

**nationaal
archief**



Introductie Digitale Archieven – Jeroen van Luin – 29-03-2019

Wie ben ik?

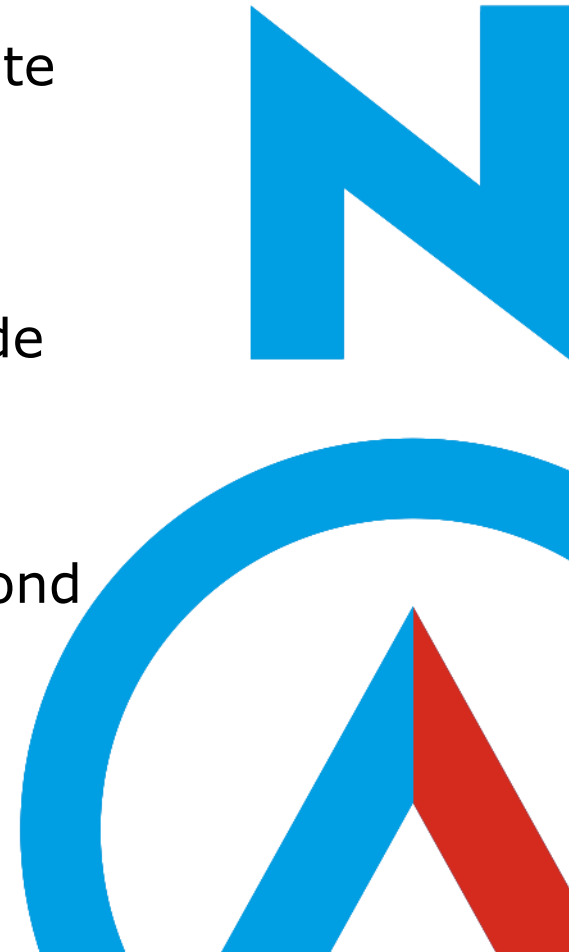
Jeroen van Luin

In 2000 afgestudeerd Technische Informatica aan de Universiteit Twente

Sinds 2006 in dienst bij Nationaal Archief

Sinds 2015 Senior Medewerker Collectiebeheer Digitale Archieven bij de afdeling Collectie van het Nationaal Archief

In 2017 de minor Archieven aan de Hogeschool van Amsterdam afgerond



Wat is archief? (1)

Archief is informatie die vastgelegd is op een medium

- Waarneming → kennis bij de waarnemer
- Kennis → informatie door het te delen, te communiceren
- Informatie → archiefstukken door het vast te leggen
- Archiefstukken → archief door het duurzaam te bewaren

Archief is procesgebonden informatie:

Binnengekomen in, gebruikt bij of opgeleverd door de uitvoering van een werkproces.

Wat is archief? (2)

Doel van archief:

1. Gebruik in de eigen dagelijkse bedrijfsvoering;
2. Aflegging van verantwoording;
3. Recht- en bewijsvinding;
4. Cultuurhistorische waarde, het persoonlijke en het gemeenschappelijke geheugen.

Wat is digitaal archief? (1)

- Digitaal archief is archief, maar dan digitaal.
- Digitaal archief is digitale procesgebonden informatie

bescheiden, ongeacht hun vorm, door een overheidsorgaan ontvangen of opgemaakt en naar hun aard bestemd daaronder te berusten

- RvS 20-03-2019: Ook WhatsApp/SMS, ook op privé-telefoons, “als deze in het kader van het werk zijn verstuurd”

Verskil analoog en digitaal archief

Er zijn ook verschillen tussen analoog archief en digitaal archief:

- Inkt zit vast op papier: de informatie en de drager zijn één geheel; digitale informatie zit los van de drager
- Inkt op papier kan direct gelezen worden; voor digitaal archief is altijd hardware en software nodig om de data weer te geven
- Digitaal archief kan ook gedrag bevatten: formules in Excel, macro's in Word, schermpjes met labels in Access, filmpjes op websites

Voor- en nadelen digitaal archief

Vergeleken met analoog archief heeft digitaal archief voor- en nadelen:

- + Digitaal archief neemt minder fysieke ruimte in, en de hoeveelheid benodigde ruimte neemt elk jaar af
- + Digitaal archief kan *full-text* doorzoekbaar zijn of worden gemaakt
- + Digitaal archief is heel gemakkelijk te kopiëren en te transporteren en via internet te raadplegen
- + De conditie van digitaal archief is automatisch te controleren
- Digitaal archief vereist altijd hardware en software om het te raadplegen
- Digitale hardware en software zijn snel aan veroudering onderhevig

Soorten digitaal archief

We kennen 2 soorten digitaal archief:

1. **Gedigitaliseerd archief**

Van een analoge gegevensdrager is een digitale reproductie gemaakt. Is de analoge drager daarna vernietigd, dan spreken we van 'vervanging' of 'substitutie'.

2. **Born-digital archief**

Het archiefstuk is digitaal gevormd en opgeslagen op een digitale gegevensdrager. Deze kan soms geprint worden.

Bij omzetting tussen papier en digitaal (digitaliseren of printen) kan informatieverlies optreden!

Wat zijn metadata?

