

OTWIERACZ.GOV/OPENER.GOV Service - job description

Problem

The widespread use of electronic documents for communication with the public administration combined with the transfer of tax obligations to remote channels caused that the National Revenue Administration (NRA) is also submitted files whose extension does not correspond to the file format, and through incorrect channels, although they must be submitted in a manner specified by law and by maintaining the required format.

Setting the file format, its visualization and further handling require to use several different specialist tools, which is time consuming and costly for the organization.

Many of the applications that are available on the market require to send a file to outside a trusted network, which is unacceptable while handling documents that are subject to any legally protected secrets.

The diversified level of digitization of the society and of their awareness of using electronic signatures, as well as the number of tools that are available on the market, makes it increasingly more difficult to verify and read the submitted files. However, it is of crucial importance for the processes handled by the NRA that the contents of the submitted documents are easily accessible, while the signature is valid and verifiable, e.g. for the signature date.

Some of the electronic documents that are processed by the NRA must be anonymized, which is a time-consuming process for which automation is recommended.

Idea

Delivering a solution to identify the actual format of the submitted electronic irrespective of the manner in which they are delivered, and to verify its correctness and display information about the electronic signature or its absence. For XML documents, it is also necessary to visualize them. An additional advantage of the solution will be that the solution will be equipped it with the 'anonymization module'.

The OTWIERACZ.GOV service solution analyses, opens and visualizes different file formats without using or sharing them with external applications, excluding those specified in the task description. The results of the verification will be presented in the form of a report.

Basic assumptions

1. Technical conditions:

- The service may be by the user in an environment that meets the following minimum requirements:
 - Windows 8 and higher,
 - 4GB RAM,
 - 1024 x 768 resolution;
- The web service is available in a minimum of three market leading browsers (Edge, Chrome, Firefox);
- The service does not require more than 20 GB of free disk space to be launched;
- The service supports electronic documents up to 500 MB. Supporting documents larger than 500 MB will be scored additionally;
- The contents of an electronic document may not be sent to third parties at any stage (including to server of a service provider or a service component), with the exception of the signature verification;
- The service being designed is desktop or web-based – for desktop solutions it is required to update on workstations;
- The service is based on open source solutions and in particular innovative solutions (not currently offered on the market as complete solutions);
- The service is functional, intuitive and user friendly, in particular it provides the user with display ergonomics and interfaces readability.

2. Supported formats

- The service recognizes:
 - ✓ As a minimum: txt, .rtf, .pdf, .xps, .odt, .ods, .odp, .doc, .xls, .ppt, .docx, .xlsx, .pptx, .csv, .jpg, .tif, .geotif, .png, .svg, .wav, .mp3, .avi, .mpg, .mp4, .ogg, .ogv, .zip, .tar, .gz, .7Z, .html, .xhtml, .css, .xml, .xsd, .gml, .rng, .xsl, .xslt, TSL, XMLsig, XAdES, PAdES, CAdES, ASIC, XMLenc;
 - ✓ the format of the documents submitted via the ePUAP platform, and provides for a friendly form of visualization, allowing the presentation of the content according to the pattern transform (e.g. without schema elements);
 - ✓ other supported formats through the verification of the compatibility of the actual format with the file extension (e.g. recognizing the document in .pdf format even if its extension is changed to .xml) and indicating the name of the supporting application.
- For electronic documents not having an extension, their format is recognized on the basis of the internal structure of the file.
- The files being archives, e.g. .zip, are not unpacked and verified for their contents, while the service recognizes such format and treats it as a single file.

3. Handling of .xml documents

The discussed problem largely concerns the documents in .xml format that are created according to the developed schema (logical structure). In most cases, both the schema and visualization are published in the Central Register of Electronic Document Templates (<https://epuap.gov.pl/wps/portal/strefa-official/other-systems/crwde>).

