



**NUPSESO**<sup>®</sup>  
a.s.



# Európsky dvor audítorov Luxemburg

- Digitalizácia dokumentov
- Indexácia
- Inteligentná konverzia OCR do

Juraj Lievaj, NUPSESO, a.s.

# Európsky dvor audítorov

- **Externý audítor** finančných prostriedkov EÚ
- **Vykonáva audit** všeobecného rozpočtu, európskych rozvojových fondov a ostatných orgánov a agentúr EÚ
- **Nezávislá inštitúcia** Európskej Únie v súlade s ustanoveniami článkov 246 – 248 Zmluvy o ES
- **Zriadený** Bruselskou zmluvou zo dňa

# Dvor audítorov uverejňuje

- **Výročné správy** – predkladajú výsledky finančných auditov vo forme vyhlásení o vierohodnosti všeobecného rozpočtu a európskych rozvojových fondov
- **Špecifické ročné správy** – predkladajú výsledky finančných auditov agentúr a orgánov Spoločenstiev
- **Osobitné správy** – predkladajú výsledky vybraných auditov výkonnosti a auditov zhody
- **Stanoviská** – k novým alebo aktualizovaným právnym predpisom s finančným dosahom

# Knižnica a archívy

- **Knižnica** Európskeho dvora audítorov bola **založená v roku 1978** pre profesionálne potreby zamestnancov Dvora audítorov
- **Knižničný fond** je špecializovaný na audit, účtovníctvo, právo EÚ a všetky oblasti politiky EÚ, ktoré zodpovedajú audítorským sektorom
- **Knižnica obsahuje** vyše 10 000 kníh a 100 rozličných predplatných novín, najmä v angličtine, francúzštine a nemčine
- **Celá knižnica bude čoskoro k dispozícii online prostredníctvom špeciálneho webového užívateľského rozhrania Web-OPAC**
- **Archívy** Európskeho dvora audítorov obsahujú všetky pracovné zložky Dvora audítorov (správy a stanoviská) od jeho založenia v roku 1977

# Rozsah digitalizácie

- **Konverzia** existujúcich obrazov vo Florencii do archívneho štandardu **PDF/A** (viac ako 400 000 strán) s fulltextovým vyhľadávaním
- **Skenovanie, indexácia a inteligentná dynamická OCR konverzia** existujúcich dokumentov do archívneho štandardu **PDF/A** (viac ako 1 000 000 strán) s fulltextovým

# Podmienky projektu

- **Neunifikované dátové podklady k indexácii**
- Texty tlačené rôznymi technológiami na **viacfarebné papiere**
- **Vysoká rôznorodosť fontov, písma a diakritiky** (latinka, cyrilika, grécke písmo...)
- **Viacjazyčné dokumenty** (niekedy aj viac ako 20 jazykov v jednom)

# Vzor viacjazyčného



CIRCULAR  
MEDDELELSE TIL PERSONALET  
MITTEILUNG AN DAS PERSONAL  
ΑΝΑΚΟΙΝΩΣΗ ΣΤΟ ΠΡΟΣΩΠΙΚΟ  
STAFF NOTICE  
COMMUNICATION AU PERSONNEL  
COMUNICAZIONE AL PERSONALE  
MEDEDELING AAN HET PERSONEEL  
COMUNICAÇÃO AO PESSOAL  
TIEDOTE HENKILÖKUNNALLE  
PERSONALMEDDELANDE

Fecha      Dato  
Date      Data  
Datum      Ημερομηνία  
Päivämäärä

12 VIII. 1997

No. 51-97  
Αριθ.

