

Ivo Zandhuis

<http://www.zandhuis.nl/>
ivo@zandhuis.nl

Armatiek BV

<http://www.armatiek.nl>
info@armatiek.nl

Handleiding EAD dienst

Versie 1.1, april 2010

Handleiding EAD dienst

Inhoudsopgave

1 Inleiding.....	3
2 Omzetten naar EAD.....	3
3 Omzetten naar PDF.....	3
4 Schermen.....	4
5 Uploaden van uw bestanden.....	5
6 Downloaden van de resultaten.....	6
7 Opbouw van het PDF LOG bestand.....	6
8 Uw profiel.....	7
9 Conventies voor MsWord documenten.....	8
10 Meldingen.....	10
11 Tips.....	11
12 Het maken van een ZIP bestand.....	12

Handleiding EAD dienst

1 Inleiding

De EAD dienst biedt de mogelijkheid om MsWord en EAD bestanden om te zetten naar EAD en PDF bestanden. Hiertoe is een website beschikbaar:

<http://www.armatiek.nl/eaddienst>

Nadat men op deze site is ingelogd kan men beide type bestanden uploaden, en wordt de omzetting op de achtergrond uitgevoerd. In de mail ontvangt men dan een link naar de plek waar het resultaat - als zip-bestand - beschikbaar is gesteld.

De omzettingen betreffen:

1. MsWord documenten – deze zijn volgens afgesproken conventies ingevoerd en worden omgezet naar EAD. Wanneer fouten optreden wordt een PDF teruggestuurd met commentaar op de plekken waar de conversie faalt.
2. EAD documenten – deze worden omgezet naar PDF. Het kan uw eigen EAD betreffen, of de EAD als resultaat van de MsWord conversie. De EAD kan middels een aparte conversie worden "voorbewerkt" voordat de PDF wordt samengesteld.

De dienst is niet gratis. Een account kan worden aangevraagd via

eaddienst@armatiek.nl

Dit start als een proefaccount, waarmee u enkele MsWord bestanden of EAD's als proef kunt uploaden. Wanneer de dienst biedt wat u zoekt wordt een offerte opgemaakt op basis van een eenmalig startbedrag en een jaarlijks abonnement.

2 Omzetten naar EAD

U kunt uw inventarissen ingevoerd in MsWord, mits deze voldoen aan enkele eisen, om laten zetten in EAD. Deze EAD voldoet aan de EAD 2002 DTD.

Alleen het inventarisdeel van de MsWord documenten wordt omgezet naar EAD. Dit is een compleet EAD document, maar zonder inleiding.

De eisen die de EAD dienst stelt aan de MsWord bestanden worden behandeld in "Conventies voor MsWord documenten".

Wanneer u reeds in het bezit bent van een EAD-bestand met basis-informatie en eventuele inleiding ("EAD-header") voor een specifieke in MsWord ingevoerde inventaris¹, kan kan deze worden samengevoegd met het EAD resultaat. Hierdoor krijgt u een complete EAD en PDF (inclusief inleiding) afgeleverd. De wijze waarop deze header moet worden meegeleverd wordt behandeld in "Uploaden van uw bestanden".

Wanneer fouten optreden worden deze als een log-bestand bijgevoegd in het resultaat. Ook worden deze in een PDF representatie van het door u aangeleverde MsWord bestand bijgevoegd als commentaar.

De wijze waarop de conversie-resultaten worden samengevoegd en afgeleverd is behandeld in "Downloaden van de resultaten".

3 Omzetten naar PDF

Uw EAD (welke is gegenereerd door de EAD dienst, of door u aangeleverd) wordt direct omgezet naar PDF, of eerst geconverteerd naar een bruikbare EAD variant en dan omgezet.