- The service also recognizes logical structures that are specific for the NRA, including the e-financial statements, e-tax returns, Uniform Control Files (JPK), etc. The service examines the compatibility of the structures with the schema provided in the Public Information Bulletin (for the financial statements - <https://www.gov.pl/web/kas/struktury-e-reporting>).
- For .xml documents, the service validates (with XSD – automatically searching for an appropriate XSD – on Central Register of Electronic Document Templates, podatki.gov.pl, Public Information Bulletin, Internet) and places a link to the schema source (Central Register of Electronic Document Templates, Public Information Bulletin) in the document verification report. In addition, it allows to generate visualizations based on XSL from the Central Register of Electronic Document Templates, or a simplified visualization in case XSL is missing, e.g. in a table. Furthermore, the export of visualizations to .pdf files will be scored.
- The service enables to extract binary elements included in the .xml file.

4. Signature information

The electronic signature means both a document signed and not signed with an electronic signature. For the documents signed with an electronic signature, information on the result of the signature verification from the external system should be made available (full report – all possible information from the application: <https://ec.europa.eu/cefdigital/DSS/webapp-demo/validation>). For the documents not signed with an electronic signature, information about that the signature is missing. For the signature verification, the service uses API to connect to the electronic signature verification tool. Additional scores will be given for extracting a file not bearing a signature from the document signed with a surrounding electronic signature, including the possibility to save such a document on the local disk under any name.

5. **The service ensures the invariability of the original electronic document at any stage of its processing and verification.**
6. **Anonymization applies only to single .pdf files.**
 - The service produces a copy of the original document and hides the contents specified by the user without interfering with the original.
 - Anonymization takes place without sending data to the servers of the vendor - the processing must take place locally because the data being processed are subject to tax secret or the protection of personal data.

The input to the task consists of the files prepared by the Ministry of Finance with various extensions. The file to be read is verified, visualized and anonymized.

The output consists of:

1. GUI for the user (the employee of the NRA) where the service allows to add a file for verification. Additional scores will be given for allowing a simultaneous verification of multiple files and for sharing a list of the recently opened files.
2. For .xml files in GUI, visualization is required at least on the basis of XSL from the Central Register of Electronic Document Templates and the tax portal, or a simplified visualization in case XSL is missing, e.g. in a table, with the result of exporting the visualized file to the .pdf format.
3. Verification report:
The verification result (report) is available for each file.
Additional scores are given for generating a bulk report for all or selected files, containing separate records for each electronic document being subject to verification.
The verification report is linked to the file being verified, available from the GUI level and printable. It is possible to save the report in an editable form, on the user's drive in the desired location.

Range of the reported information:

