

Ewidencja PPK

Projekt techniczny interfejsu, formatu i zakresu wymiany danych z Państwową Inspekcją Pracy

v. 1.0



Spis treści

| | | |
|-------|---------------------------------|---|
| 1 | Wstęp..... | 3 |
| 2 | Założenia ogólne..... | 4 |
| 2.1 | Format plików..... | 4 |
| 2.2 | Format danych..... | 4 |
| 2.3 | Nagłówek pliku | 4 |
| 2.4 | Nazwy plików..... | 5 |
| 2.5 | Konto na serwerze SFTP | 6 |
| 2.6 | Dostarczenie pliku na SFTP..... | 6 |
| 3 | Pliki danych..... | 8 |
| 3.1 | Umowy o Zarządzanie..... | 8 |
| 3.1.1 | Plik wyjściowy | 8 |



1 Wstęp

Dokument opisuje interfejs plikowej wymiany danych pomiędzy Systemem Ewidencji PPK (dalej System) a Państwową Inspekcją Pracy (dalej PIP). Wymiana danych następuje w formie plików raportów przekazywanych z Systemu do PIP. Pliki są przekazywane z użyciem SFTP.



2 Założenia ogólne

Niniejszy rozdział opisuje ogólne założenia dotyczące plików raportów (format plików, format danych, nazewnictwo) i sposobu ich przekazywanie do/z Systemu. Dla każdego raportu wysłanego do Systemu zostanie zwrócony raport z informacją zwrotną o statusie przetwarzania.

2.1 Format plików

- Pliki są w postaci plików CSV (Comma Separated Values).
- Plik jest zmiennopozycyjny. W wszystkich wierszach oprócz nagłówka będzie tyle samo kolumn. Każda kolumna może mieć inną maksymalną długość. Wartości w danej kolumnie mogą mieć różne długości (brak dopełniania do maksymalnej długości).
- Separator pól to znak pipe - „|”.
- Separator w polach z listą wartości (wielowierszowe) to hash - "#". Np. w lista Id EPPK Umów o Zarządzanie: "ID-1#ID-2#ID-3".
- Separator dziesiętnej liczb to kropka - ".".
- Separator daty to myślnik - "-".
- Brak kwalifikacji tekstu w polach cudzysłowem.
- Każdy plik zaczyna się od linii nagłówka. Zawartość nagłówka została szczegółowo opisana w rozdziale **Nagłówek pliku**.
- Kodowanie znaków to UTF-8 (bez BOM).
- Znak końca linii to '\n' (Unixowy znak końca linii).
- Każda linia (oprócz pierwszej) rozpoczyna się od liczby porządkowej rekordu (numer kolejny zaczynając od 1).
- Nazwa pliku musi być unikalna. Drugi plik o tej samej nazwie zostanie odrzucony jako duplikat.

2.2 Format danych

- Pola zawierające datę są w formacie yyyy-MM-dd. Np. 2019-12-31 lub 2021-01-01.
- Pola zawierające timestamp są w formacie yyyy-MM-dd HH:mm:ss. Np. 2019-12-31 23:59:59 lub 2021-01-01 12:03:04.

2.3 Nagłówek pliku

Wszystkie pliki w pierwszej linii powinny posiadać linię z następującymi polami:

- Wersja formatu pliku (obecnie 1.0); pole ma umożliwić w przyszłości płynne przechodzenie na nowe wersje formatów raportów (dodawanie/usuwanie pól). Wersja jest globalna dla całego interfejsu/API. Zmiana w jednym pliku powoduje, że zmienia się również wersja we wszystkich pozostałych plikach.



- Liczba rekordów w pliku (nie licząc pierwszej linii)
- Powyższy opis pierwszej linii dotyczy również raportów zwrotnych generowanych przez System.

