

Eliminating papers and creating electronic archives by using clouds

(eDOX Archiver)

E-Group ICT Software Zrt.

Áron Szabó

Péter Tihanyi

Gábor Garami

June 3-5, 2013



SCI-BUS is supported by the FP7 Capacities Programme under contract nr RI-283481

Introduction

- paperless office

Does it remain a „buzzword” or will there be any improvement in the business or governmental sector?

economic crisis, cost-effectiveness → eliminate paper warehouses
both legal and technical background exist
immediate cost reduction (no fees for hiring, easier look-ups)



Legal background

- paperless office

legal regulations:

decree No. 13/2005. (X. 27.) of the Ministry of Information and Communications on the Rules of Digitization of Paper-based Documents

decree No. 83/2012. (IV. 21.) of the Government on the Mandatory Services of Government and on the Regulated Electronic Public Administration Services

decree No. 78/2012. (XII. 28.) of the Ministry of Interior on Unified Classes of Records at Local Governments



Legal background

- paperless office

legal regulations:

The list of **Regulated Electronic Public Administration Services** that must be provided:



38. The register of administration instructions of citizens'
39. The register of record validity
40. Governmental Certificate Authority (CA)
41. Governmental electronic signature verification service
42. Central identification, authentication and authorization
43. Unified service for sending filled forms
44. Central data authentication (e-Notary)
45. Authentic printing of electronic documents
46. Authentic digitization of paper-based documents
47. Electronic payment and settlement system

Technical background

- paperless office

technical regulations:

retrieved and attached **metadata**

(e.g. title, validity period)

advanced **electronic signature**

qualified **timestamp**



**Decree No. 13/2005. (X. 27.) of the Ministry of Information and Communications
on the Rules of Digitization of Paper-based Documents**

4. §

(3) The digitized copy of a document must be created with the following metadata attached unambiguously:

- a) the title of paper-based document
- b) the physical size of paper-based document
- c) the name of entity responsible for digitization
- d) the digitization policy identifiers
- e) the scanning date of paper-based document
- f) the creation date of paper-based document



Technical background

- paperless office

technical regulation:

Retrieving metadata from the document itself is computation intensive (based on elements of **Dublin Core** - ISO 15836 - **MIReG**) and needs **OCR** functionality and intelligent relevancy **algorithms**.



Technical background

- paperless office

technical regulation:

The documents and records can be classified based on retrieved metadata (described in **decree No. 78/2012. (XII. 28.) of the Ministry of Interior**) and on other „best practices” of business sector.

identifier	title	retention period	archivation
[...]			
U111	Local government decree	non-discardable	15
[...]			
U212	Audit documentation of National Local Governments	non-discardable	15
[...]			
U311	Complaint, request, proposal, notification	2	-
[...]			
U402	Workflow management (documentation of workflows, supporting applications)	non-discardable	15
[...]			
U503	Wage board, salary register	75	-
[...]			
U601	Taxation	5	-
[...]			

Technical background

- paperless office

technical regulation:

digitization of documents (scan) (.pdf, .tiff)

output of optical character recognition (OCR) (.txt, .html, .docx),

retrieving metadata (MIReG) (.xml)

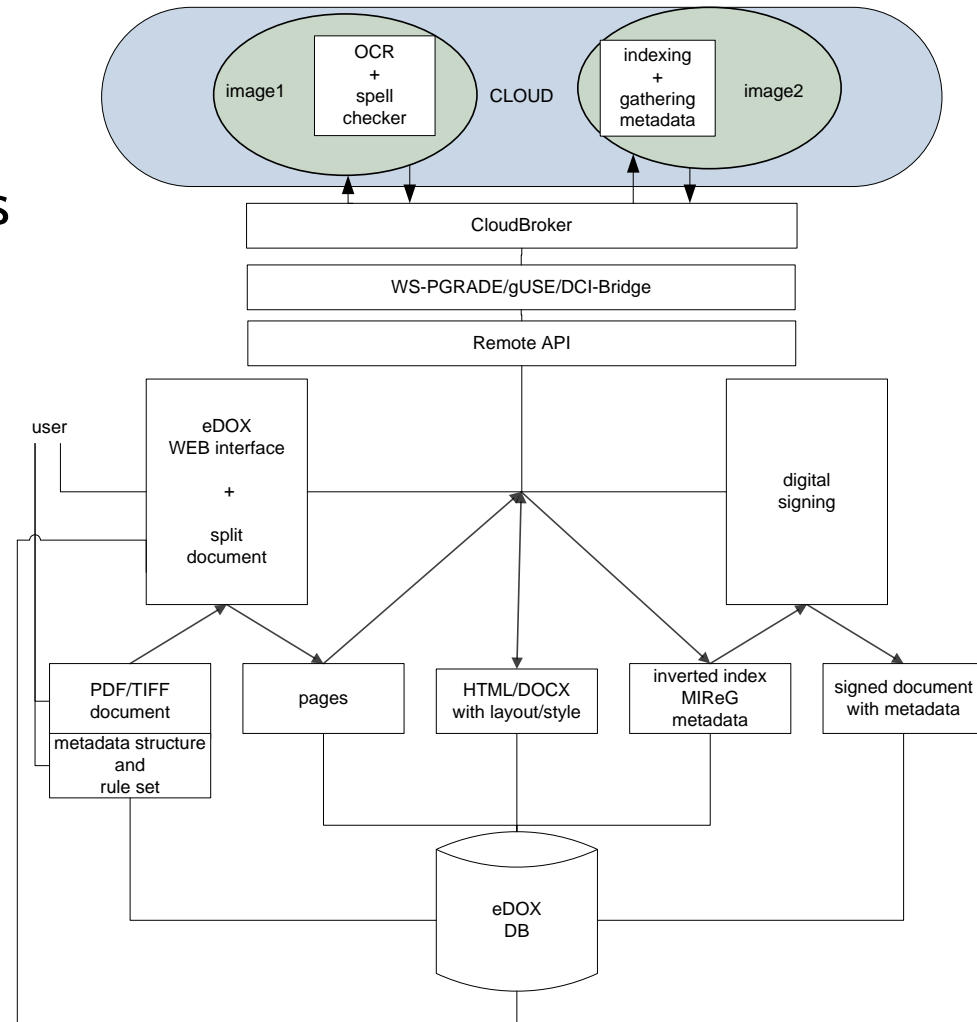
→ protect them together with electronic signature (XAdES)



Technical background

- cloud environment

The **computation intensive** steps of document processing (OCR, spell-checking, word analysis, indexing, retrieving metadata, document classification) are executed in the cloud.