De EAD wordt omgezet naar PDF vorm die gebruikt wordt door het Nationaal Archief en als vrije software beschikbaar is gesteld:

1 Deze zou uit ABS-Archeion kunnen zijn geëxporteerd.

Handleiding EAD dienst

<http://www.taskforce-archieven.nl/dossiers/tienvragen/eaddownloads>

De aangeleverde EAD moet valide zijn. Uitgangspunt voor de conversies vormt de EAD 2002 DTD met de public identifier:

+//ISBN 1-931666-00-8//DTD ead.dtd (Encoded Archival Description (EAD) Version 2002)//EN
(<http://www.loc.gov/ead/ead2002.zip>)

Ook EAD's die voldoen aan het EAD schema worden geaccepteerd, wanneer valide:

urn:isbn:1-931666-22-9

(<http://www.loc.gov/ead/ead.xsd>)

Daarnaast moet de EAD voldoen aan de EAD richtlijnen van het NA:

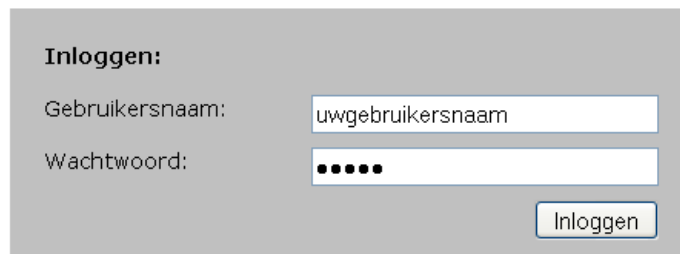
<http://www.promead.org/nationaal-archief-ead-richtlijnen>

Uiteindelijk stelt de PDF conversie enkele beperkingen aan de EAD die het gevolg zijn van keuzes, die gemaakt zijn in de ontwikkeling van de PDF generator.

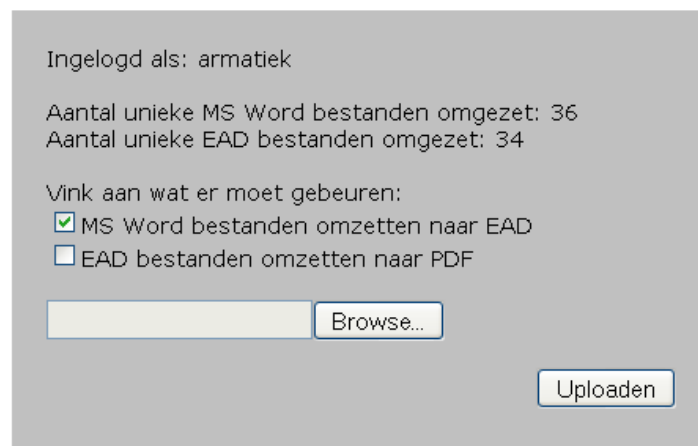
Wanneer fouten optreden worden deze als een log-bestand bijgevoegd in het resultaat.

4 Schermen

U logt in op het systeem met de aan u toegekende gebruikersnaam.



U bent nu bekend, en uw persoonlijke instellingen worden geladen. Het volgende scherm wordt getoond:



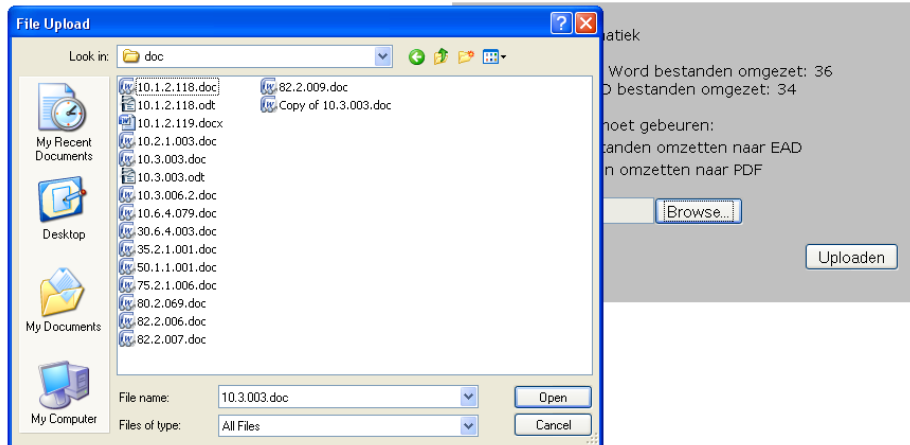
Het aantal bestanden (MsWord en/of EAD) dat u heeft aangeboden wordt op basis van de bestandsnaam door het systeem geteld. Er is geen grens aan het aantal bestanden dat u kunt converteren.