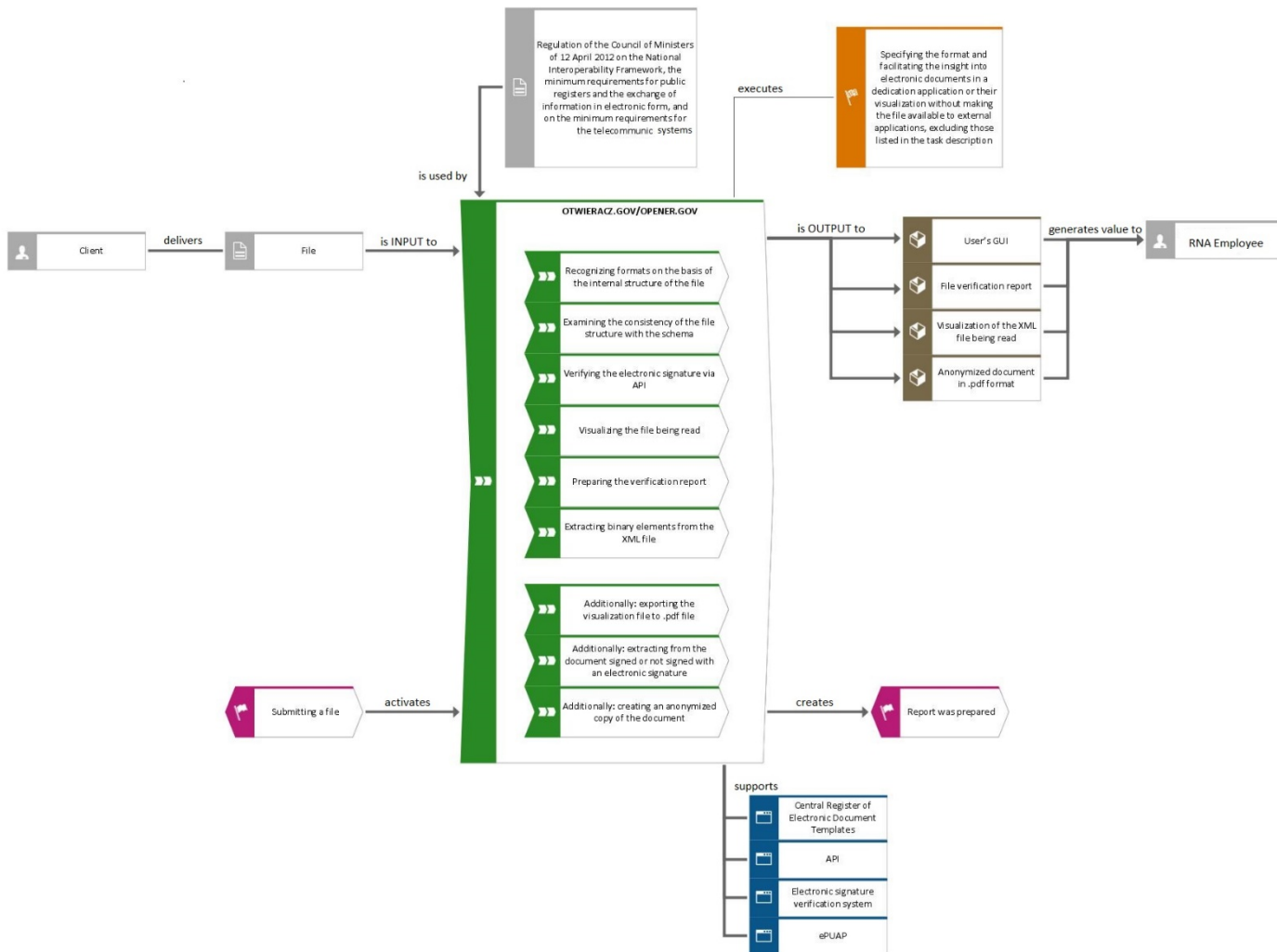
- ✓ Date of verification,
 - ✓ Document name,
 - ✓ Size,
 - ✓ Electronic signature data, including the verification result including the data of the electronic signature verification indicating all the data received from API,
 - ✓ Compatibility of the format with the extension by returning both the extension and the actual format,
 - ✓ Indication of the name of the application that supports the file with an extension,
 - ✓ For the bulk report, an unambiguous link of the item of the report with the file being verified (optional),
 - ✓ For .xml files:
 - Compatibility of .xml with XSD with the schema name or information about non-compliance. Additional scores will be given for a report listing errors for the electronic document not being consistent with the schema,
 - Direct link to the source of the schema,
 - List of binary items containing links to download the attached files.
4. Anonymization module (additionally scored):
A special advantage of the service for .pdf files will be the possibility to make them anonymous. The user's GUI contains a function that enables to type any string of characters that are hidden. The service selects and obscures the specified areas. The user's approval of the anonymized document prevents its

modification and insight into the anonymized contents. The hidden contents may not be removed with any application.

Development of the solution

The solution is planned to ultimately cover all the electronic documents the organizational units of the NRA receive. Therefore, it is necessary to ensure that it can also handle other formats, excluding those listed in the description, while assuming a minor interference in the code. It is also desirable that it is integrated with systems for processing electronic documents and document flow channels.

The National Revenue Administration plans to place the service in the Catalogue of Public Services of the NRA in future so that it may be used by taxable persons and third parties, including the bodies of the public administration.