Volgens de ISO 15489 norm:

gegevens die context, structuur en inhoud van archiefbescheiden en hun beheer door de tijd heen beschrijven

- Bij, op of aan het archiefstuk
 - Paperclips, nietjes, post-it's, etiketten of bijgeschreven aantekeningen
- Om het archiefstuk
 - Ordners, snelhechters, maar ook locatie archiefstukken ten opzichte van elkaar
- In een apart registratiesysteem,
 - Postregistratie, documentregistratie, archiefinventarissen op papieren archief

Doelen van metadata

- **Context**

Hoe zag de wereld om het archiefstuk heen er uit ten tijde van de aanmaak en gedurende diens *life-cycle*: wie, wat, waar, waarom, met welk mandaat, binnen welk werkproces, onder welke regelgeving

- **Structuur**

Wat is structuur binnen het archiefstuk zelf (interne structuur), en wat is de relatie en samenhang met andere archiefstukken (externe structuur). Binnen de externe structuur onderscheiden we de *logische* structuur (in de inventaris) en de *fysieke* structuur (in de kast).

- **Inhoud**

Waar gaat het archiefstuk over: omschrijving, datering, trefwoorden, vorm

Nederlandse Christelijke Radio Vereniging

NCRV

postbus 25000
1202 HB Hilversum
Nederland

bezoekadres:
Bergweg 30, Hilversum

telefoon (035) 6 71 99 11

telefax (035) 6 71 92 85

Het Algemeen Rijksarchief
tav: Mw. [redacted]
Prins Willem-Alexanderhof 20
2905 LM 's-GRAVENHAGE

onze ref. .AFD. CULTUUR & GODSDIENST

betreft .Langs Heilige Huisjes

Hilversum, .21 augustus 1998

Geachte mevrouw [redacted],

Zoals afgesproken treft u bijgaand 1 vhs-band aan van ons programma Langs Heilige Huisjes / afl. Den Helder.

Er komen veel positieve reacties op de uitzending binnen.

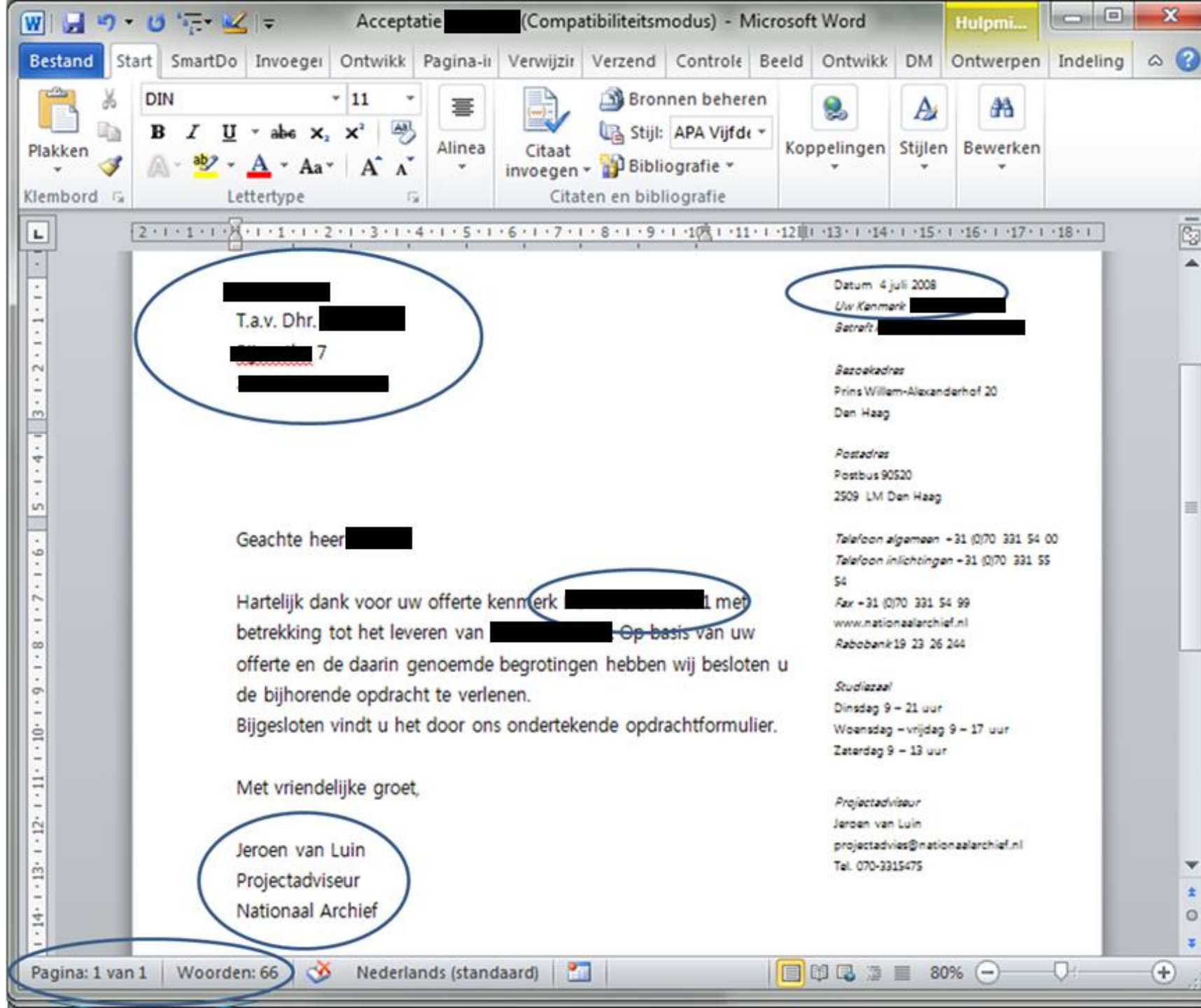
Mede namens ons regie- en redactieteam nogmaals onze hartelijk dank voor uw medewerking.