Handleiding EAD dienst

Wanneer U alleen EAD's wenst terug te ontvangen vinkt u het eerste vakje aan.

Wanneer U (ook) PDF versies van de EAD's wenst terug te ontvangen vinkt u het tweede vakje aan.

De browse knop brengt u in een dialoog zoals u dat gewend bent op uw besturingssysteem. In Windows is dat:



U kunt 1 bestand selecteren: een MsWord bestand, een EAD bestand of een ZIP bestand, zoals behandeld in "Uploaden van uw bestanden".

Vervolgens start u de conversie door op de knop "Uploaden" te klikken. U krijgt een melding te zien dat de upload is gestart. U kunt onmiddellijk daarna evt. volgende upload-opdrachten uitvoeren.

Het systeem beëindigt de sessie wanneer u lang geen interactie heeft gepleegd met het systeem. U moet dan opnieuw inloggen.

5 Uploaden van uw bestanden

Er zijn drie soorten upload mogelijk.

1. U wijst een MsWord bestand aan. U wenst dit dan om te zetten naar EAD en evt. naar PDF. U vinkt de gewenste vakjes aan:

Vink aan wat er moet gebeuren:

- MS Word bestanden omzetten naar EAD
- EAD bestanden omzetten naar PDF

2. U wijst een EAD bestand aan. U wenst dit dan om te zetten naar PDF. U vinkt minstens het tweede vakje aan:

Vink aan wat er moet gebeuren:

- MS Word bestanden omzetten naar EAD
- EAD bestanden omzetten naar PDF

3. U wijst een ZIP bestand aan. Dit bestand bevat dan een aantal bestanden dat moet worden verwerkt (zie "Het maken van een ZIP bestand" voor een korte uitleg over het maken van een zip file). De opbouw van het zip bestand *moet* -afhankelijk van wat u wilt doen- bestaan uit tenminste één van de volgende folders:

- **doc** – wanneer u meerdere MsWord bestanden wenst om te zetten naar EAD.
- **ead** – wanneer u door u beheerde EAD bestanden wenst om te zetten naar PDF.

Handleiding EAD dienst

- **header** – wanneer u de door de MsWord conversie gegenereerde EAD bestanden wenst te integreren met de header. (Let op: u kunt niet een eigen aangeleverde EAD inventaris in engere zin laten samenvoegen met de EAD header.)

De bestandsnaam van de header moet gelijk zijn aan die van het MsWord bestand. Dus bijvoorbeeld:

In folder **doc** staat **inventaris1.doc** (alleen de inventaris)

In folder **header** staat **inventaris1.xml** (alleen de header)

De geleverde ZIP bevat dan het bestand **inventaris1.xml** (volledige EAD)

U vinkt de vakjes aan die van toepassing zijn.

De header moet een valide EAD document zijn.

Bij het samenvoegen van de header met de inventaris in engere zin zoals gegenereerd door de MsWord conversie gebeurt het volgende.

- Het gegenereerde <dsc>-gedeelte op basis van het Word-bestand (*.doc) in de **doc**-folder van ZIP, wordt geplaatst op de juiste plaats in het EAD-bestand (*.xml) in de folder **header**.
- Vervolgens wordt het resulterende EAD bestand teruggegeven met de extensie *.xml in de resultaat-ZIP.
- Bij deze operatie gaan de eventuele gegevens in de <dsc> van de EAD header (inventaris in engere zin) en de gegevens *buiten* de <dsc> gegenereerd op basis van het Word bestand verloren. Die laatste gegevens betreffen de sleutelwoorden Titel=, Id= etc.. (Zie "Conventies voor MsWord documenten")

6 Downloaden van de resultaten

Het resultaat-bestand is altijd een ZIP bestand. In het ZIP bestand treft u alle bestanden aan.