**Θέμα:** Σχολικό επίδομα (άρθρο 3, παράρτημα III του Κανονισμού υπηρεσιακής κατάστασης και γενικές εκτελεστικές διατάξεις)

1. Γνωστοποιείται στους μόνιμους και έκτακτους υπαλλήλους ότι οι αιτήσεις για τη χορήγηση του σχολικού επιδόματος πρέπει να υποβληθούν για κάθε παιδί για το σχολικό έτος 1997-98 (1η Σεπτεμβρίου 1997 έως 31 Αυγούστου 1998).\*
2. Το συνημμένο έντυπο θα πρέπει να αποσταλεί δεόντως συμπληρωμένο και υπογεγραμμένο στην αρμόδια υπηρεσία, μαζί με τα παρακάτω δικαιολογητικά:

# Vzor viacjazyčného

ΕΒΡΟΠΕΪΣΚΑ ΣΜΕΤΗΑ ΠΑΛΑΤΑ  
TRIBUNAL DE CUENTAS EUROPEO  
EVROPSKÝ ÚČETNÍ DVŮR  
DEN EUROPÆISKE REVISIONSRET  
EUROPÄISCHER RECHNUNGSHOF  
EUROOPA KONTROLLIKODA  
ΕΥΡΩΠΑΪΚΟ ΕΛΕΓΚΤΙΚΟ ΣΥΝΕΔΡΙΟ  
EUROPEAN COURT OF AUDITORS  
COUR DES COMPTES EUROPÉENNE  
CŪIRT INIÚCHÓIRÍ NA HEORPA



CORTE DEI CONTI EUROPEA  
EIROPAS REVĪZIJAS PALĀTA  
EUROPOS AUDITO RŪMAI

EURÓPAI SZÁMVEVŐSZÉK  
IL-QORTI EWROPEA TA' L-AWDITURI  
EUROPESE REKENKAMER  
EUROPEJSKI TRYBUNAŁ OBRACHUNKOWY  
TRIBUNAL DE CONTAS EUROPEU  
CURTEA DE CONTURI EUROPEANĂ  
EURÓPSKY DVOR AUDÍTOROV  
EVROPSKO RAČUNSKO SODIŠČE  
EUROOPAN TILINTARKASTUSTUOMIOISTUIN  
EUROPEISKA REVISIONSRÄTTEN

CPT006989FR03.doc

## **Décision de la Cour des comptes n° 43-2007 relative à son Instance spécialisée en matière d'irrégularités financières**

### **LA COUR DES COMPTES DES COMMUNAUTÉS EUROPÉENNES,**

- VU les articles 60, paragraphe 6, et 66, paragraphe 4, du règlement financier<sup>1</sup>,
- VU les articles 74 et 75, paragraphe 2, des modalités d'exécution du règlement financier<sup>2</sup>,
- VU la décision de la Cour du 24 octobre 2002 par laquelle elle a établi son Instance spécialisée en matière d'irrégularités financières<sup>3</sup> ;



# Príprava, indexy, databázy

- **Import zdrojových dát z neunifikovaných excelovských tabuliek do SQL**
- **Unifikácia rôznorodeho formátovania buniek, zalomenia riadkov, nepovolených znakov...**
- **Tlač a lepenie čiarových kódov**
- **Doindexovanie počas procesu skenovania**

# Vzor – Excel súbor

CCE\_RAS\_EURATOM\_1992-1997.xls [režim kompatibility] - Microsoft Excel

Soubor Domů Vložení Rozložení stránky Vzorce Data Revize Zobrazení

Vložit Vymout Kopírovat Kopírovat formát

Schránka Písmo Zarovnání Číslo

Číslo % 000 +,0 -,0

Podmíněné formátování Formátovat jako tabulku Styly

K56 RAS sur les comptes de l'Agence d'approvisionnement d'Euratom 1997 accompagné des réponses de l'Agence - Version néerlanda

Folder	FILE	TITLE	AUTHOR	SUBJECT	KEYWORDS
1	CCE_RAS_EURATOM_1996_1.pdf	Bordereau du rapport sur le comptes de l'Agence d'approvisionnement d'Euratom relatif à l'exercice 1996	Converted by NUPSESO, a s	Scanned document converted to PDF/A file	ENERGY, EURATOM
1	CCE_RAS_EURATOM_1996_2.pdf	GA 50/97	Converted by NUPSESO, a s	Scanned document converted to PDF/A file	ENERGY, EURATOM
1	CCE_RAS_EURATOM_1996_3.pdf	GA 59/97 FINAL (PV)	Converted by NUPSESO, a s	Scanned document converted to PDF/A file	ENERGY, EURATOM
1	CCE_RAS_EURATOM_1996_4.pdf	DEC 49/97	Converted by NUPSESO, a s	Scanned document converted to PDF/A file	ENERGY, EURATOM
1	CCE_RAS_EURATOM_1996_5.pdf	Réponse de l'Agence Euratom au projet de RAS de la Cour pour 1996	Converted by NUPSESO, a s	Scanned document converted to PDF/A file	ENERGY, EURATOM
1	CCE_RAS_EURATOM_1996_6.pdf	DEC 49/97 add. 1	Converted by NUPSESO, a s	Scanned document converted to PDF/A file	ENERGY, EURATOM
1	CCE_RAS_EURATOM_1996_7.pdf	DEC 49/97 (R)	Converted by NUPSESO, a s	Scanned document converted to PDF/A file	ENERGY, EURATOM
1	CCE_RAS_EURATOM_1996_8.pdf	DEC 54/97 FINAL (PV)	Converted by NUPSESO, a s	Scanned document converted to PDF/A file	ENERGY, EURATOM