Technical background

- cloud environment

communications:

SZTAKE - Remote API

web service for uploading

SZTAKE - gUSE

workflow management

SZTAKE - DCI Bridge

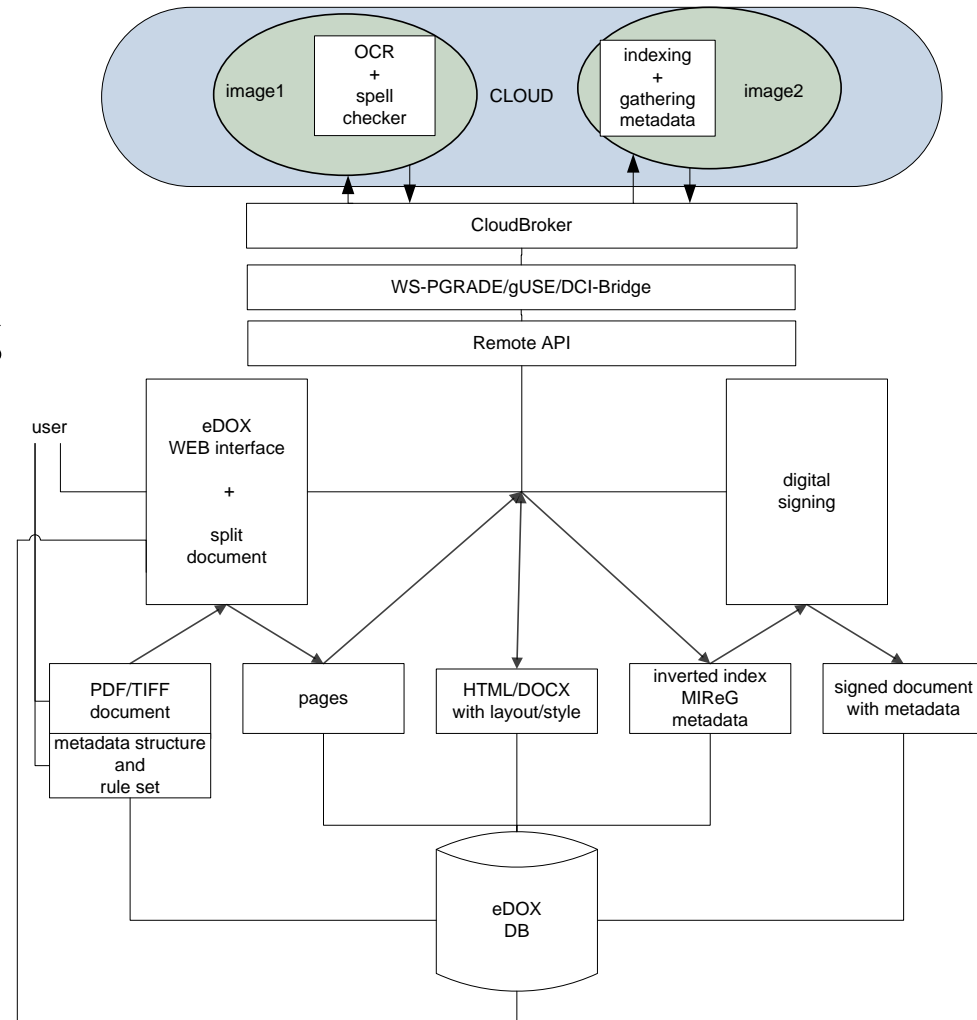
setting cloud parameters

CloudBroker


management of code
running in the cloud

Amazon - EC2


cloud resource provider



eDOX Archiver



eDOX ARCHIVER



E-GROUP
SOFTWARE & BEYOND

Log in - EDOX Archiver

Log in

Username

Password

New user: register [here](#).

