situation of China's regional insurance development is

$$\mathrm{GiNi} = \frac{2}{n^2 u_y} \sum_i i y_i - \frac{n+1}{n} \tag{1}$$

In formula (1), n represents the total number of insurance regions in China, which can be divided into 3 or 31 according to different research levels. y_i is the number of the i-th regional indicators after ranking according to the indicators from high to low, and u_y is the mean value of the selected indicators. In the process of specific calculation, the difference degree of relevant indicators is based on different regions or different income classes. The calculation formulas of Theil index and logarithmic mean deviation are listed in (2) and (3).

$$\mathrm{GE}_0 = \frac{1}{N}\sum_i LN\left(\frac{u}{y_i}\right) \tag{2}$$

$$\mathrm{GE}_1 = \frac{1}{N}\sum_i \frac{y_i}{u} LN\left(\frac{y_i}{u_i}\right) \tag{3}$$

In the formulas (2) and (3), N is the total number of regions, u is the mean value of regional extraction index, and y_i is the measurement index of regional insurance development level of the i-th region. Logarithmic mean deviation and Theil index. On the basis of the above formula, the operation process is shown in formula (4).

$$\begin{aligned}
\mathrm{GE}_0 &= \frac{1}{N}\sum_i LN\left(\frac{u}{y_i}\right) \\
&= \frac{1}{n}\sum_{t=1}^{k}\sum_{i\in N_t} LN\frac{u}{y_i} \\
&= \sum_{t=1}^{k}\frac{n_t}{n}\frac{1}{n_t}\sum_{i\in N_t} LN\frac{u_t}{y_i} + \frac{1}{n}\sum_{t=1}^{k}\sum_{i\in N_t} LN\frac{u}{u_t} \\
&= \sum_{t=1}^{k} Z_t GE_0\left(y^t\right) + \sum_{t=1}^{k} Z_t LN\frac{u}{u_t}
\end{aligned} \tag{4}$$

On the basis of formula (4), the operation of formula (5) is carried out.

$$\begin{aligned}
\mathrm{GE}_1 &= \frac{1}{N}\sum_i \frac{y_i}{u} LN\left(\frac{y_i}{u_i}\right) \\
&= \frac{1}{n}\sum_{t=1}^{k}\sum_{i\in N_l} \frac{y_i}{u} LN\frac{y_i}{u} \\
&= \sum_{t=1}^{k}\frac{n_t}{n}\frac{u_t}{u}\frac{1}{n_t}\sum_{i\in N_i}\frac{y_i}{u_t} LN\frac{y_i}{u_t} + \frac{1}{n}\sum_{t=1}^{k}\sum_{i\in N_t}\frac{u_t}{u} LN\frac{u_t}{u} \\
&= \sum_{i=1}^{k} Z_t\frac{u_t}{u} GE_1\left(y^t\right) + \sum_{t=1}^{k} Z_t\frac{u_t}{u} LN\frac{u}{u_t} \\
&= \mathrm{I} + \mathrm{O}
\end{aligned} \tag{5}$$

Due to the different sensitivities of the indicators, the mean logarithmic deviation and Theil index are very sensitive to the changes of the indicators of the ethnic group with the lowest proportion and the highest proportion, respectively. Thus, it can make up for the deficiency the Gini coefficient is sensitive to the changes of indicators in the medium level, so that we can more comprehensively analyze the degree of regional differentiation of China's insurance. Each city formulates its own accounting subjects according to its specific situation. The accounting subjects of each city are different, which needs to be transformed into a unified accounting subject. In the data deletion function, the data management organization users delete the problematic and incomplete data submitted by the lower level data center users or municipal agency users through this function.

The amount of data transmission between reporting task r_{i} and information resource r_j is recorded as $F\left(r_{\mathrm{i}}, r_j\right)$. after obtaining all resources, task r_{i} can perform the communication function. After the completion of the task, all resources are released to become available resources. The execution time of reporting task r_{i} on all resources E_{k} is $T_{\mathrm{c}}(r_{\mathrm{i}}, E_{\mathrm{k}})$, and the traffic process mainly depends on the calculation time and the transmission time between resources. For the part where calculation and reporting cannot overlap, $T_{\mathrm{c}}(r_{\mathrm{i}}, E_{\mathrm{k}})$ can be defined as:

$$T_{\mathrm{c}}(r_{\mathrm{i}}, E_{\mathrm{k}}) = R_c(r_{\mathrm{i}}, E_{\mathrm{k}}) + \sum_{r_j} \left(F\left(r_{\mathrm{i}}, r_j\right) \times Q_c\left(r_{j\mathrm{i}}, E_{\mathrm{k}}\right)\right) \tag{6}$$

In formula (6): $\sum_{r_j} \left(F\left(r_{\mathrm{i}}, r_j\right) \times Q_c\left(r_{j\mathrm{i}}, E_{\mathrm{k}}\right)\right)$ is the total time of resource transmission; $R_c(r_{\mathrm{i}}, E_{\mathrm{k}})$ is the time of calculation. The average execution time of reporting task r_{i} is shown in formula (7).