Met vriendelijke groet,

[redacted]

Metadata?





tionaal
archief



Digitale metadata

ISO 23081, is uitgewerkt in:

→ Richtlijn Metagegevens Overheidsinformatie, is uitgewerkt in:

→ Toepassingsprofiel Rijk /

Toepassingsprofiel Metagegevens Lokale Overheden

→ ToPX

- Voorbeeld machine-leesbare metadata bij vergunningaanvraag

<aggregatie>

<identificatiekenmerk>**101541**</identificatiekenmerk>

<aggregatieniveau>**Dossier**</aggregatieniveau>

<naam>**Vergunningaanvraag sloop CBS-gebouw Voorburg**</naam>

<dekking>

<inTijd>

<begin>

<jaar>**2018**</jaar>

</begin>

<eind>

<jaar>**2019**</jaar>

</eind>

</inTijd>

<geografischGebied>**Het gebied in Voorburg afgekaderd door de Prinses Beatrixlaan, de Nicolaas Beetslaan, de Prinses Irenelaan en de Bruijnings Ingenhoeslaan**

</geografischGebied>

</dekking>

<eventGeschiedenis>

<datum>**2018-05-07**</datum>

<type>**Creatie**</type>

<verantwoordelijkeFunctionaris>**Roelfsema, M.A.**</verantwoordelijkeFunctionaris>