eDOX Archiver



eDOX™ ARCHIVER



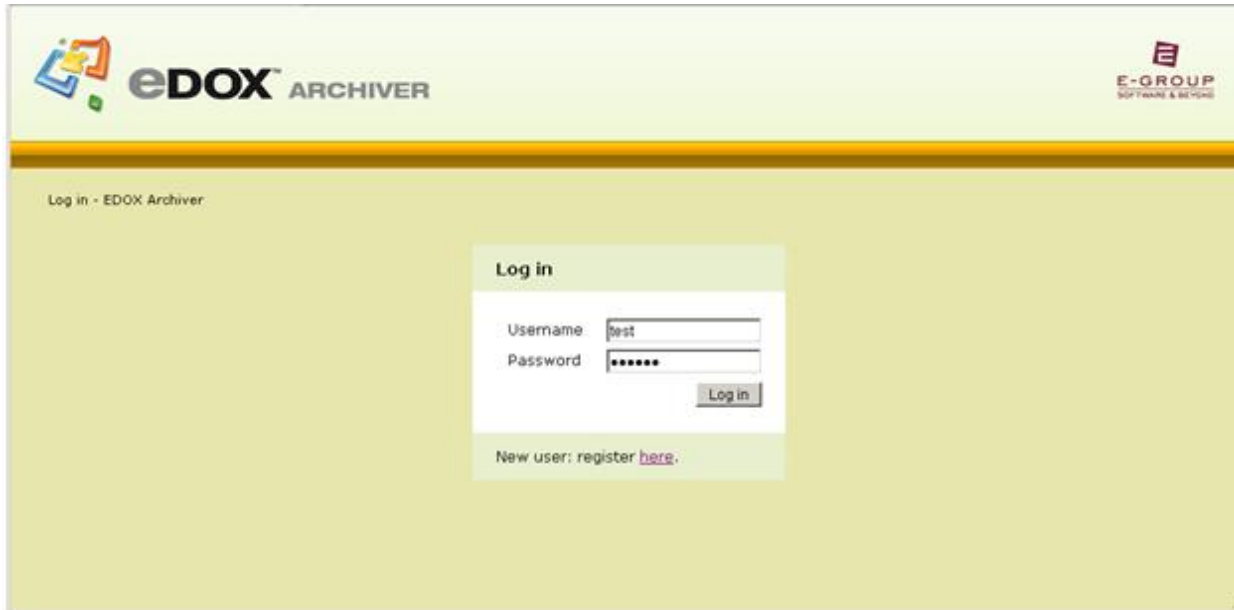
E-GROUP
SOFTWARE & BEYOND

Register - EDOX Archiver

Register

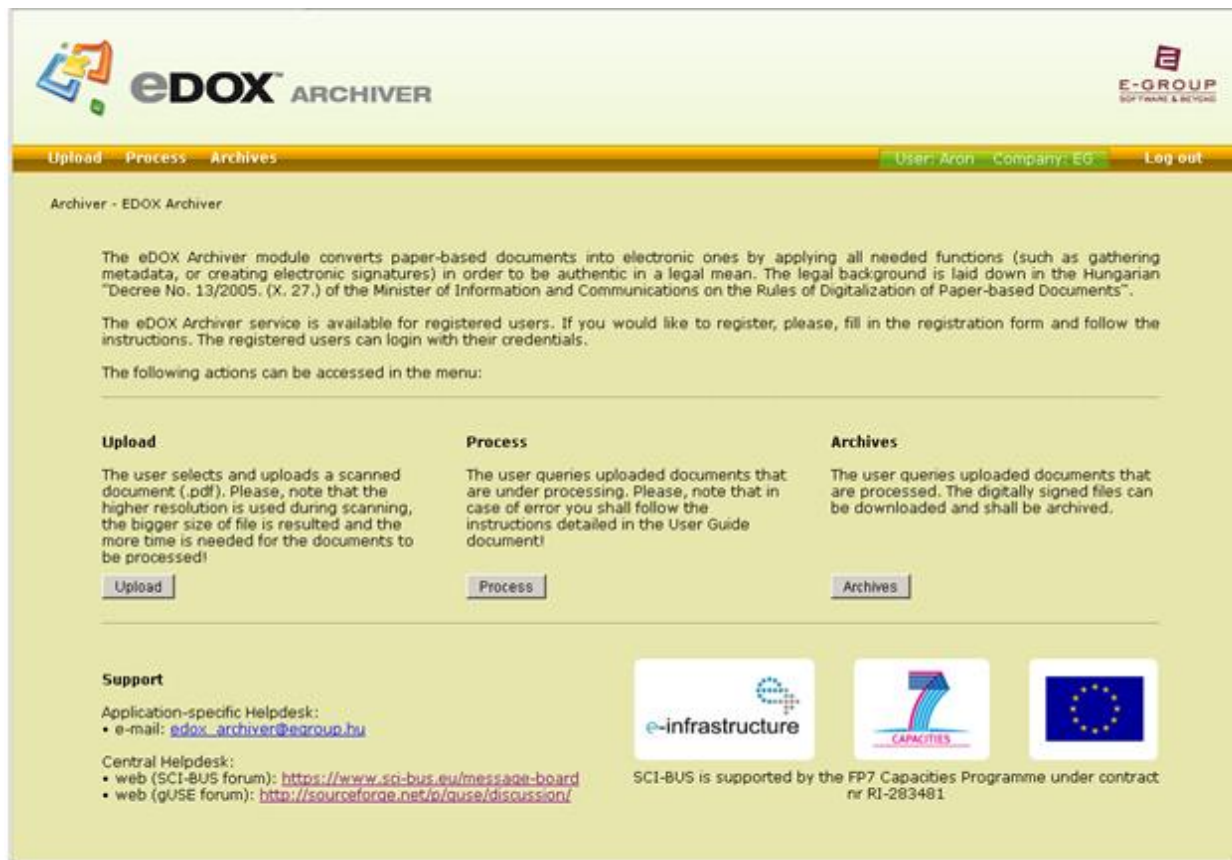
Username	<input type="text" value="test"/>
Password	<input type="password" value="*****"/>
Confirm password	<input type="password" value="*****"/>
Name	<input type="text" value="Aron"/>
Company	<input type="text" value="EG"/>
Archivation policy	<input type="text" value="http://www.egroup.hu/policyxr"/>

eDOX Archiver



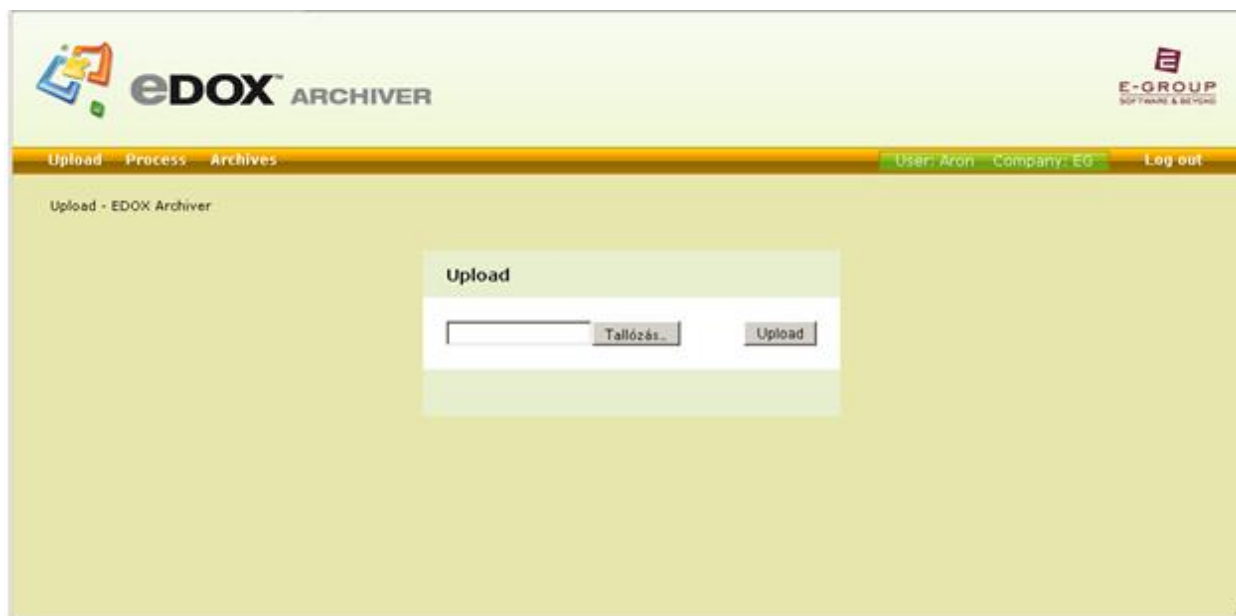
The screenshot shows the login interface of the eDOX Archiver. At the top left is the eDOX ARCHIVER logo, and at the top right is the E-GROUP SOFTWARE & BEYOND logo. Below the logos, the text "Log in - EDOX Archiver" is displayed. In the center, there is a "Log in" section with two input fields: "Username" containing the text "test" and "Password" containing seven dots. A "Log in" button is positioned to the right of the password field. Below the login fields, a link for new users is provided: "New user: register [here](#)."