Het betreft voor een aangeboden MsWord bestand met de naam **inventaris1.doc**:

- **inventaris1.ead.xml** – het EAD bestand dat is gegenereerd of samengesteld uit de MsWord inventaris en een eventueel aangeleverde header.
- **inventaris1.log.pdf** – een PDF versie van uw aangeleverde MsWord bestand, met daarin commentaar. Zie "Opbouw van het PDF LOG bestand".
- **inventaris1.ead.pdf** – de PDF versie van de EAD
- **log-word2ead.txt** – een verzameling van alle meldingen voor alle files binnen deze stap.

Het betreft voor een aangeboden EAD bestand met de naam **inventaris1.xml**:

- **inventaris1.ead.pdf** – de PDF versie van de EAD.
- **log-ead2pdf.txt** – een verzameling van alle meldingen voor alle files binnen deze stap.

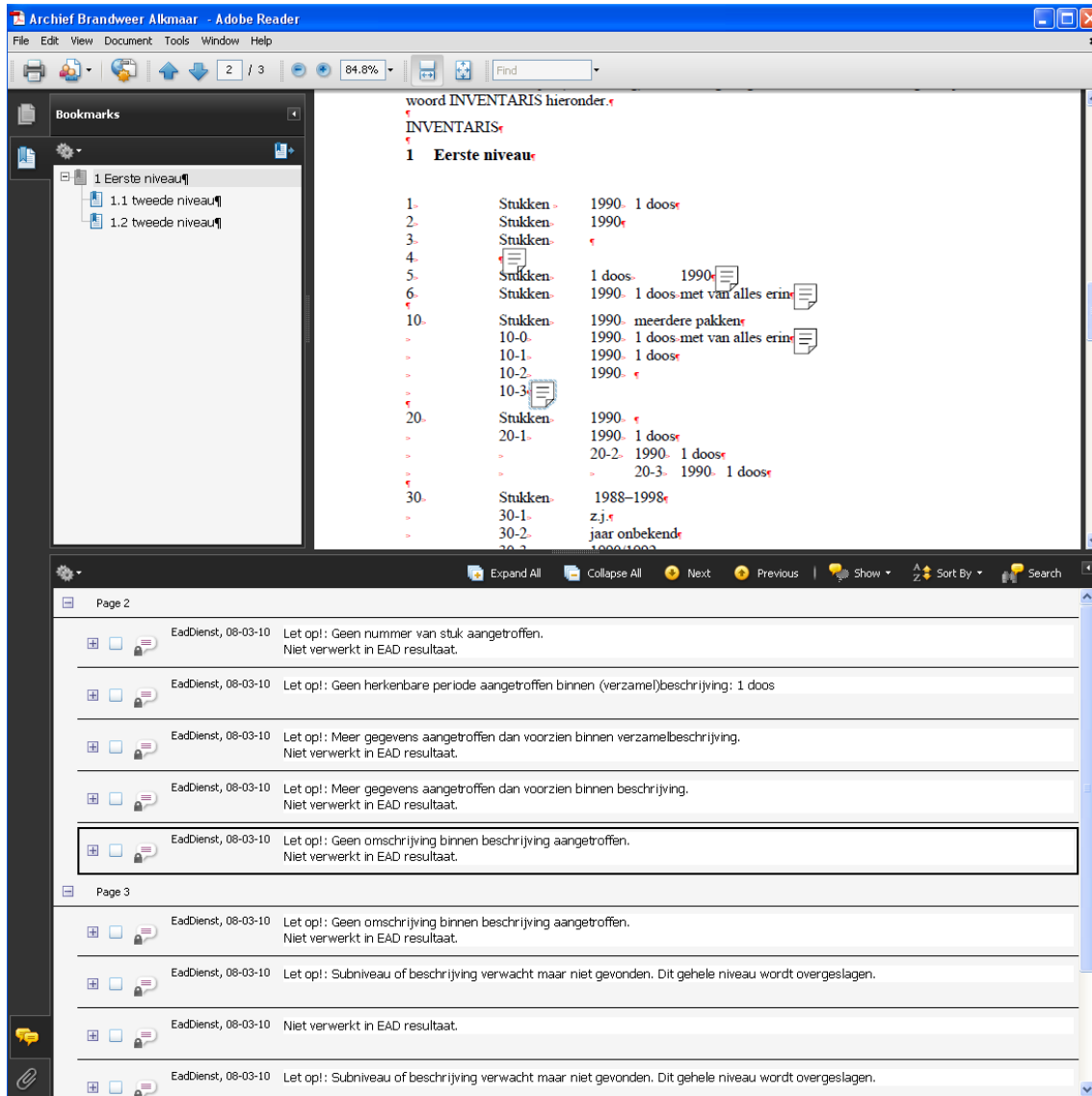
7 Opbouw van het PDF LOG bestand

De conversie naar EAD vraagt om een specifieke opbouw van de MsWord bestanden. Deze is beschreven in "Conventies voor MsWord documenten".

Handleiding EAD dienst

In het bestand ***.log.pdf** vindt u per bestand weergegeven welke problemen zijn aangetroffen in de MsWord. Op basis van deze meldingen kunt u bepalen of de conversie geslaagd is. Natuurlijk is een controle van de EAD zelf, en van de PDF resultaten, ook van belang. Maar evidente problemen in de MsWord kunt u zo snel ontdekken.

De PDF opent u in een PDF reader zoals Adobe Acrobat. U geeft hierin aan dat u de commentaren (comments) wenst te zien. Binnen Acrobat ziet dat er zo uit:



U ziet de commentaren door op de tekstballonnetjes te klikken of via menu view/navigation panels/comments te selecteren.