</eventGeschiedenis>

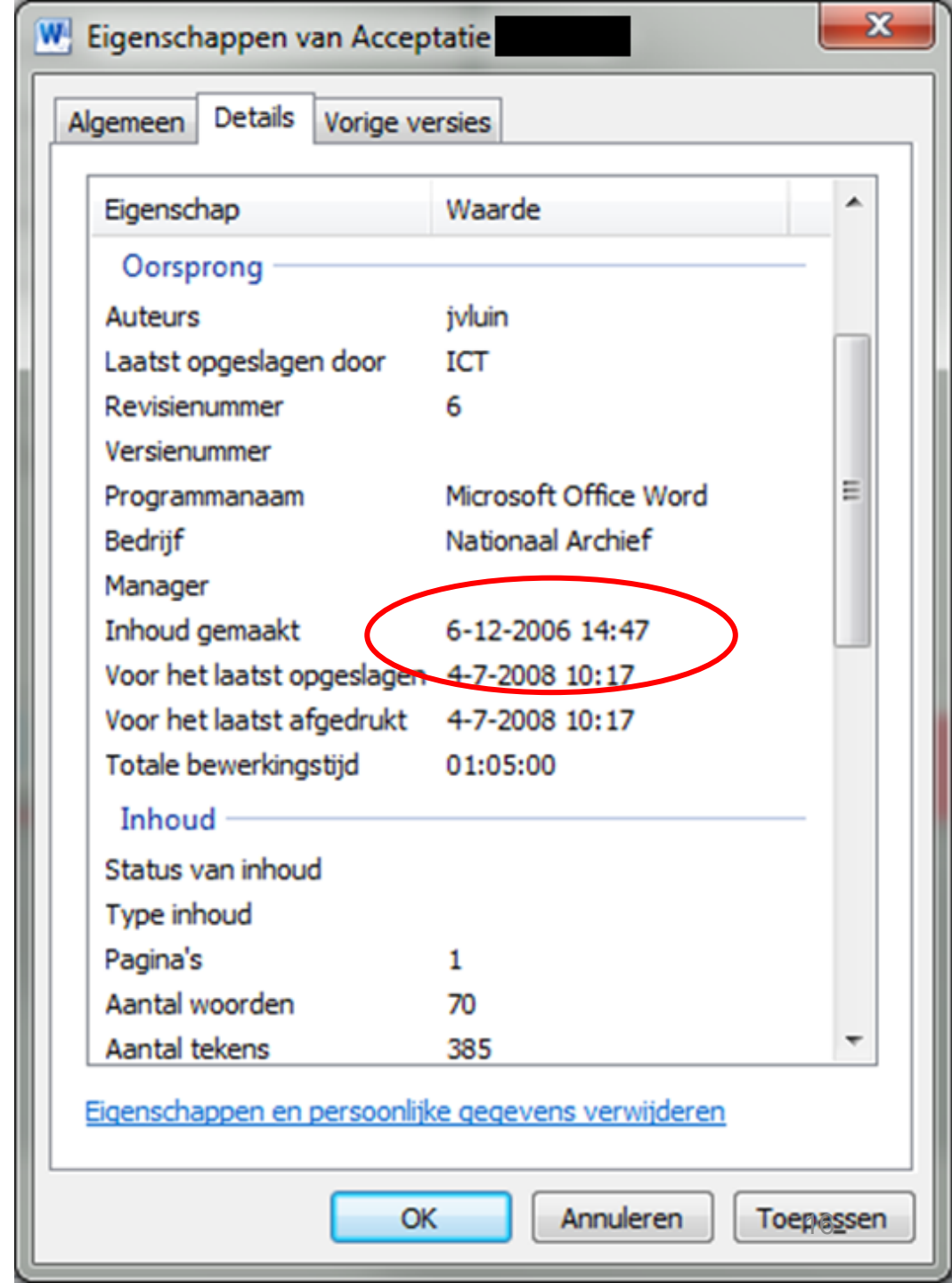
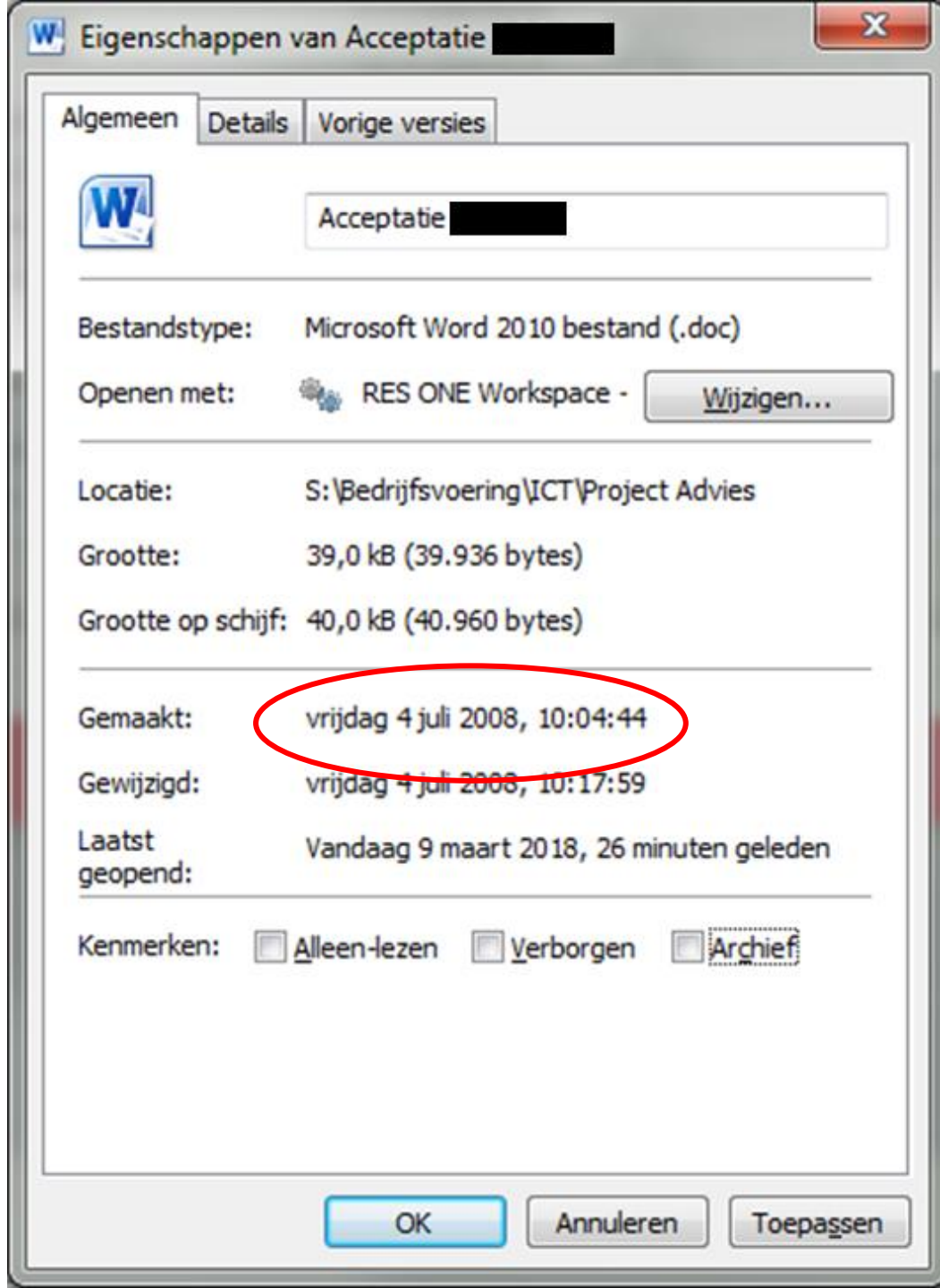
<openbaarheid>

<omschrijvingBeperkingen>**Openbaar**</omschrijvingBeperkingen>

<datumOfPeriode>

Technische metadata

- Bij digitale bestanden zit ook technische metadata
- Deze wordt automatisch toegevoegd door het systeem
- Is dus ook afhankelijk van de instellingen van het systeem
 - Bijv. een foutief ingestelde datum op het systeem
 - Bijv. de auteur van een sjabloon blijft geregistreerd als auteur van een document
 - Bijv. een document dat anderhalf jaar eerder geschreven is dan het bestand oud is, en dat ruim 1 uur bewerkt is in de 13 minuten tussen creatie en laatste wijziging.