eDOX Archiver

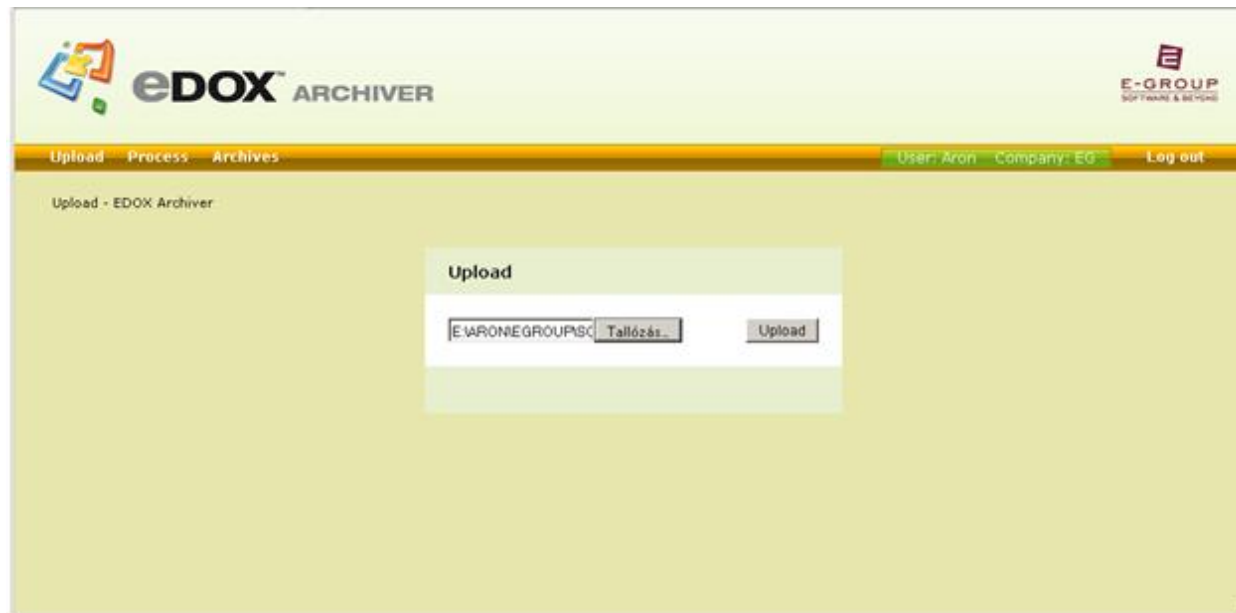


The screenshot displays the eDOX Archiver web application. At the top, there is a header with the eDOX ARCHIVER logo on the left and the E-GROUP SOFTWARE & BEYOND logo on the right. Below the header is a navigation bar with tabs for 'Upload', 'Process', and 'Archives'. On the right side of the navigation bar, there is a user status bar showing 'User: Aron', 'Company: EG', and a 'Log out' button. The main content area is titled 'Archiver - EDOX Archiver'. It contains a paragraph explaining the module's function: converting paper-based documents into electronic ones by applying all needed functions (such as gathering metadata, or creating electronic signatures) in order to be authentic in a legal mean. The legal background is laid down in the Hungarian "Decree No. 13/2005. (X. 27.) of the Minister of Information and Communications on the Rules of Digitalization of Paper-based Documents". Another paragraph states that the eDOX Archiver service is available for registered users, and if they would like to register, please, fill in the registration form and follow the instructions. The registered users can login with their credentials. A third paragraph states that the following actions can be accessed in the menu: Upload, Process, and Archives. Below this text, there are three columns, each with a title and a description. The 'Upload' column describes the process of selecting and uploading a scanned document (.pdf), noting that higher resolution is used during scanning, resulting in a bigger file size and more time needed for processing. Below the description is an 'Upload' button. The 'Process' column describes querying uploaded documents that are under processing, noting that in case of error, the user shall follow the instructions detailed in the User Guide document. Below the description is a 'Process' button. The 'Archives' column describes querying uploaded documents that are processed, noting that the digitally signed files can be downloaded and shall be archived. Below the description is an 'Archives' button. At the bottom of the page, there is a 'Support' section. It includes 'Application-specific Helpdesk' with an email address: edox_archiver@egroup.hu. It also includes 'Central Helpdesk' with two web links: <https://www.sci-bus.eu/message-board> and <http://sourceforge.net/fo/quse/discussion/>. To the right of the support section, there are three logos: 'e-infrastructure', 'CAPACITIES', and the European Union flag. Below these logos, it states: 'SCI-BUS is supported by the FP7 Capacities Programme under contract nr RI-283481'.

eDOX Archiver



eDOX Archiver



eDOX Archiver




eDOX Archiver




The screenshot shows the eDOX Archiver web application. At the top, there is a header with the eDOX ARCHIVER logo on the left and the E-GROUP SOFTWARE & BEYOND logo on the right. Below the header is a navigation bar with tabs for 'Upload', 'Process', and 'Archives'. The 'Process' tab is currently selected. In the top right corner of the navigation bar, there is a user status bar showing 'User: Aron' and 'Company: EG', along with a 'Log out' button. Below the navigation bar, the main content area displays 'Processes - EDOX Archiver'. A section titled 'Processes' contains a table with the following data:

File name	Start time	Physical Size	User	Action
test_01_300dpi.pdf	2013.01.24. 15:25:34	A4	User	
test_01_300dpi.pdf	2013.01.23. 11:17:52	A4	User	

eDOX Archiver



eDOX ARCHIVER



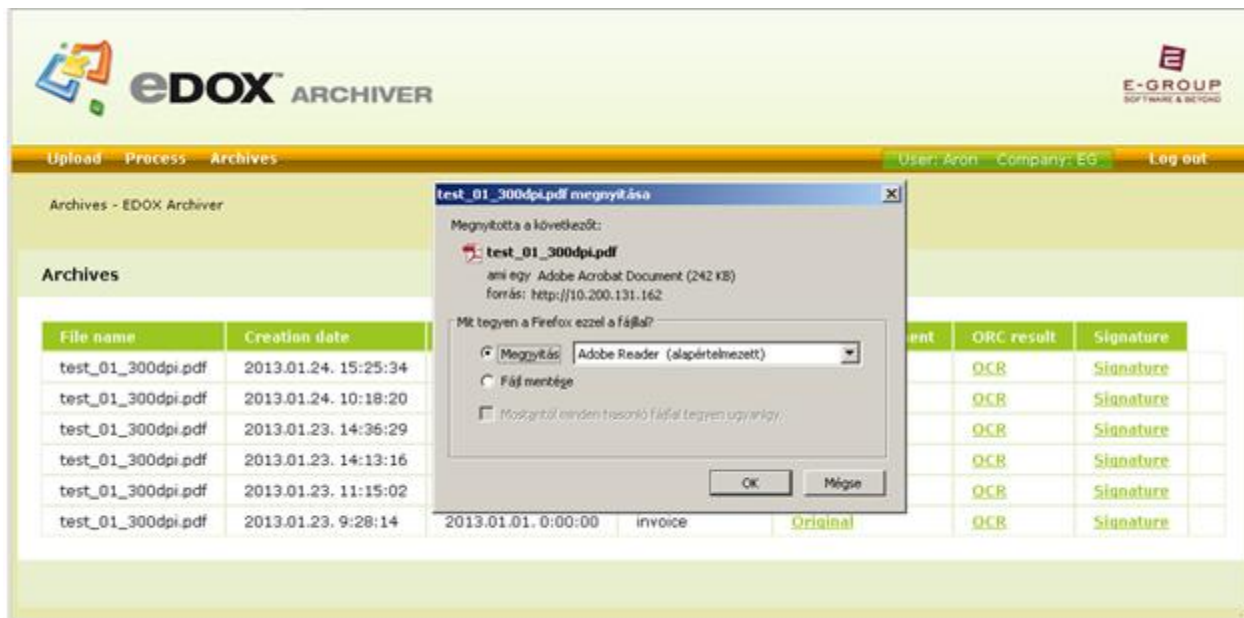
Upload Process Archives User: Aron Company: EG Log out

Archives - EDOX Archiver

Archives

File name	Creation date	Archivation date	Document type	Original document	ORC result	Signature
test_01_300dpi.pdf	2013.01.24. 15:25:34	2013.01.01. 0:00:00	invoice	Original	OCR	Signature
test_01_300dpi.pdf	2013.01.24. 10:18:20	2013.01.01. 0:00:00	invoice	Original	OCR	Signature
test_01_300dpi.pdf	2013.01.23. 14:36:29	2013.01.01. 0:00:00	invoice	Original	OCR	Signature
test_01_300dpi.pdf	2013.01.23. 14:13:16	2013.01.01. 0:00:00	invoice	Original	OCR	Signature
test_01_300dpi.pdf	2013.01.23. 11:15:02	2013.01.01. 0:00:00	invoice	Original	OCR	Signature
test_01_300dpi.pdf	2013.01.23. 9:28:14	2013.01.01. 0:00:00	invoice	Original	OCR	Signature

eDOX Archiver



eDOX ARCHIVER

Upload Process Archives User: Aron Company: EG Log out


Archives - EDOX Archiver

Archives

File name	Creation date
test_01_300dpi.pdf	2013.01.24. 15:25:34
test_01_300dpi.pdf	2013.01.24. 10:18:20
test_01_300dpi.pdf	2013.01.23. 14:36:29
test_01_300dpi.pdf	2013.01.23. 14:13:16
test_01_300dpi.pdf	2013.01.23. 11:15:02
test_01_300dpi.pdf	2013.01.23. 9:28:14

test_01_300dpi.pdf megnyitása

Megnyitotta a következőt:

 **test_01_300dpi.pdf**
ami egy Adobe Acrobat Document (242 KB)
Forrás: http://10.200.131.162

Mit tegyen a Firefox ezzel a fájjal?

☒ **Megnyitás** Adobe Reader (alapértelmezett)

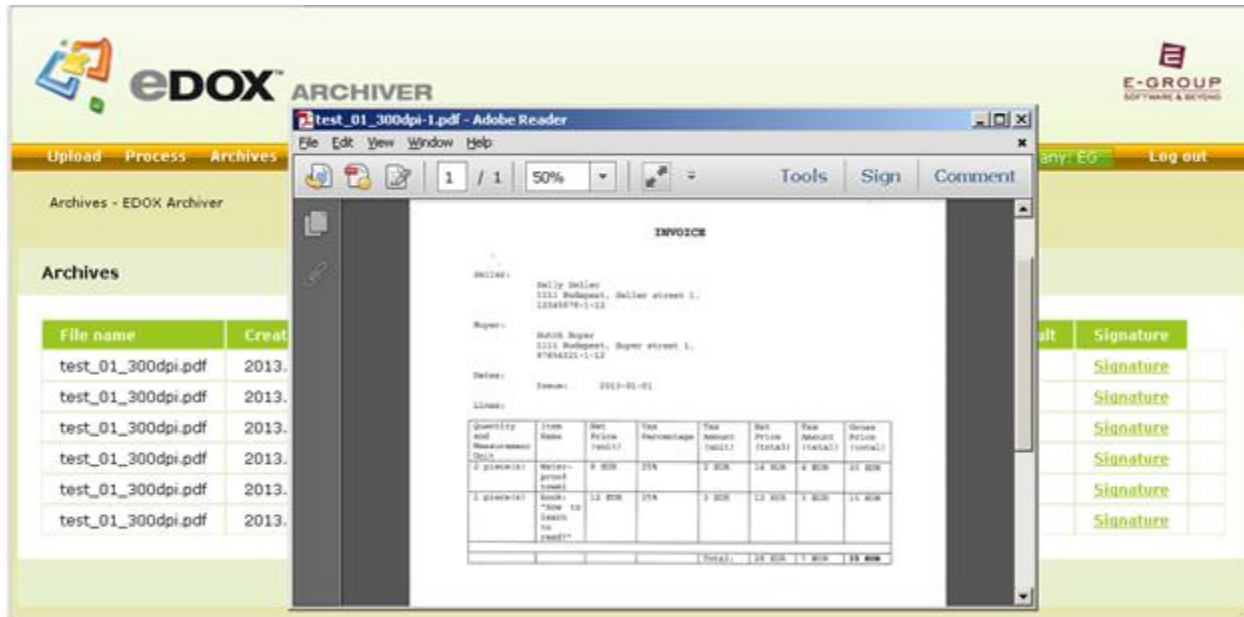
☐ **Fájl mentése**

☐ Mostantól minden hasonló fájlra tegyen ugyanígy.