U kunt door de commentaren klikken of door de pagina's wandelen en met de muis over de tekstdocumentjes gaan. De melding wordt dan als "tooltip" getoond.

Zie "Meldingen" voor de lijst van mogelijk meldingen.

8 Uw profiel

Wanneer u zich abonneert op de EAD dienst wordt voor u een profiel opgesteld.

Handleiding EAD dienst

Dit profiel bevat de volgende gegevens:

- De volledige naam van Uw archiefdienst, zoals dit zal verschijnen in EAD en PDF.
- Naam en contact gegevens van uw contactpersoon.
- Of de EAD die u zelf aanlevert volgens het EAD schema of de EAD DTD is opgezet.
- De email-adressen waarnaar de resultaten moeten worden gezonden.
- De XSLT stylesheet waarmee u de door u aangeleverde EAD wilt laten omzetten naar een variant. Deze variant kan voor uw eigen beheer van belang zijn, en/of gewenst zijn voor het samenstellen van de PDF.

Deze stylesheet kan door uzelf worden aangeleverd, maar ook door Armatiek BV op verzoek worden ontwikkeld.

- Een ODF template welke gebruikt moet worden bij het genereren van de PDF. Hierin kunt U o. a. keuzes maken over font (DIN, Arial of Times Roman), opzet van de voorpagina, en uw logo.

Dit template wordt door Armatiek BV samengesteld in overleg met u.

9 Conventies voor MsWord documenten

Aangeleverde MsWord bestanden kunnen alleen de volgende MsWord versies betreffen.

- MsWord 2000 (*.doc)
- MsWord 2002/XP (*.doc)
- MsWord 2003 (*.doc)
- MsWord 2007 (*.docx)

Omzetting van de MsWord bestanden betreft alleen de inventaris/plaatsingslijst.