# Inteligentné, dynamické

- **Parametrizovanie OCR servera** zvlášť pre každý jednotlivý dokument
- **Dynamické nastavovanie** jazykov, druhu tlače, kvality predlohy a iných parametrov pomocou tzv. XML Ticket-ov
- **Nadprahová kontrola** procesu OCR operátorom verifikačnej stanice
- **Dosahovaná presnosť** cca. 1,4 % neistých znakov (chybovosť býva štatisticky nižšia)

# Vzor – XML Ticket input

```
16 <<Language>Italian</Language>
17 <<Language>French</Language>
18 </RecognitionParams>
19 <ExportParams DocumentSeparationMethod="MergeIntoSingleFile">
20 <<ExportFormat OutputFileFormat="PDF/A" KeepLastModifiedDate="false" OutputFlowType="SharedFolder
21 <<OutputLocation>D:\RecognitionServer\ECA_Default\RS_OUTPUT\Box0012\</OutputLocation>
22 <<NamingRule>CCE_COMPERS_1997_48_FR.pdf</NamingRule>
23 <Summary>
24 <SummaryInfoItem>
25 <<Name>Author</Name>
26 <<Value>BEUROTTE</Value>
27 </SummaryInfoItem>
28 <SummaryInfoItem>
29 <<Name>Keywords</Name>
30 <<Value>Joint European Torus; JET; Energy; Euratom</Value>
31 </SummaryInfoItem>
32 <SummaryInfoItem>
33 <<Name>Subject</Name>
34 <<Value>Décision n° 97-34 modifiant la décision n°90-35 portant réglementation pour le rembourse
35 </SummaryInfoItem>
36 <SummaryInfoItem>
37 <<Name>Title</Name>
38 <<Value>Communication au personnel 1997 48/97</Value>
39 </SummaryInfoItem>
40 </Summary>
41 <<Producer>Scanned and converted by NUPSESOS a.s.</Producer>
42 </ExportFormat>
43 </ExportParams>
44 </XmlTicket>
```

# Vzor – XML Ticket output

```
<UserDictionaryFile>Eca.txt{0FAA02AC-A7A4-42B5-943E-16FA55527550}</UserDictionaryFile>
</RecognitionParams>
- <ExportParams DocumentSeparationMethod="MergeIntoSingleFile" XMLResultPublishingMethod="XMLResultToFolder" JobsOrderMe
  <SeparationScriptLanguage />
  <PublishingScriptLanguage />
  <FailedScriptLanguage />
- <ExportFormat Id="{B2E5E3EA-A0D8-4824-9357-25C99A402FA0}" OutputFileFormat="PDFA" KeepLastModifiedDate="false" C
  ExportMode="ImageOnText" PictureFormat="Automatic" Quality="100" PictureResolution="-1" UseExplicitDocumentInfo="true"
  KeepOriginalHeadersFooters="true" PdfVersion="Auto" Write1ACompliant="false">
  <OutputLocation>\\Server\RecognitionServer\ECA_Default\RS_OUTPUT\Box0043</OutputLocation>
  <NamingRule>CCE_RAS_CECA Etats financieres _1991_14__EN.pdf</NamingRule>
- <Summary>
- <SummaryInfoItem>
  <Name>Author</Name>
  <Value>Lambert</Value>
</SummaryInfoItem>
- <SummaryInfoItem>
  <Name>Keywords</Name>
  <Value>CECA, Rapport, financières</Value>
</SummaryInfoItem>
- <SummaryInfoItem>
  <Name>Subject</Name>
  <Value>Rapport relatif aux états financiers de la CECA au 31 décembre 1991 - Version anglaise</Value>
</SummaryInfoItem>
- <SummaryInfoItem>
  <Name>Title</Name>
  <Value>RAS CECA Etats financiers 1991 - Version anglaise</Value>
</SummaryInfoItem>
</Summary>
<Producer>Scanned and converted by NUPSESO@ a.s.</Producer>
</ExportFormat>
<XMLResultLocation>\\Server\RecognitionServer\ECA_Default\RS_XML</XMLResultLocation>
</ExportParams>
<Statistics TotalCharacters="33948" UncertainCharacters="542" PagesArea="31" />
</XmlResult>
```