Technische metadata — Checksums

- Wiskundige formules met twee belangrijke eigenschappen:
 1. Berekend op hetzelfde bronbestand krijg je altijd hetzelfde controlegetal
 2. Berekend op een ander bestand krijg je altijd een volledig ander controlegetal

- Bijvoorbeeld “Archief is procesgebonden informatie” in SHA-256:

2168F5B970D889A9418FB7C242AB21EC81C1E04E2BA4DAD2C494C09F8C37521B

- Maar “archief is procesgebonden informatie” (met kleine ‘a’):

7EDE0A40C8A715F2C106F1CED0C42416F26CE1221536688D22D124D12A120885

Het nut van checksums

- Checksums kunnen worden gebruikt om veranderingen aan bestanden te constateren: controle van de integriteit.
- Bijvoorbeeld:
 - Veranderingen die hebben plaatsgevonden bij kopiëren, of bij versturen over internet
 - Veranderingen doordat de gegevensdrager een fout heeft opgelopen
 - Veranderingen omdat iemand opzettelijk het bestand heeft gewijzigd

Wat is een e-depot?

- Een e-depot is een speciaal soort archiveringsstelsel
- Een archiveringsstelsel is een speciaal soort informatiesysteem
- Dus beginnen we met de vraag: wat is een informatiesysteem?

(Dit loopt iets vooruit op de module 'Archiveringsstelsels' in het volgende blok)

Wat is een informatiesysteem?

Het geheel van bestanden, metadata, procedures, apparaten en daarbij benodigd hulpmiddelen, ingericht door een persoon, groep personen of organisatie ten behoeve van de uitvoering van zijn of haar taken

- Registratie en opslag
- Vastleggen van de 'opsteller'
- Toekennen toegangsrechten
- Documenten zoeken, delen, routeren, advies vragen, etc.
- Uitchecken en inchecken, versies maken

Wat is een archiveringssysteem?

Het geheel van documenten, metadata, processen, methoden, procedures, kennis, regels, middelen en mensen waarmee een persoon of organisatie zich voorziet van betrouwbare en duurzame informatie ten behoeve van bedrijfsvoering, herinnering en verantwoording.

Extra t.o.v. informatiesysteem:

- Bevriezen van documenten en hele dossiers
- Bijhouden van de audit-trail over alles wat er met een document of dossier gebeurt, door wie dat is gedaan en wanneer dat is gedaan.
- Bijhouden van extra metadata (bijv. bewaartermijn, classificatie, vertrouwelijkheid)

Wat is nu een e-depot?

Het geheel van organisatie, beleid, processen en procedures, financieel beheer, personeel, databeheer, databeveiliging en aanwezige hard- en software, dat duurzaam beheren en raadplegen van te bewaren digitale archiefbescheiden mogelijk maakt.

Extra t.o.v. archiveringsstelsel: preservation-functies voor lange termijn duurzame toegankelijkheid van bestanden (≥ 10 jaar).

Let op: de term "e-depot" wordt ook vaak gebruikt voor alleen het softwarepakket dat onderdeel is van het geheel: e-depot in "nauwe" zin.

Waarom een e-depot? (1)

- 4 bedreigingen van toegankelijkheid van digitaal archief

1. Digitale opslagmedia kunnen verouderd raken

- **Probleem:**
Lezers voor floppy disks, diskettes, zip-disks en cd-roms zitten vaak niet meer in de huidige computers
- **Oplossing:**
Overzetten naar moderne opslagmedia, wanneer die moderne opslagmedia zelf weer verouderd zijn, de informatie weer verder verhuizen

Waarom een e-depot? (2)

2. De bits en bytes van een bestand kunnen beschadigd raken

Digitaal bestand: bitstream op een gegevensdrager of in een netwerk

- **Probleem:**
Tijdens de opslag en tijdens het transport kunnen bitstreams beschadigd raken, waardoor informatie verloren gaat
- **Oplossing:**
Een e-depot bewaart meerdere kopieën (minstens 3) van een bitstream op meerdere locaties en vergelijkt deze regelmatig (bijv. m.b.v. checksums). Bij geconstateerde fout wordt een goed exemplaar teruggezetzet.

Waarom een e-depot? (3)

3. De hardware en software kunnen verouderd raken

- **Probleem:**

Levensduur van hardware en software is relatief kort.

20 jaar terug: documenten in WordPerfect, spreadsheets in Lotus 123 en databases in DBase 3

Nu: PDF, of Word, Excel en Access, of OpenOffice versies hiervan

- **Oplossing:**

Migratie: omzetten naar ander formaat (bijv. Word → PDF)

Emulatie: nabootsen van oude hardware / software [\(voorbeeld\)](#)

Waarom een e-depot? (4)

4. De gebruiker verandert

- **Probleem:**
Gebruikers van het archief veranderen. Terminologie die nu gebruikt wordt, zal door het publiek van over 100 jaar wellicht niet meer gesnapt worden.
- **Oplossing:**
Archiefdiensten moeten hun publiek blijven kennen en ondersteunen, met extra uitleg, trainingen, beschrijvingen, woordenboeken, onderzoeksgidsen.