Submit your project proposal with the following files:

1. PROJECT NAME _ Project description.ZIP [mandatory]:

- Description of the project in English or Polish,
- Max. 500 words.

2. PROJECT NAME_PRESENTATION. ZIP [mandatory]:

- Folder containing information, materials, descriptions, screenshots, presentations, links to films and other visualizations of your project, prototype interface, etc.,
- Presentation (in English or Polish, .pdf format, max. 10 slides).

3. PROJECT NAME _ MANUAL.ZIP: (mandatory)

Video URL (mandatory)

- Prepare a video about the project or a video where you will talk about your project,
- Max. 60 sec.,
- in English or Polish,

Link to Project Demo (mandatory)

- Remember to transfer the login data to the JURY,

URL repository (mandatory)

- The whole in one repository (if your project consists of several parts, modules, make sure you put them in separate folders),
- Ensure that they may be verified online,
- Repository containing the solution design or a link thereto, including the step-by-step description of how to launch the solution in Windows or Linux,
- Recommended process automation script,
- List of frameworks, compilers, virtual machines, containers, launch environment, hardware requirements,
- List of additional tools, libraries and additional programs (e.g. databases), which the service uses,
- Launch manual – what needs to be done to launch the project,
- Operating manual – for the administrator,
- Basic, shortened user manual – how to use the tool,
- Description of architecture.

4. OTHER Files.ZIP (mandatory):

- Generated verification report for the delivered files,
- Result of the visualization,
- In case the anonymization module is prepared – the anonymized document in .pdf format.

Usługa OTWIERACZ.GOV/OPENER.GOV – opis zadania

Problem:

Powszechne zastosowanie dokumentu elektronicznego w komunikacji z administracją publiczną połączone z przeniesieniem obowiązków podatkowych do kanałów zdalnych spowodowało, że do Krajowej Administracji Skarbowej (KAS) wpływają również pliki z rozszerzeniem nieodpowiadającym formatowi pliku oraz niewłaściwymi kanałami, pomimo obowiązku wysyłania ich w zgodny z prawem sposób i w określonym formacie.

Ustalenie formatu pliku, jego wizualizacja i dalsza obsługa wymaga korzystania z kilku różnych, specjalistycznych narzędzi, co jest pracochłonne i kosztowne dla organizacji.

Dostępne na rynku aplikacje niejednokrotnie wymagają wysłania pliku poza zaufaną sieć, co w przypadku obsługi dokumentów objętych tajemnicami prawnie chronionymi jest niedopuszczalne.

Zróżnicowany poziom cyfryzacji społeczeństwa i świadomości korzystania z podpisów elektronicznych, a także liczba dostępnych na rynku narzędzi, powoduje, że przesyłane pliki coraz trudniej zweryfikować i odczytać. Natomiast łatwy dostęp do treści przesyłanych dokumentów oraz potwierdzenie prawidłowości podpisu i weryfikacja, np. daty złożenia podpisu, jest kluczowa w procesach KAS.

Część dokumentów elektronicznych przetwarzanych w KAS wymaga anonimizacji co jest procesem czasochłonnym, dla którego zalecana jest automatyzacja.

Pomysł:

Dostarczenie rozwiązania umożliwiającego ustalenie faktycznego formatu przesłanych dokumentów elektronicznych, bez względu na sposób dostarczenia i poprawność ich utworzenia oraz pokazanie informacji o podpisie elektronicznym, bądź jego braku. W przypadku dokumentów w formacie .xml, konieczna jest również ich wizualizacja. Dodatkowym atutem rozwiązania będzie wyposażenie go w „moduł anonimizacji”. Rozwiązanie w formie usługi OTWIERACZ.GOV analizuje różne formaty plików, otwiera je i wizualizuje, bez wykorzystywania lub udostępniania zewnętrznym aplikacjom, innym niż wskazanym w opisie zadania. Wyniki weryfikacji przedstawia w formie raportu.