$$\overline{T}_{\mathrm{c}}(r_{\mathrm{i}}) = \sum_{m=1}^{n} T_{\mathrm{c}}(r_{\mathrm{i}}, E_{\mathrm{k}}) / n \tag{7}$$

If the escalation task r_{i} is mapped to all resources E_{k}, then make $T_{\mathrm{a}}(r_{\mathrm{i}}, E_{\mathrm{k}})$ and $T_{\mathrm{b}}(r_{\mathrm{i}}, E_{\mathrm{k}})$ the start and end time of the task execution, and $T_{\mathrm{a}}(r_{\mathrm{i}}, E_{\mathrm{k}})$ can be expressed as formula (8).

$$T_{\mathrm{a}}(r_{\mathrm{i}}, E_{\mathrm{k}}) = \max\left\{E_{\mathrm{a}}(E_{\mathrm{k}}), D_{pred}(r_{\mathrm{i}}, E_{\mathrm{k}})\right\} \tag{8}$$

In formula (8), $D_{pred}(r_{\mathrm{i}}, E_{\mathrm{k}})$ is the time of all data obtained by the reporting task from the start task r_{i} set. And $T_{\mathrm{b}}(r_{\mathrm{i}}, E_{\mathrm{k}})$ can be expressed as formula (9).

$$T_{\mathrm{b}}(r_{\mathrm{i}}, E_{\mathrm{k}}) = T_{\mathrm{a}}(r_{\mathrm{i}}, E_{\mathrm{k}}) + T_{\mathrm{c}}(r_{\mathrm{i}}, E_{\mathrm{k}}) \tag{9}$$

According to the start time and end time of the reporting task, the task is mapped to the resource target, and the matching time between the task and the computing resource is determined. According to the current situation reported, parallel distribution scheme implementation. According to the requirement analysis mentioned above, entities are summarized in each subsystem, and the relationship between entities is determined. Based on this, the system structure function points are described, as shown in Table 2. Submit the verified data to the superior company. If the superior company returns the submitted data, the system will display "re submit data". The returned data must go

Table 2. System structure function point description

Function point	Describe
Entrance	Is the access to the system
Legitimacy check	Determine whether the user exists
Safety inspection	Judge whether the user password is correct
Authority check	Check the permission that the user has

through the verification step again before it can be submitted again. Note: once the data is submitted to the superior company, it cannot be submitted again, and the system will automatically control the data after it is submitted, and it is not allowed to upload and modify the data. If errors are found after submitting to the superior company, the data status of the company shall be released to the "not submitted" state through the operation of the users of the superior company at the provincial level, so as to re-operate. See Table 3 for the function points of data reporting module.

Table 3. System module function point description

Function point	Describe
Data verification status check	Check whether the data of the organization passes the verification
Data reporting status check	Check whether the data of the organization is reported
Data summary status check	Check whether the data is "full summary" or "temporary summary"
Data number check	Check for data

Select the data item to be summarized in the interface of displaying summary status and summary records, click summary and select the subordinate, and STS sum action will process the data summary request of provincial and head office. After judging the summary type, accumulate the submitted data item by item and return to the summary result display page. So as to effectively guarantee the operation effect of the system and improve the operation efficiency of the system.

3 Analysis of Experimental Results

The configuration requirements for the experimental platform to be built are: a database server with a simulated minicomputer, 8 × 4.0 GB, 64-bit dual-core or above. Memory ≥ 128 GB, support memory mirroring and protection. 4 pieces of 8 TB capacity hard drives, hard drive speed ≥ 16 000 r/min, SCSI8068, support Raid-0, 1, 5, 6, 10. Application-side server 1 emulation minicomputer, processor, 4 GB, 36 MB cache. ≥32 GB fully buffered DDR4 memory, which can be expanded to 128 GB, supports memory mirroring and memory protection. Two hard drives with 4 TB capacity, rotating

speed ≥ 20000 r/min. Support Raid-0, 1, 5, 6, 10. In the above-mentioned experimental environment, the method of literature [2] is used as a traditional comparison method, and the experiment is compared with the text method. All the test data are collected into the virtual branch; the business system is being tested, and the data of new order, renewal, preservation, surrender and insurance payment are generated; the system imports the relevant data generated by the business system. System input capital investment, in the reporting system, new institutions: branch, number of documents: 440300. Operators of new branch company. To generate business system report data, execute business system "data report" program; click "generate business report data" to output file. In the system, the corresponding responsibilities are used to execute the "operation report", select the accounting period and rewrite section "40300…", and the report output is "report submitted by CIRC (balance sheet)", "report submitted by CIRC (monthly profit and loss statement)" and "report submitted by CIRC (cash flow statement)". The report output is processed in Excel, and the specific steps are shown in Table 4.