OK Mégse

ent	ORC result	Signature
	OCR	Signature
	OCR	Signature
	OCR	Signature
	OCR	Signature
	OCR	Signature
	OCR	Signature
2013.01.01. 0:00:00	invoice	Original
	OCR	Signature

eDOX Archiver



eDOX ARCHIVER

Upload Process Archives

Archives - EDOX Archiver

Archives

File name	Create
test_01_300dpi.pdf	2013.
test_01_300dpi.pdf	2013.
test_01_300dpi.pdf	2013.
test_01_300dpi.pdf	2013.
test_01_300dpi.pdf	2013.
test_01_300dpi.pdf	2013.

test_01_300dpi-1.pdf - Adobe Reader

File Edit View Window Help

1 / 1 50% Tools Sign Comment

INVOICE

Bill to: BILLY BELLER
1111 Beller Street, Beller Street 1,
1111 Beller Street 1-1111

Bill to: BILLY BELLER
1111 Beller Street, Beller Street 1,
1111 Beller Street 1-1111

Bill to: BILLY BELLER
1111 Beller Street, Beller Street 1,
1111 Beller Street 1-1111

Quantity	Unit Price	Tax Percentage	Tax Amount	Net Price	Tax Amount	Gross Price
1	14.000	10%	1.400	12.600	1.400	14.000
1	14.000	10%	1.400	12.600	1.400	14.000
			Total:	14.000	1.400	15.400

any: EDOX Log out

Signature

Signature

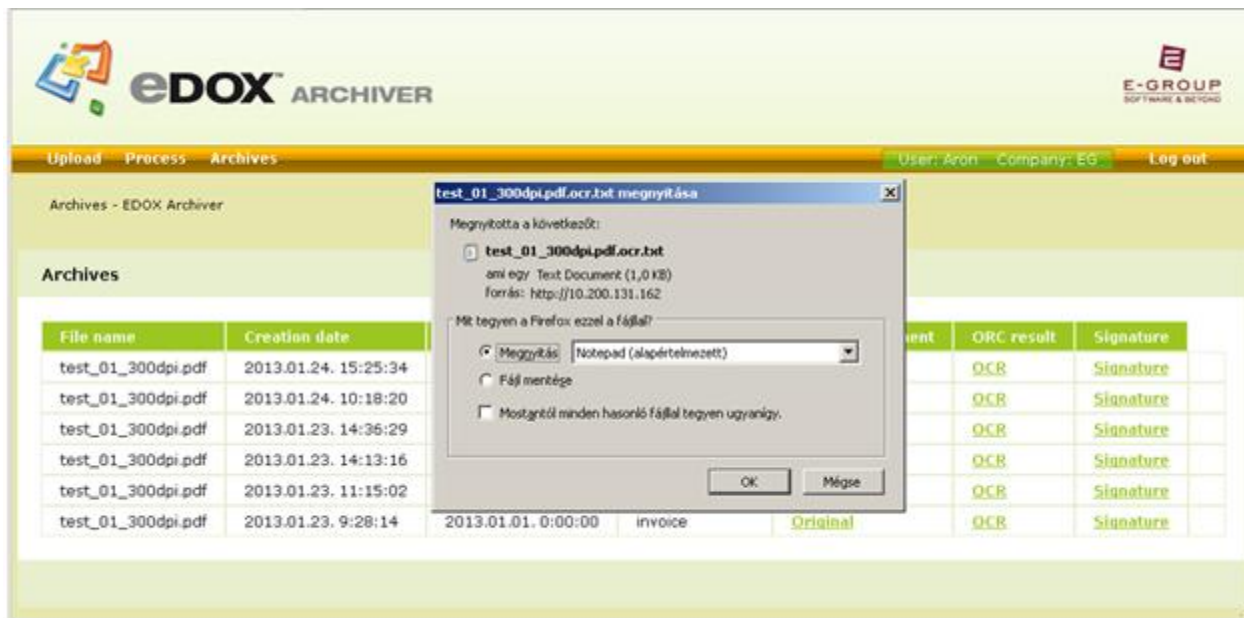
Signature

Signature

Signature

Signature

eDOX Archiver



Archives - EDOX Archiver

Archives

File name	Creation date
test_01_300dpi.pdf	2013.01.24. 15:25:34
test_01_300dpi.pdf	2013.01.24. 10:18:20
test_01_300dpi.pdf	2013.01.23. 14:36:29
test_01_300dpi.pdf	2013.01.23. 14:13:16
test_01_300dpi.pdf	2013.01.23. 11:15:02
test_01_300dpi.pdf	2013.01.23. 9:28:14

test_01_300dpi.pdf.ocr.txt megnyitása

Megnyitotta a következőt:

- test_01_300dpi.pdf.ocr.txt
ami egy Text Document (1,0 KB)
Forrás: http://10.200.131.162

Mit tegyen a Firefox ezzel a fájlal?