Założenia podstawowe

1. Warunki techniczne:

- Usługa jest możliwa do uruchomienia przez użytkownika w środowisku spełniającym następujące wymagania minimalne:
 - ☐ Windows 8 i wyższe,
 - ☐ 4 GB RAM,
 - ☐ Rozdzielczość 1024x768
- Usługa web jest dostępna na minimum trzech wiodących na rynku przeglądarkach (Edge, Chrome, Firefox),
- Usługa dla uruchomienia nie wymaga więcej niż 20 GB wolnej przestrzeni na dysku,
- Usługa obsługuje dokumenty elektroniczne o rozmiarze do 500 MB. Obsługa dokumentów większych niż 500 MB będzie dodatkowo punktowana,
- Na żadnym etapie zawartość dokumentu elektronicznego nie może być wysyłana na zewnątrz (w tym np. na serwer dostawcy usługi lub komponentu usługi), za wyjątkiem weryfikacji podpisu,

- Projektowana usługa jest desktopowa albo webowa – dla rozwiązań desktopowych wymagania jest aktualizacja na stacjach roboczych,
- Usługa jest oparta o rozwiązania open source, a w szczególności o rozwiązania innowacyjne (nieoferowane aktualnie na rynku jako kompletne rozwiązania).
- Usługa jest funkcjonalna, intuicyjna i przyjazna, w szczególności zapewnia użytkownikowi ergonomię ekranu prezentującego oraz czytelność interfejsów.

2. Obsługiwane formaty:

- Usługa rozpoznaje:
 - ✓ co najmniej: .txt, .rtf, .pdf, .xps, .odt, .ods, .odp, .doc, .xls, .ppt, .docx, .xlsx, .pptx, .csv, .jpg, .tif, .geotif, .png, .svg, .wav, .mp3, .avi, .mpg, .mp4, .ogg, .ogv, .zip, .tar, .gz, .7Z, .html, .xhtml, .css, .xml, .xsd, .gml, .rng, .xsl, .xslt, TSL, XMLsig, XAdES, PAdeES, CAdeES, ASIC, XMLenc;
 - ✓ format dokumentów przesyłanych za pośrednictwem platformy ePUAP i udostępnia przyjazną formę wizualizacji, umożliwiającą prezentowanie treści zgodnie z transformatą wzoru (np. bez elementów schemy);
 - ✓ pozostałe formaty obsługiwane poprzez weryfikację zgodności faktycznego formatu z rozszerzeniem pliku (np. rozpoznanie dokumentu w formacie .pdf nawet jeśli jego rozszerzenie zmieniono na .xml) i wskazaniem nazwy aplikacji obsługującej.
- Dla dokumentów elektronicznych bez rozszerzenia, rozpoznawanie formatów odbywa się na podstawie struktury wewnętrznej pliku.
- Pliki będące archiwami np. .zip; nie podlegają rozpakowywaniu i weryfikacji pod względem zawartości, usługa rozpoznaje taki format i traktuje jak pojedynczy plik.

3. Obsługa dokumentów w formacie .xml

Opisywany problem w znacznej mierze dotyczy dokumentów w formacie .xml, tworzonych wg opracowanej schemy (struktury logicznej). W większości przypadków zarówno schema, jak i wizualizacja są publikowane w Centralnym Rejestrze Wzorów Dokumentów Elektronicznych (<https://epuap.gov.pl/wps/portal/strefa-urzednika/inne-systemy/crwde>).

- Usługa rozpoznaje również struktury logiczne specyficzne dla KAS, w tym: e-sprawozdania finansowe, e-deklaracje, JPK itp. – usługa bada zgodność struktur ze schemą udostępnioną na BIP (dla sprawozdań finansowych - <https://www.gov.pl/web/kas/struktury-e-sprawozdan>).
- Dla dokumentów w formacie .xml usługa sprawdza poprawność (z XSD – automatyczne wyszukanie odpowiedniego XSD – na CRWDE, podatki.gov.pl, BIP, Internet) oraz umieszcza w raporcie z weryfikacji dokumentu link do źródła schemy (CRWDE, BIP). Ponadto umożliwia wygenerowanie wizualizacji na podstawie XSL z CRWDE albo w razie braku XSL uproszczoną wizualizację, np. w tabeli. Dodatkowo punktowany będzie eksport wizualizacji do pliku w formacie .pdf.
- Usługa umożliwia wyodrębnienie elementów binarnych zawartych w pliku .xml.