Table. 4 Test process

Serial number	Test Case Specification
1	1 → 9 → 13 → 14 → 15 → 16 . . . Business system data is reported separately
2	10 → 11 → 12 → 13 → 14 → 15 → 16 . . . Financial system data is reported separately
3	1 → 8 → 11 → 12 → 13 → 14 → 15 → 16 . . . Report through the interface

In order to analyze the data upload processing performance of this traditional system and this system more intuitively, the paper compares and analyzes the efficiency of file processing, and the results are shown in Fig. 4.

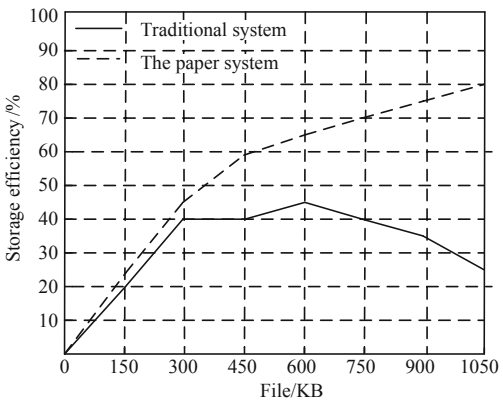

Fig. 4. Comparison results of data processing performance of two systems

Compare the time of file reporting between the traditional system and the system in this paper, and verify the stability of the system in this paper. The results are shown in Fig. 5.

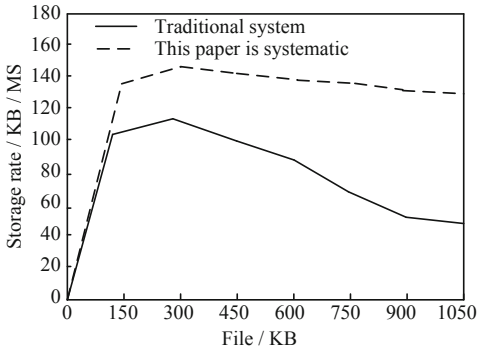

Fig. 5. Comparison results of reporting rate between two systems

Based on the above test results, it can be seen that the human resources social insurance data remote reporting system based on big data technology has high timeliness and stability in the practical application process, which fully meets the research requirements.

4 Conclusion

Based on big data technology, the paper designs the remote reporting system of human resources social insurance data. The paper analyzes the mechanism of the spatial differentiation of insurance industry from the measurement level, expands the research on the influencing factors of the differentiation of insurance industry from analytical description to measurement level, and improves the theory of insurance space to some extent. The results of the application of the remote reporting system of human resources social insurance data based on big data technology are proved by experiments. The research shows that the system of this design has a theoretical research and empirical test on the evolution trend of the spatial differentiation of insurance industry in China, and discusses and forecasts the trend of the differentiation of insurance industry in a long period.

References

1. Dhieb, N., Ghazzai, H., Besbes, H., et al.: A secure AI-driven architecture for automated insurance systems: fraud detection and risk measurement. IEEE Access **8**, 58546–58558 (2020)
2. Wu, J., Wang, J., Liu, Y.: Design and research of insurance survey claims system based on Big Data analysis. In: 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), pp. 211–214. IEEE (2019)
3. Li, Z., Xiao, Z., Xu, Q., et al.: Blockchain and IoT data analytics for fine-grained transportation insurance. In: 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS), pp. 1022–1027. IEEE (2018)

4. Bo, Z.A., Cdb, A., Tz, C., et al.: An IGAP-RBFNN-based secondary control strategy for islanded microgrid-cyber physical system considering data uploading interruption problem. Neurocomputing **397**(6), 422–437 (2020)
5. Liu, D., Xu, Y., Xu, Y., et al.: Opportunistic data collection in cognitive wireless sensor networks: air-ground collaborative online planning. IEEE Internet Things J. **7**, 8837–8851 (2020)
6. Usman, M., Jan, M.A., Jolfaei, A., et al.: A distributed and anonymous data collection framework based on multi-level edge computing architecture. IEEE Trans. Ind. Inf. **16**, 6114–6123 (2019)
7. She, W., Gu, Z.H., Lyu, X.K., et al.: Homomorphic consortium blockchain for smart home system sensitive data privacy preserving. IEEE Access **7**(6), 62058–62070 (2019)
8. Liu, S., Fu, W., He, L., et al.: Distribution of primary additional errors in fractal encoding method. Multimedia Tools Appl. **76**(4), 5787–5802 (2017)
9. Honeycutt, W.T., Kim, T., Ley, M.T., et al.: Sensor array for wireless remote monitoring of carbon dioxide and methane near carbon sequestration and oil recovery sites. RSC Adv. **11**(12), 6972–6984 (2021)
10. Liu, S., Bai, W., Liu, G., et al.: Parallel fractal compression method for big video data. Complexity **2018**, 1–16 (2018)