☒ Megnyitás: Notepad (alapértelmezett)

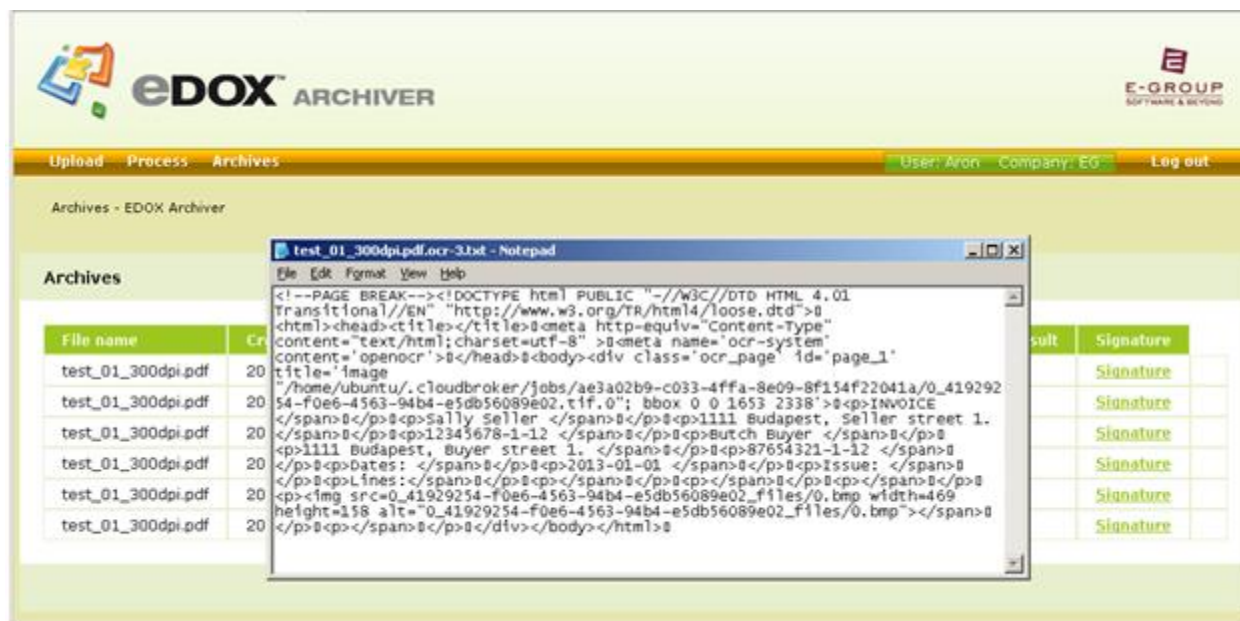
☐ Fájlmentő

☐ Mostantól minden hasonló fájlra tegyen ugyanígy.

OK Mégse

File name	Creation date	ORC result	Signature
test_01_300dpi.pdf	2013.01.24. 15:25:34	OCR	Signature
test_01_300dpi.pdf	2013.01.24. 10:18:20	OCR	Signature
test_01_300dpi.pdf	2013.01.23. 14:36:29	OCR	Signature
test_01_300dpi.pdf	2013.01.23. 14:13:16	OCR	Signature
test_01_300dpi.pdf	2013.01.23. 11:15:02	OCR	Signature
test_01_300dpi.pdf	2013.01.23. 9:28:14	OCR	Signature

eDOX Archiver



eDOX ARCHIVER

Upload Process Archives User: Aron Company: EG Log out

Archives - EDOX Archiver

Archives

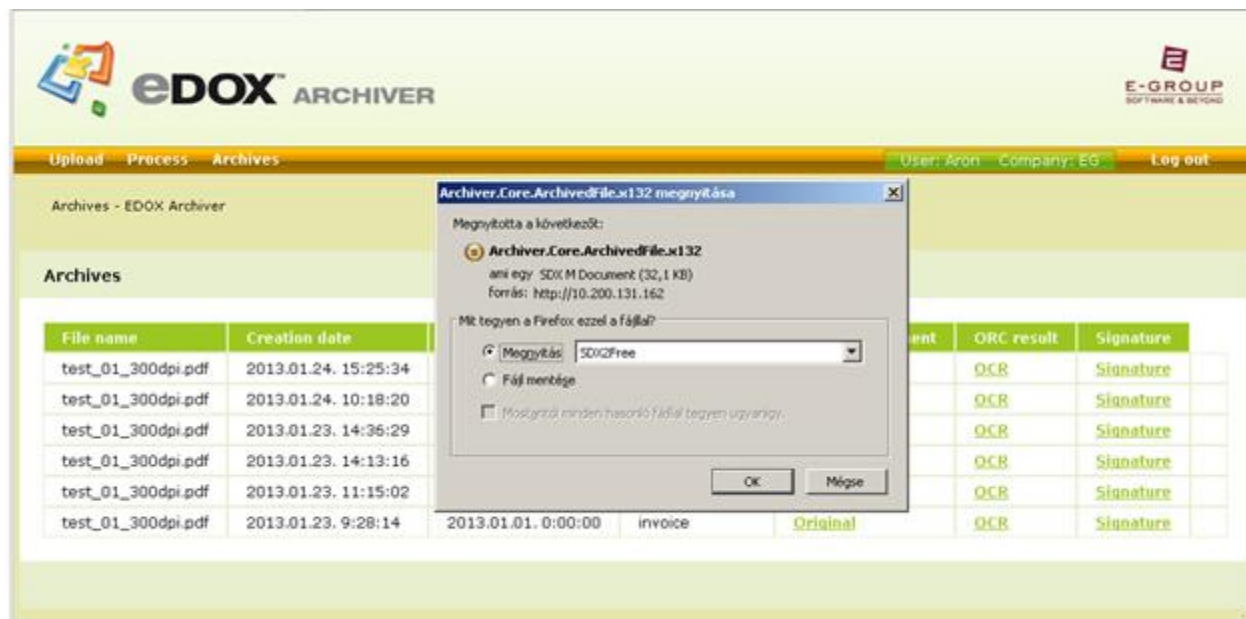
File name	Cr
test_01_300dpi.pdf	20
test_01_300dpi.pdf	20
test_01_300dpi.pdf	20
test_01_300dpi.pdf	20
test_01_300dpi.pdf	20
test_01_300dpi.pdf	20

test_01_300dpi.pdf.ocr-3.txt - Notepad

```
<!--PAGE BREAK--><!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01
Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
<html><head><title></title><meta http-equiv="Content-Type"
content="text/html; charset=utf-8" ><meta name="ocr-system"
content="openocr"></head><body><div class="ocr_page" id="page_1"
title="image"
"/home/ubuntu/.cloudbroker/jobs/ae3a02b9-c033-4ffa-8e09-8f154f22041a/0_419292
54-f0e6-4563-94b4-e5db56089e02.tif.0"; bbox 0 0 1653 2338"><p>INVOICE
</span></p><p>Sally Seller </span></p><p>1111 Budapest, Seller street 1.
</span></p><p>12345678-1-12 </span></p><p>Butch Buyer </span></p>
<p>1111 Budapest, Buyer street 1. </span></p><p>87654321-1-12 </span>
</p><p>Dates: </span></p><p>2013-01-01 </span></p><p>Issue: </span>
</p><p>Lines: </span></p><p></span></p><p></span></p><p></span></p>
<p></span>
</p><p></span></p></div></body></html>
```