De inventaris moet binnen MsWord worden opgebouwd op de volgende wijze, in deze volgorde (> is: tab, *schuin* is: optioneel):

- Alinea met de exacte tekst "INVENTARIS" or "PLAATSINGSLIJST". Wanneer een van deze twee woorden niet wordt aangetroffen wordt ervan uitgegaan dat de inventaris start bij het eerst aangetroffen kopniveau.
- Daaronder per rubriek: Kopniveau. Wanneer geen enkel kopniveau wordt aangetroffen wordt kopniveau 1 verondersteld.
- Daaronder ingebedde subniveaus 2, 3, 4 etc. of beschrijvingen. Er mogen geen subniveaus worden overgeslagen, dus kop1 gevolgd door kop3 is niet goed.
- Per verzamelbeschrijving:
 - Reeks > beschrijving > *datering* > *vorm van bestanddelen*
- Per deelbeschrijving:
 - > inventarisnummer > beschrijving > *vorm*
- Een verzamelbeschrijving moet worden gevolgd door 1 of meer deelbeschrijvingen.
- Per enkelvoudige beschrijving:
 - Inventarisnummer > beschrijving > *datering* > *vorm*

Handleiding EAD dienst

Wanneer tussen deze blokken andere gegevens worden aangetroffen die niet als NB zijn herkend, wordt dit als waarschuwing aangemeld.

Een voorbeeld is (de tekens > en ¶ maken deze speciale karakters zichtbaar, en keren terug in het logbestand):

2.1 >	Verslagen van het Bestuur ¶ (kopniveau 2)	
2-3 >	Agenda's, verslagen en besluitenlijsten van de Voorbereidingscommissie.>	2001 – 2002.>
	2 omslagen¶ (verzamelbeschrijving)	
>	N.B. De Voorbereidingscommissie was het bestuur van de vereniging.¶ (nota bene)	
>	2 >	2001 > 1 omslag¶ (deelbeschrijving)
>	3 >	2002 > 1 omslag¶ (deelbeschrijving)

Let op! Het gebruik van automatische nummering in MsWord om langere reeksen beschrijvingen te nummeren voldoet dus niet. Een inventarisnummer moet tenslotte voorafgegaan worden door een tab!

Nota bene wordt herkend op basis van de letters NxBx (niet letterkast gevoelig, waarbij x is optionele combinatie van spaties en punten), evt. interpunctie, en een eventueel opvolgend nummer gevolgd door interpunctie. Varianten zijn:

- Nb 1: noottekst;
- N.B. noottekst;
- NB - noottekst; etc.

De nota bene heeft betrekking op de voorgaande (verzamel)beschrijving.

Bij het omzetten worden periodes herkend en in genormaliseerde vorm opgenomen in de EAD. Periodes moeten de vorm hebben van JJJJ maandnaam DD, een periode gescheiden door een koppelteken, of een variant daarvan. Bijv.

- 1894 april 23 wordt omgezet naar 1894-04-23
- 1894 apr. - nov. wordt omgezet naar 1894-04-01/1894-10-30

De genormaliseerde waarden zijn altijd complete en op dag nauwkeurige specificaties. Dus ook:

- 2001 wordt omgezet naar 2001-01-01/2001-12-31

Let op: deze normalisatie wordt niet zichtbaar in de PDF, maar wel in de EAD.

De volgende constructies worden expliciet uitgesloten:

- Tekst die buiten de aangegeven structuur valt. Dit omvat tevens alle andere vormen van plaatsingslijsten (zoals tabellen). Dit levert een waarschuwing op.
- Opmaak van de tekst. Dit levert geen waarschuwingen op, maar wordt stilletjes overgeslagen.
- Noten opgenomen in de lopende tekst, inline figuren. Zulke constructies worden dmv. een waarschuwing doorgegeven, en overgeslagen.