4. Informacja o podpisie

Za dokument elektroniczny uważany jest zarówno dokument podpisany podpisem elektronicznym jak i niepodpisany.

W przypadku dokumentów podpisanych podpisem elektronicznym należy udostępnić informację o wyniku weryfikacji podpisu z zewnętrznego systemu (pełny raport – wszystkie możliwe informacje z aplikacji: <https://ec.europa.eu/cefdigital/DSS/webapp-demo/validation>). Dla dokumentów

niepodpisanych - informację o braku podpisu. W zakresie weryfikacji podpisów usługa poprzez API łączy się z narzędziem weryfikującym podpisy elektroniczne.

Dodatkowo punktowane będzie wyodrębnienie z dokumentu podpisanego podpisem elektronicznym otaczającym, pliku bez podpisu, z możliwością zapisania takiego dokumentu na dysku lokalnym pod dowolną nazwą.

5. Usługa zapewnia niezmiennosc oryginalnego dokumentu elektronicznego na każdym etapie przetwarzania i weryfikacji.

6. Anonimizacja dotyczy tylko pojedynczych plików w formacie .pdf.

- Usługa wytwarza kopię oryginalnego dokumentu i zakrywa wskazane przez użytkownika treści bez ingerencji w oryginał.
- Anonimizacja odbywa się bez wysyłania danych na serwery dostawców – przetwarzanie musi się odbywać lokalnie, ponieważ przetwarzane dane są objęte tajemnicą skarbową lub ochroną danych osobowych.

Wejściem do zadania są przygotowane przez Ministerstwo Finansów pliki z różnymi rozszerzeniami. Weryfikacji, wizualizacji i anonimizacji podlega czytany plik.

Wyjściem jest:

1. GUI dla użytkownika (pracownika KAS), w którym usługa umożliwi dodanie pliku do weryfikacji. Dodatkowo punktowanym rozwiązaniem będzie umożliwienie jednoczesnej weryfikacji wielu plików oraz udostępnienie listy ostatnio otwieranych plików.
2. Dla plików .xml w GUI wymagana jest wizualizacja co najmniej na podstawie XSL z CRWDE i portalu podatkowego albo w razie braku XSL uproszczona wizualizacja, np. w tabeli, wraz z wynikiem wyeksportowania zwizualizowanego pliku do formatu .pdf.
3. Raport z weryfikacji:
Wynik weryfikacji (raport) dostępny jest dla każdego pliku.
Dodatkowo punktowane jest wygenerowanie zbiorczego raportu dla wszystkich lub wybranych plików, zawierającego osobne rekordy dla każdego weryfikowanego dokumentu elektronicznego.
Raport weryfikacji jest powiązany z weryfikowanym plikiem, dostępny z poziomu GUI i możliwy do wydrukowania. Istnieje możliwość zapisania raportu w formie edytowalnej na dysku użytkownika w wybranej przez użytkownika lokalizacji.

Zakres raportowanej informacji:

- ✓ Data weryfikacji,
- ✓ Nazwa dokumentu,
- ✓ Rozmiar,
- ✓ Dane o podpisach elektronicznych, w tym wynik weryfikacji obejmujący dane z weryfikacji podpisu elektronicznego ze wskazaniem wszystkich danych otrzymanych z API.
- ✓ Zgodność formatu z rozszerzeniem z podaniem zarówno rozszerzenia jak i faktycznego formatu,
- ✓ Wskazanie nazwy aplikacji obsługującej plik o danym rozszerzeniu,
- ✓ Dla raportu zbiorczego jednoznaczne powiązanie pozycji w raporcie z weryfikowanym plikiem (opcjonalnie),
- ✓ W przypadku plików .xml:

- ☒ Zgodność .xml z XSD z nazwą schemy albo informacja o braku zgodności, dodatkowo punktowany jest raport zawierający listę błędów w przypadku dokumentu elektronicznego niezgodnego ze schemą,
- ☒ Bezpośredni link do źródła schemy,
- ☒ Listę elementów binarnych zawierających linki umożliwiające pobranie załączonych plików.