Result	Signature
	Signature
	Signature
	Signature
	Signature
	Signature
	Signature

eDOX Archiver



eDOX ARCHIVER

Upload Process Archives User: Aron Company: EG Log out

Archives - EDOX Archiver

Archives

File name	Creation date
test_01_300dpi.pdf	2013.01.24. 15:25:34
test_01_300dpi.pdf	2013.01.24. 10:18:20
test_01_300dpi.pdf	2013.01.23. 14:36:29
test_01_300dpi.pdf	2013.01.23. 14:13:16
test_01_300dpi.pdf	2013.01.23. 11:15:02
test_01_300dpi.pdf	2013.01.23. 9:28:14

Archiver.Core.ArchivedFile.x132 megnyitása

Megnyitotta a következőt:

- Archiver.Core.ArchivedFile.x132
ami egy SDX M Document (32,1 KB)
Forrás: http://10.200.131.162

Mit tegyen a Firefox ezzel a fájlal?

☒ Megnyitás SDX2Free

☐ Fáj mentése

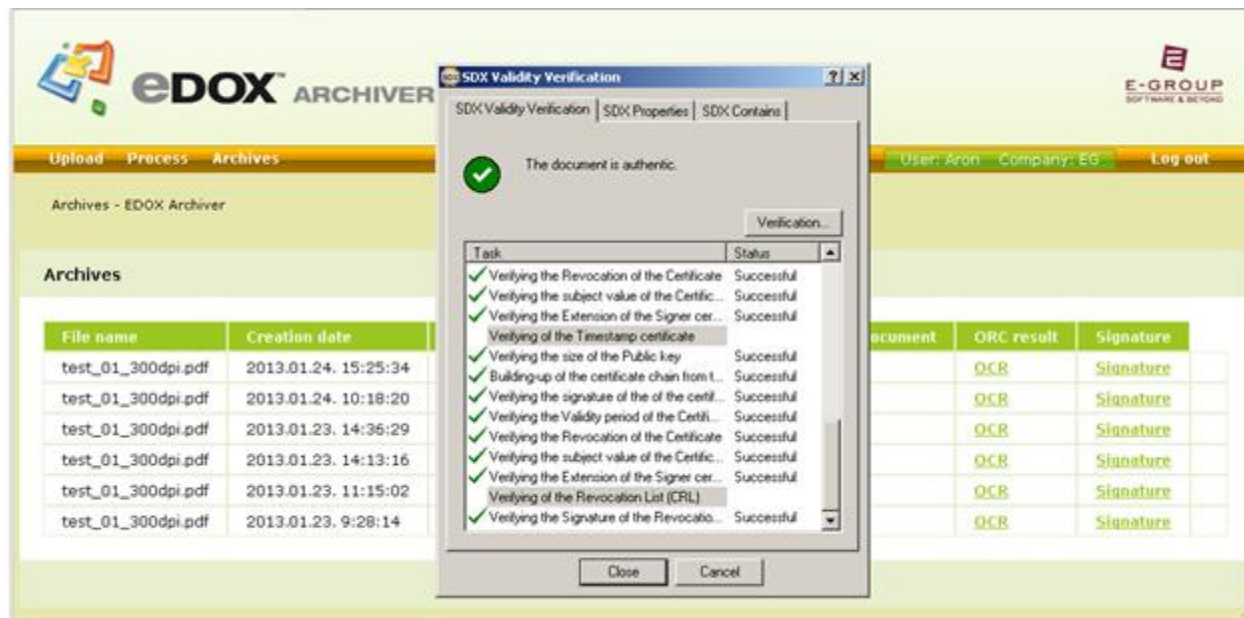
☐ Mielőtt minden használt fájlt tegyen vázlati...

OK Mégse

ent	ORC result	Signature
	OCR	Signature
	OCR	Signature
	OCR	Signature
	OCR	Signature
	OCR	Signature
	OCR	Signature

2013.01.01. 0:00:00 invoice Original OCR Signature

eDOX Archiver



The screenshot displays the eDOX Archiver web interface. A modal dialog titled "SDX Validity Verification" is open, showing a green checkmark and the message "The document is authentic." Below this, a table lists verification tasks and their statuses, all of which are "Successful".

Archives

File name	Creation date
test_01_300dpi.pdf	2013.01.24. 15:25:34
test_01_300dpi.pdf	2013.01.24. 10:18:20
test_01_300dpi.pdf	2013.01.23. 14:36:29
test_01_300dpi.pdf	2013.01.23. 14:13:16
test_01_300dpi.pdf	2013.01.23. 11:15:02
test_01_300dpi.pdf	2013.01.23. 9:28:14

SDX Validity Verification

SDX Validity Verification | SDX Properties | SDX Contents

✓ The document is authentic.

Verification...

Task	Status
✓ Verifying the Revocation of the Certificate	Successful
✓ Verifying the subject value of the Certificate	Successful
✓ Verifying the Extension of the Signer certificate	Successful
Verifying of the Timestamp certificate	
✓ Verifying the size of the Public key	Successful
✓ Building-up of the certificate chain from it...	Successful
✓ Verifying the signature of the certificate	Successful
✓ Verifying the Validity period of the Certificate	Successful
✓ Verifying the Revocation of the Certificate	Successful
✓ Verifying the subject value of the Certificate	Successful
✓ Verifying the Extension of the Signer certificate	Successful
Verifying of the Revocation List (CRL)	
✓ Verifying the Signature of the Revocation...	Successful

Close Cancel

Thank you!



Secure and Authentic Paper Documents



mailto: aron.szabo@egroup.hu
mailto: peter.tihanyi@egroup.hu
mailto: gabor.garami@egroup.hu

web: <http://www.egroup.hu>