Header informatie

Wanneer de EAD is gegenereerd vanuit MsWord (dus alleen de inventaris is omgezet) wordt een header samengesteld die als "minimum" variant kan worden gebruikt bij het samenstellen

Handleiding EAD dienst

van een PDF.

U kunt een aantal gegevens mbt. de header toch invoeren om een bruikbare PDF versie samen te stellen. Dit heeft de vorm van een sleutelwoord dat start met een hoofdletter, daarna kleine letters en daaraan vast het "="-teken. Na het "="-teken staat een waarde binnen het document (bij voorkeur aan de kop van het document). Mogelijke sleutelwoorden zijn:

Titel= ...Titel van dit archief....

Als deze parameter ontbreekt: "Inventaris *documentnaam*"

Id= ... EAD ID van dit archief....

Als deze parameter ontbreekt: de documentnaam.

Inhoud= ... Kopje boven de inventaris....

Als deze parameter ontbreekt: "Beschrijving van de archiefbestanddelen"

Kortetitel= ... korte titel van het archief (bovenaan de pagina's)

Als deze parameter ontbreekt: de titel.

Auteur= ... de auteur van het archief

Als deze parameter ontbreekt: *leeg*

10 Meldingen

Mogelijke meldingen zijn de volgende. Let op: voor de oplossing van de genoemde problemen verwijzen we naar "Conventies voor MsWord documenten".

Boven in het PDF logbestand is aangegeven hoeveel problemen werkelijk zijn aangetroffen:

Hieronder zijn X (uit Y) waarschuwingen of fouten geplaatst als commentaar in de tekst.

Het maximum aantal weergegeven meldingen is 100. De laatste melding is in dat geval:

"Hierna volgende meldingen worden niet verder weergegeven".

Let op! Wanneer u geen fouten meer aantreft in het logbestand wil dit niet altijd zeggen, dat de conversie 100% geslaagd is. De EAD dienst kan niet garanderen dat alle relevante informatie is verwerkt, en dat deze goed is verwerkt. U moet dus altijd een visuele inspectie uitvoeren op de resultaat PDF, en een inhoudelijke controle uitvoeren op de EAD.

Dit kan leiden tot aanvullende correcties in het Word-document, waardoor het EAD bestand nog beter wordt.

melding	uitleg
Alinea niet herkend.	Het systeem kan niet vaststellen hoe deze alinea is opgebouwd.
Deze constructie is niet herkend, en overgeslagen	Een constructie is aangetroffen waar het systeem niet bekend mee is.
Een constructie is verwijderd uit de alinea. Check of hiermee gegevens verloren zijn geraakt.	Er is een constructie, zoals een commentaar of een bookmark verwijderd uit deze alinea. Dat hoeft niet te betekenen dat er iets mis is; de tekst van deze constructie is wel gehandhaafd.