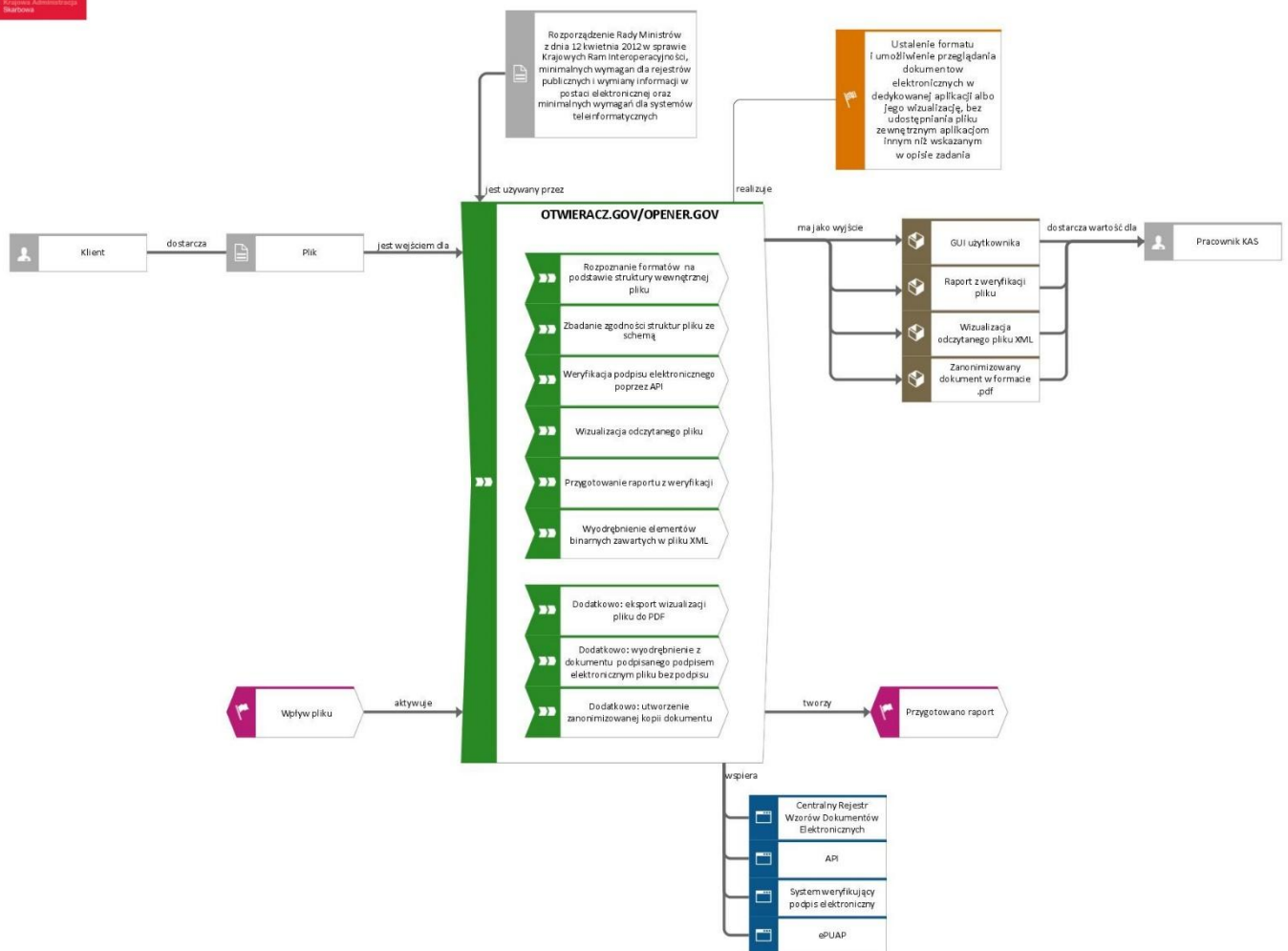
4. Moduł anonimizacji (dodatkowo punktowany):