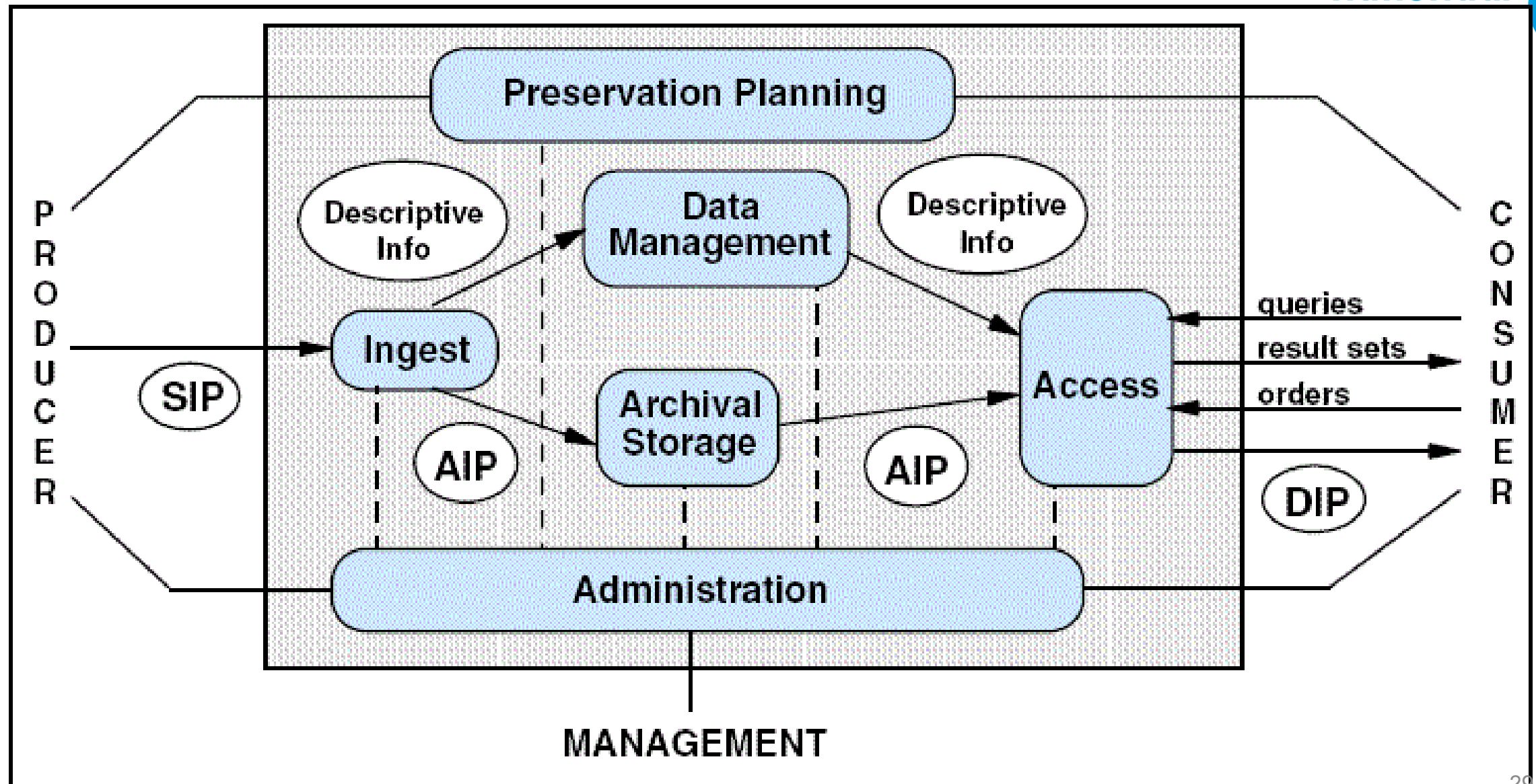
Norm voor e-depots: Open Archival Information Systems

- Oorspronkelijk ontworpen in 2000 voor archieven m.b.t. ruimtevaart
- Nu een ISO-norm (ISO 14721) die beschrijft hoe een OAIS e-depot moet werken
- De norm beschrijft:
 - De verantwoordelijkheden van een OAIS e-depot
 - Een functioneel model voor de functies waaraan zo'n e-depot moet voldoen
 - Een begrippenlijst voor veel voorkomende onderdelen van een e-depot
 - De omgeving van het e-depot: de archiefvormers, het management van het e-depot en de gebruikers van het e-depot.

OAIS verantwoordelijkheden

1. Zorgen voor goede afspraken met de archiefvormers zodat alle benodigde informatie wordt aangeleverd bij een overbrenging of uitplaatsing.
2. Zorgen voor voldoende (juridisch) mandaat om het beheer te kunnen en mogen uitvoeren.
3. Bepalen van de beoogde gebruikersgroep en de verwachting over de voorkennis van deze gebruikers
4. Hebben van een duidelijk beleid met goed beschreven procedures en processen
5. Zorgen voor voldoende toegankelijkheid van de informatie in het e-depot
6. Bewaren en kunnen aantonen van de authenticiteit van stukken in het e-depot