Przykład nagłówka, z którego wynika, że plik jest zgodny z wersją 1.0 i zawiera 536 wierszy:

```
1.0|536
```

2.4 Nazwy plików

Nazwy plików raportów wysyłanych przez System do PIP mają format:
`{typ_raportu}_{kwartał}_{data}_{wersja}.csv`

Gdzie:

- {typ_raportu} to jeden z napisów:
 - „uoz” - dla pliku o strukturze „Umowy o Zarządzenie”
- {kwartał} to:
 - dla typu raportu „uoz” jest to kwartał, którego dotyczy raport, w formacie YYYYKN, gdzie YYYY to rok, K to napis (od słowa Kwartał), a N to numer kwartału;
- {data} to dzień wygenerowania raportu formacie YYYYMMDD
- {wersja} to wersja raportu za dany dzień; za zwyczaj powinno być tu 1, aczkolwiek w przypadku przysłania błędnego raportu można będzie dostarczyć nowy raport podając inną nową wersję pliku.

Przykład:

- uoz_2019K4_20201024_1.csv



2.5 Konto na serwerze SFTP

PFR dostarczy nazwę użytkownika do serwera SFTP dla PIP. PIP generuje klucze RSA o długości 4096 bitów (ssh-keygen) i przekazuje do PFR klucz publiczny, który zostanie dodany do `authorized_keys` na SFTP. Uwierzytelnianie do SFTP będzie realizowane poprzez nazwę użytkownika i klucze. Brak uwierzytelnienia poprzez nazwę użytkownika i hasło.

Wymagane jest dostarczenie przez PIP adresu IP serwera z którego PIP będzie łączył się do serwera SFTP (będzie wykona restrykcja adresów IP). Logowanie PIP będzie możliwe wyłącznie ze zdefiniowanego adresu IP. Połączenia z innych adresów IP będą odrzucane.

Kanał komunikacji będzie szyfrowany (SFTP), a przesyłane pliki nie będzie dodatkowo szyfrowany.

System będzie umieszczał pliki raportów w katalogu „wyjscie”. W przypadku gdyby zasłała potrzeba przekazywania raportów z PIP do Systemu, to PIP będzie przysyłał swoje raporty do katalogu "wejscie".

2.6 Dostarczenie pliku na SFTP

Komunikacja pomiędzy PIP a Systemem będzie się odbywać w sposób symetryczny, czyli zarówno odbieranie jak i wysyłanie plików będzie działać zgodnie z opisanym niżej schematem.

Założenia dotyczące wymiany plikowej:

- Strona, która wysyła plik generuje skrót pierwotnego pliku raportu (pliku csv)
- Strona, która wysyła plik zobowiązana jest do wykonania kompresji pliku raportu
- Strona, która wysyła plik na czas transmisji raportu tworzy plik *.blokada, aby strona obierająca mogła pominąć ten plik w przetwarzaniu
- Strona obierająca usuwa pliki z serwera SFTP, po jego skopiowaniu

Algorytm przygotowania oraz wysłania pliku z raportem na serwer SFTP dla pliku raportu o nazwie `uoz_2019K4_20200201_1.csv`:

1. System generuje plik w formacie CSV.
Przykład nazwy pliku: `uoz_2019K4_20200201_1.csv`
2. System generuje plik z sumą kontrolną dla raportu przy użyciu sha512. Wynikowy plik musi mieć taką samą nazwę jak plik CSV z dodaną końcówką ".sha512".
Przykład nazwy pliku: `uoz_2019K4_20200201_1.csv.sha512`
3. System wykonuje kompresję pliku CSV z wykorzystaniem formatu GZIP. Wynikowy plik musi mieć taką samą nazwę jak plik CSV z dodaną końcówką ".gz".
Przykład nazwy pliku: `uoz_2019K4_20200201_1.csv.gz`
4. System w pierwszej kolejności tworzy lub kopiuje na serwer SFTP do katalogu „wyjscie” pusty plik blokady, sygnalizując że trwa operacja wystawienia plików. Plik powinien mieć nazwę zgodną z plikiem CSV z dodaną końcówką ".blokada".
Przykład nazwy pliku: `uoz_2019K4_20200201_1.csv.blokada`
5. Następnie System kopiuje na SFTP właściwy plik raportu oraz plik z sumą kontrolną.