Handleiding EAD dienst

melding	uitleg
Een lijst item wordt teruggebracht tot haar inhoud	Lijsten worden niet ondersteund; lijst inhoud wordt opgevat als gewone alinea's.
Een zacht regeleinde aangetroffen; dit wordt overgeslagen	Een zacht regeleinde ("soft return") wordt niet ondersteund. U kunt dit vervangen door een echt regeleinde.
Geen herkenbare periode aangetroffen binnen (verzamel)beschrijving	De periode van de verzamelbeschrijving moet op de 3e positie worden geplaatst; dat is hier niet gebeurd.
Geen INVENTARIS of PLAATSINGSLIJST in alinea, en geen kop-niveau alinea aangetroffen	Het systeem kan niet vaststellen waar het moet starten.
Geen omschrijving binnen beschrijving aangetroffen.	Er is geen beschrijving van het stuk aangetroffen
Leeg niveau. Dit niveau wordt overgeslagen.	Er is een niveau zonder inhoud aangetroffen.
Meer gegevens aangetroffen dan voorzien binnen beschrijving.	Er zijn meer dan 3 stukjes informatie gescheiden door tabs aangetroffen in de verzamelbeschrijving. Het mogen er maximaal 3 zijn.
Meer gegevens aangetroffen dan voorzien binnen verzamelbeschrijving.	Er zijn meer dan 4 stukjes informatie gescheiden door tabs aangetroffen in de verzamelbeschrijving. Het mogen er maximaal 4 zijn.
Niet verwerkt in EAD resultaat	Deze melding wordt overal toegevoegd waar de betreffende aangetroffen alinea niet is overgenomen in de gegenereerde EAD.
Niveau niet herkend: is hier een niveau overgeslagen?	het lijkt erop dat een niveau mist. U bent bijvoorbeeld in de tekst van kopniveau 2 naar 4 gesprongen. Dit accepteert het systeem niet. Niveaus moeten aansluitend zijn opgezet.
Tekst aangetroffen binnen dit niveau, maar geen subniveau of beschrijving. Dit niveau wordt overgeslagen.	Er is een kopniveau aangetroffen waarbinnen geen beschrijvingen of subniveaus voorkomen, maar wel tekst.
Tabellen worden overgeslagen	Tabellen worden niet ondersteund. De hele tabel wordt overgeslagen.
Voet/eindnoot verwijderd	Voet/eindnoten worden niet omgezet. Deze zijn dus verwijderd uit de tekst.

11 Tips

1. Let op de kopjes-structuur van uw document. Binnen de inventaris moeten de niveaus alleen betrekking hebben op series. Zij moeten ook correct zijn ingebed.

Een kopniveau is soms niet visueel te onderscheiden van een gewone alinea met (bijv. dikgedrukte) opmaak. Voor het EAD systeem is zo'n gewone alinea niet als kopniveau herkenbaar. De typische manier om deze kopniveaus in te voegen is dmv. toekenning van de stijl Heading 1..5 (Kop 1...5).

In MsWord kunt u de kopstructuur zichtbaar maken d.m.v. View/Document map.

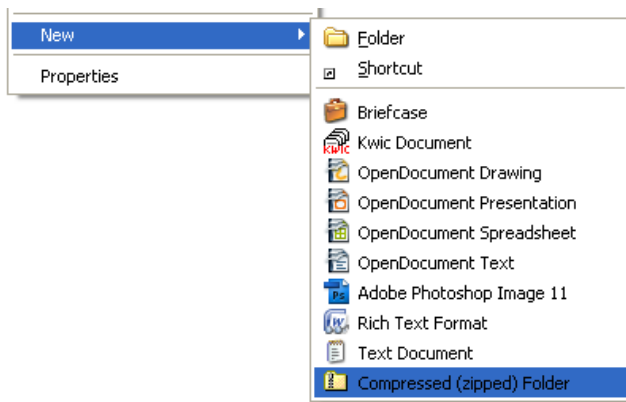
Handleiding EAD dienst

2. Houd u zich strikt aan de conventies zoals afgesproken in "Conventies voor MsWord documenten". Toon de werkelijke MsWord tekststructuur van tabs en geregeleinden door in MsWord op het Paragraaf-icoontje ¶ te klikken.
3. Kies ervoor altijd de inventaris te beginnen met de tekst INVENTARIS. Het systeem zal anders uitgaan van het eerst aangetroffen kopniveau, en dat kan abusievelijk een niveau in de inleiding zijn.

12 Het maken van een ZIP bestand

U maakt de folders **doc**, **ead** en/of **header** aan binnen een gewone folder. Daarin plaatst u (een kopie van) de te converteren bestanden.

Vervolgens maakt u een nieuw ZIP bestand. U kunt onder windows een "compressed folder" aanmaken binnen een andere (gewone) folder.



U geeft de compressed folder een naam (eindigend op .zip, bijv. **eaddienst.zip**) en kopieert vervolgens de genoemde folders naar deze compressed folder.

Uiteindelijk wijst u in het upload scherm van de EAD dienst naar het zip bestand als geheel (**eaddienst.zip**).