OAIS Functioneel Model



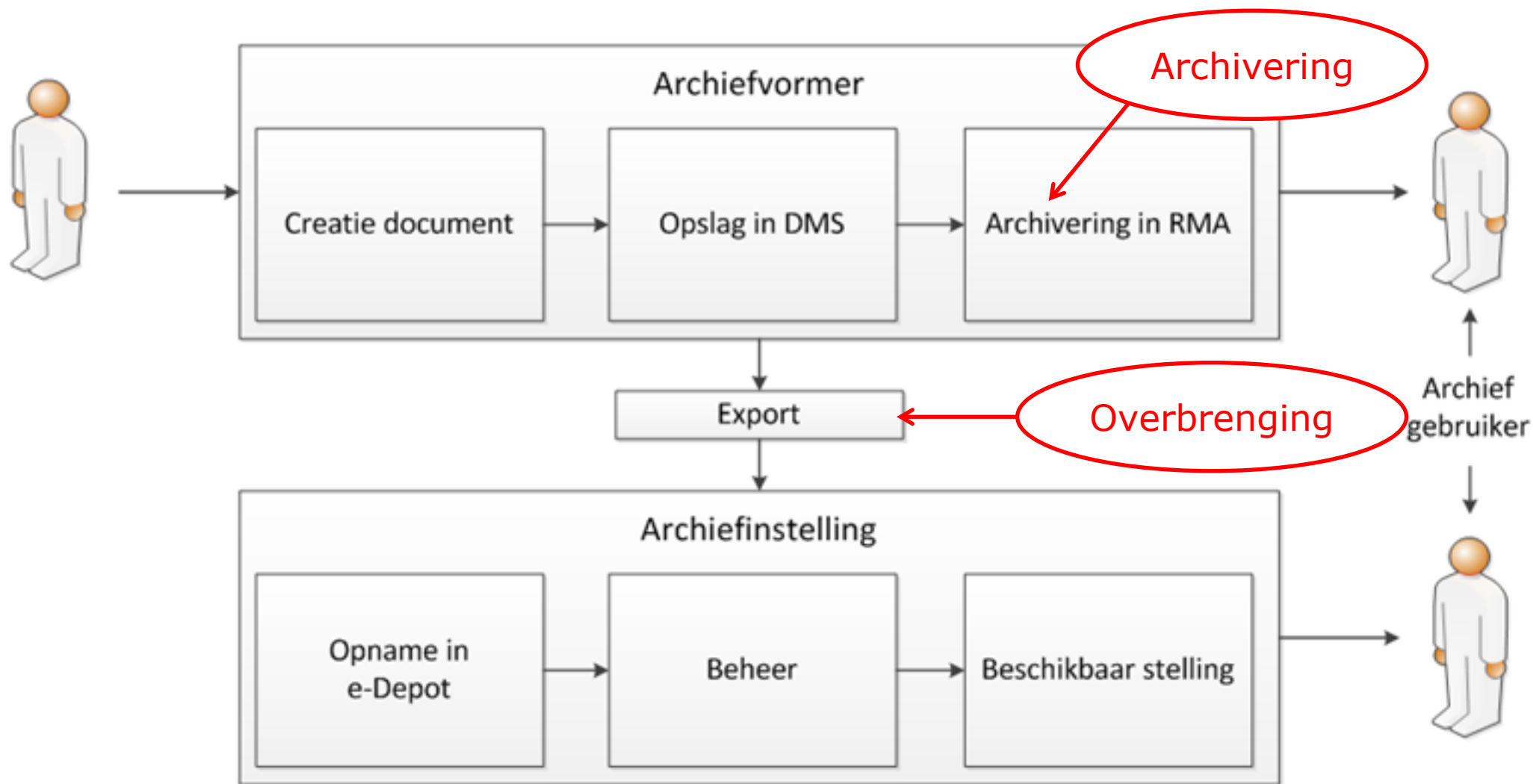
OAIS-begrippen

- **SIP** = Submission Information Package, een informatiepakket dat opgenomen gaat worden in het e-depot
- **AIP** = Archival Information Package, een informatiepakket dat beheerd wordt in het e-depot
- **DIP** = Dissemination Information Package, een informatiepakket dat verspreid gaat worden
- **Ingest** = het proces van opnemen van een SIP in het e-depot
- **Access** = het zorgen dat bestanden op aanvraag van de gebruiker vindbaar, bruikbaar en interpreteerbaar zijn
- **Preservation (Planning)** = (het plannen van) acties om bestanden duurzaam toegankelijk te houden d.m.v. bijv. migratie en emulatie

Designated Community

- De gebruikers van de informatie in een e-depot beschikken wel of niet over bepaalde voorkennis.
- Bijv.: kunnen lezen, kennis van Engels, Frans, Latijn of Papiamentu, weten hoe je een PDF-bestand opent, weten hoe je zoekt met *wildcards*.
- Welke gebruikers je wilt bedienen, en van welke voorkennis je uitgaat, bepaalt de gebruikersgroep op wie je jouw e-depot richt: de *designated community (DC)*.
- Bij alle e-depot informatieobjecten die niet zondermeer door een lid van de DC kunnen worden geïnterpreteerd, is extra beschrijvende info nodig.
- De Designated Community verandert door de tijd heen.