# Výstup a kontroly

- „**Image quality check**“ – kontrola kvality skenovaného obrazu
- **Nadprahová kontrola OCR** operátorom, manuálne korekcie
- „**Preflight check**“ – kontrola validácie formátu
- „**Integrity check**“ – kontrola integrity digitálneho objektu

# Verify Conformance of PDF/

The screenshot shows the Adobe Acrobat Pro Extended interface. The title bar reads 'CCE\_COMPERS\_1997\_73\_FR.pdf - Adobe Acrobat Pro Extended'. The menu bar includes File, Edit, View, Document, Comments, Forms, Tools, Advanced, Window, and Help. The toolbar contains icons for Create, Combine, Collaborate, Secure, Sign, Forms, Multimedia, and Comment. The status bar shows '1 / 3' pages, a zoom level of '63,2%', and a 'Find' search box. A blue banner at the top of the document area states 'You are viewing this document in PDF/A mode.'

On the left, the 'Standards' panel is open, displaying the following information:

- Conformance**
  - Standard: PDF/A-1B
  - ISO Name: ISO 19005-1
  - Status: verification succeeded
  - [Verify Conformance...](#)
- OutputIntent**
  - Identifier: Generic RGB Profile
  - Info: Generic RGB Profile
  - [Open Preflight](#)

The main document content is a circular notice from the 'UNIVERSITÉ DE LUXEMBOURG' (University of Luxembourg). The text is multilingual, including:

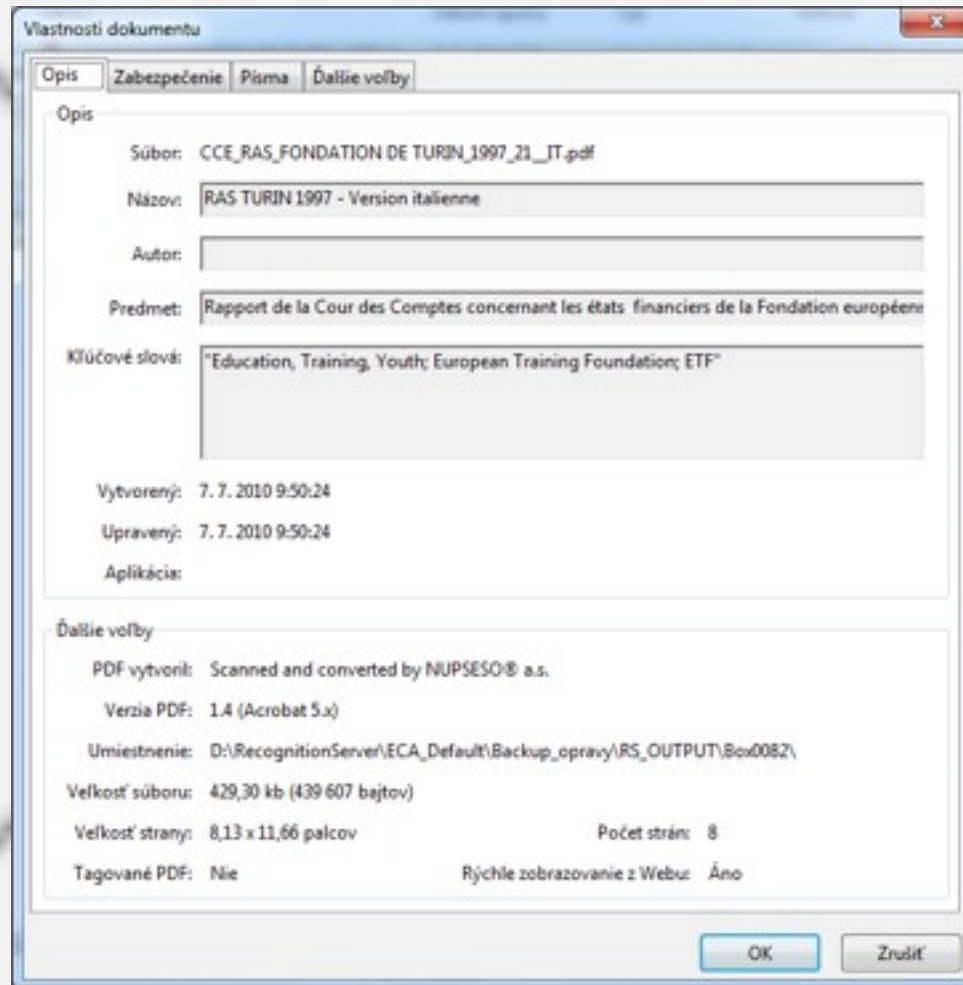
- MEDELELSE TIL PERSONALET
- MITTEILUNG AN DAS PERSONAL
- ΑΝΑΚΟΙΝΗΣΗ ΣΤΟ ΠΡΟΣΩΠΙΚΟ
- CIRCULAR
- COMMUNICATION AU PERSONNEL
- STAFF NOTICE
- COMUNICAZIONE AL PERSONALE
- MEDEDELING AAN HET PERSONEEL
- COMUNICAÇÃO AO PESSOAL
- TIEDOTE HENKILÖKUNNALLE
- PERSONALMEDDELANDE

A table below the header provides details:

Fecha	Date	N°
Date	Data	Ap18. 73-97
Datum	Ημερομηνία	Ηρ.
Päivämäärä		N°

The date is listed as 11. XI. 1997. Below the table, the subject is 'Objet : Promotions 1997.' The main body of the text reads: 'Je vous prie de trouver ci-joint, la liste des promotions que j'ai décidées en ma qualité d'AIPN pour l'exercice 1997.' and 'Les décisions formelles seront notifiées aux intéressé(e)s dans les meilleurs délais.'

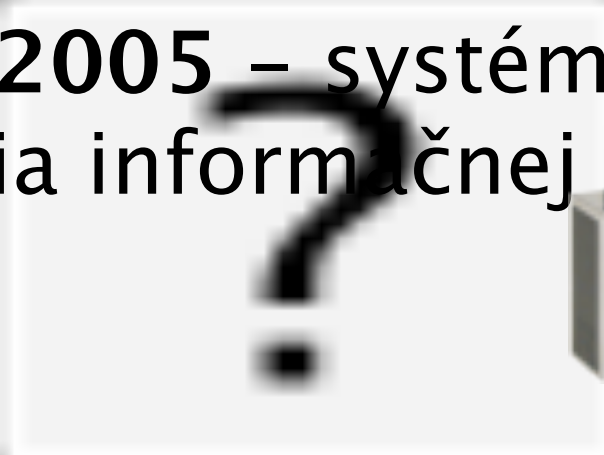
# Indexy priamo v PDF/A





# Celková kvalita a

- **ISO 9001:2008** – systém manažovania kvality
- **ISO 27001:2005** – systém manažovania informačnej bezpečnosti
- **System riadenia vstupov v objekte ID kartami**



# Zaujímavosť na záver



- **MicroSDHC**; rozmer 15x11x1mm; kapacita 32GB, cca. 500 tisíc strán (PDF/A+text; B/W; 400dpi)
- **Zápalková krabička**; rozmer 48x36x15,5mm; kapacita cca. 5TB, cca. 80 miliónov strán (PDF/A+text; B/W; 400dpi)



# Ďakujem za pozornosť

Juraj Lievaj  
NUPSESO, a.s.

[jlievaj@nupseso.eu](mailto:jlievaj@nupseso.eu)