6. Po zakończeniu kopiowania plików wymienionych w punkcie 5, System usuwa plik blokady. Od tego momentu PIP może rozpocząć przetwarzanie pliku CSV.

Ten sam algorytm będzie stosować PIP do ewentualnego przekazywania swoich raportów do PFR.

Podsumowując, na serwerze SFTP w momencie wysyłania plików będą maksymalnie 3 pliki:

- uoz_2019K4_20200201_1.csv.blokada
- uoz_2019K4_20200201_1.csv.gz
- uoz_2019K4_20200201_1.csv.sha512

Po zakończeniu wysyłania plików na serwer SFTP będą dokładnie 2 pliki:

- uoz_2019K4_20200201_1.csv.gz
- uoz_2019K4_20200201_1.csv.sha512

System i PIP będą ignorować pliki dla których:

- nie został usunięty plik „blokada”
- brakuje pliku „.gz” lub „.sha512”
- plik ma niepoprawną nazwę lub rozszerzenie

Zakłada się, że po pobraniu plików z katalogu „wyjscie” PIP będzie usuwać pobrane pliki (w tym plik .sha512). Symetrycznie, System będzie usuwać pobrane pliki z katalogu „wejscie” (w tym plik .sha512).



3 Pliki danych

Niniejszy rozdział opisuje zawartość poszczególnych plików.

3.1 Umowy o Zarządzanie

Raport zawiera listę pracodawców przechowywaną w Systemie wraz z informacją o zawartych umowach o zarządzanie (dalej UOZ lub uoz).

3.1.1 Plik wyjściowy

| Nazwa pola | Wymagalność | Min | Max | Walidacje (regex, dozwolone znaki, słownik wartości, suma kontrolna) | Uwagi |
|---|-------------|-----|------|--|--|
| Liczba porządkowa rekordu (numer kolejny zaczynając od 1) | T | 1 | 10 | Liczba. Niedozwolone są zera wiodące. Wartość w pierwszym wierszu wynosi 1, w każdym kolejnym n+1. | |
| NIP | T | 0 | 10 | NIP pracodawcy. | |
| REGON | T | 9 | 14 | REGON pracodawcy (Centrali lub oddziału) | |
| Identyfikator wewnętrzny ZUS | N | 0 | 11 | | Jest to identyfikator wewnątrz płatnika w bazie ZUS. System dopasowuje zgłaszane przez PZIF UOZ do pracodawców raportowanych przez ZUS po NIP, stąd możliwe są przypadki, kiedy w Systemie będzie istnieć UOZ bez identyfikatora wewnętrznego ZUS. |
| Nazwa pracodawcy | T | 0 | 2000 | | |
| Nazwisko | N | 0 | 160 | | Nazwisko płatnika – pole wypełnianie w przypadku płatnika będącego osobą fizyczną. |
| Imię | N | 0 | 60 | | Imię płatnika – pole wypełnianie w przypadku płatnika będącego |



| Nazwa pola | Wymagalność | Min | Max | Walidacje (regex, dozwolone znaki, słownik wartości, suma kontrolna) | Uwagi |
|-------------|-------------|-----|-----|--|--|
| | | | | | osobą fizyczną. |
| Zawarte UOZ | N | 0 | 512 | Lista zawartych UOZ w formacie: EPPKID_UOZ:data Gdzie: <ul style="list-style-type: none"> EPPKID_UOZ jest identyfikatorem UOZ nadanym przez System dwukropek jest separatorem pomiędzy EPPKID_UOZ a datą data jest w formacie yyyy-MM-dd. Poszczególne rekordy w liście są rozdzielone znakiem #. | Lista zawartych i wciąż aktywnych Umów o Zarządzanie. System dopuszcza istnienie 2-ch aktywnych umów o zarządzanie dla jednego pracodawcy na potrzeby obsługi zmiany UOZ przez pracodawcę. Jeśli różnica pomiędzy poszczególnymi datami jest większa niż 2 tygodnie, to taki pracodawca prawdopodobnie nie dopełnił obowiązku wypowiedzenia umowy z jednym z PZIF. |