Szczególnym atutem usługi dla plików w formacie .pdf, będzie możliwość anonimizacji. GUI użytkownika zawiera funkcję umożliwiającą wpisanie dowolnego ciągu znaków podlegającego ukryciu. Usługa zaznacza wskazane obszary i je zasłania. Zatwierdzenie przez użytkownika zanonimizowanego dokumentu uniemożliwia jego modyfikację i wgląd do zanonimizowanych treści. Zakryte treści są nieusuwalne przy użyciu jakiegokolwiek aplikacji.

Rozwój rozwiązania

Planuje się, że rozwiązanie docelowo obejmie swoim zasięgiem wszystkie dokumenty elektroniczne wpływające do jednostek organizacyjnych KAS, dlatego konieczne jest aby mogło przy nieznacznej ingerencji w kod obsłużyć także inne, niż wskazane w opisie, formaty. Wskazana jest również integracja z systemami obsługującymi dokumenty elektroniczne i kanałami wpływu tych dokumentów.

Krajowa Administracja Skarbowa planuje w przyszłości umieścić usługę w Katalogu Usług Publicznych KAS, tak aby korzystali z niej podatnicy i podmioty zewnętrzne, w tym organy administracji publicznej.



Złóż swój projekt wraz z następującymi plikami:

1. PROJECT NAME_ Project description.ZIP [obowiązkowy]:

- Opis projektu w jęz. angielskim lub jęz. Polskim,
- Max 500 słów.

2. PROJECT NAME_ PRESENTATION. ZIP [obowiązkowy]:

- Folder zawierający informacje, materiały, opisy, screenshots, prezentacje, linki do filmów i innych wizualizacji Twojego projektu, interfejs prototypu itp.,
- Prezentacja (w jęz. angielskim lub jęz. polskim, format .pdf, max 10 slajdów).

3. PROJECT NAME_ MANUAL.ZIP : (obowiązkowy)

Video URL (obowiązkowo)

- Przygotuj video dot. projektu lub video gdzie opowiecie o Waszym projekcie,
- Max 60 sec.,
- w jęz. angielskim lub jęz. Polskim,

Link do Demo projektu (obowiązkowy)

- Pamiętaj o przekazaniu danych do logowania dla JURY,

Repozytorium URL (obowiązkowy)

- Całość w jednym repozytorium (jeśli Twój projekt składa się z kilku części, modułów upewnij się, że zamieściłeś je w oddzielnych folderach),
- Zapewnij możliwość sprawdzenia on-line,
- Repozytorium zawierające projekt rozwiązania lub link do niego wraz z opisem krok po kroku jak uruchomić rozwiązanie w Windows lub Linux,
- Zalecany skrypt automatyzujący proces,
- Lista frameworków, kompilatory, maszyny wirtualne, kontenery, środowisko uruchomieniowe, wymagania sprzętowe,
- Lista dodatkowych narzędzi, bibliotek i dodatkowych programów (np. bazy danych), z których korzysta usługa,
- Instrukcja uruchomienia – co należy zrobić, żeby uruchomić projekt,
- Instrukcja obsługi – dla administratora,
- Podstawowa, skrócona instrukcja obsługi dla użytkownika – jak posługiwać się narzędziem,
- Opis architektury.

4. OTHER Files. ZIP (obowiązkowy):

- Wygenerowany raport z weryfikacji dostarczonych plików,
- Wynik wizualizacji,
- W przypadku przygotowania Modułu anonimizacji - zanonimizowany dokument w formacie .pdf.