Proces van overbrenging



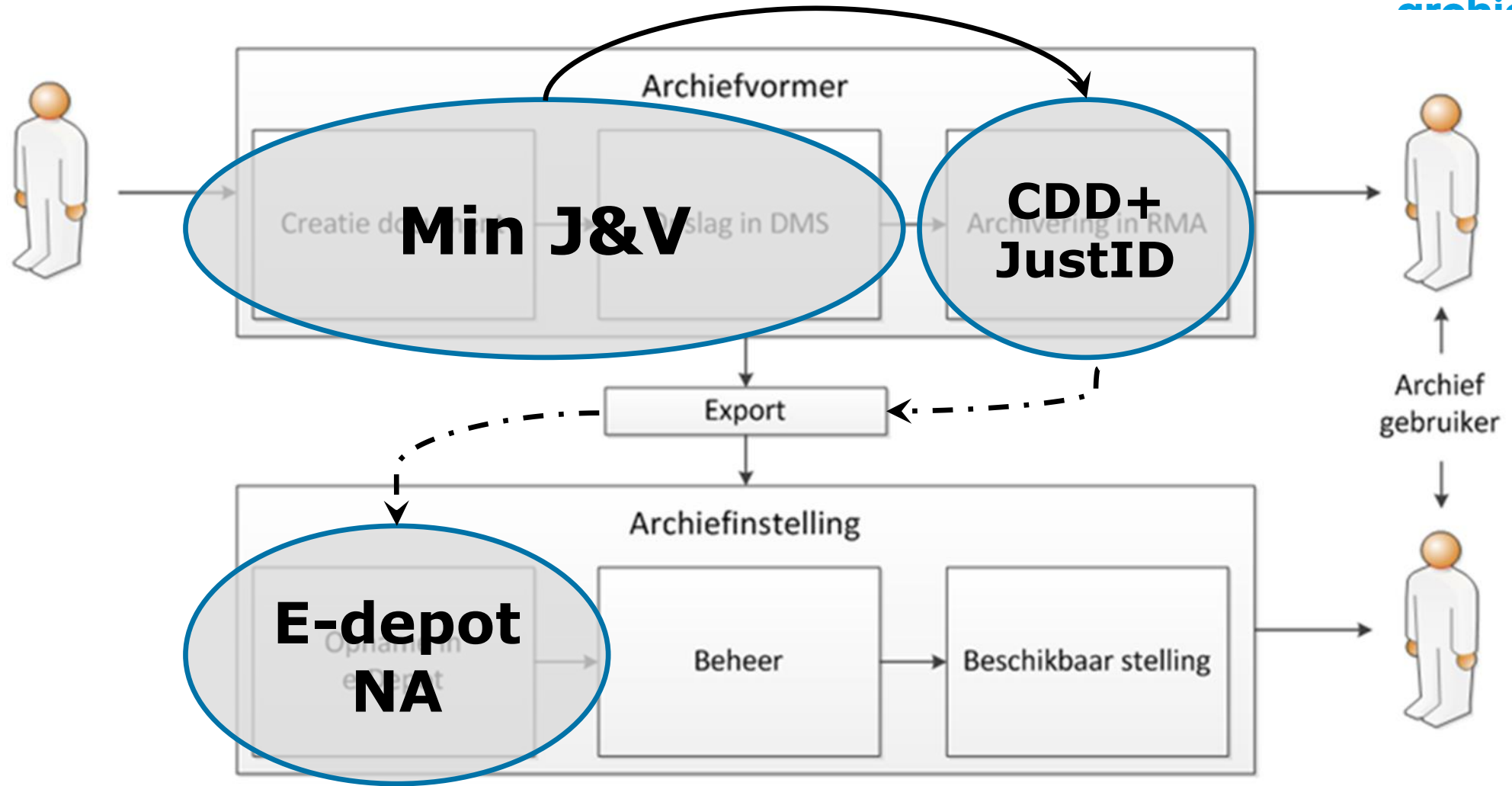
Focus nu op proces van opname in e-depot

- Overbrenging van DMS / RMA naar e-depot is een kritiek moment
- Na overbrenging worden de bestanden uit het bronsysteem verwijderd
- Fouten tijdens de overbrenging die niet opgemerkt worden zijn moeilijk of helemaal niet meer te herstellen
- Eenmaal goed opgenomen kan het beheer en de beschikbaar stelling altijd nog worden verbeterd.

Aansluiting bronsysteem op e-depot

- Bronsysteem = informatiesysteem of archiveringsysteem waarin de over te brengen stukken zitten
- Bronsystemen zijn ontwikkeld om informatie in op te slaan
- Exporteren voor overdracht naar e-depot is vaak geen bestaande functie
- Hulp van leverancier is nodig om export-functie toe te voegen
- Vraag: Wat is er nodig om het bronsysteem aan te sluiten op het e-depot?
- Antwoord komt via een impactanalyse

Voorbeeld



Deelnemers impactanalyse

- Vanuit NA:
 - Recordkeepers (voor metadata-mapping)
 - Digitale archivarissen (voor context, structuur en inhoudelijke beschrijvingen)
 - Preservation onderzoekers (voor bestanden)
 - Projectleider
 - Nu nog: ontwikkelaars E-depot (voor evt. aanpassingen ingest-workflow)
- Vanuit J&V:
 - Specialisten van de afdeling die het archief gevormd hebben
- Vanuit JustID:
 - Specialisten van de CDD+-applicatie

Opbrengst van een impactanalyse

Antwoord op de vragen:

- Wat moet er gebeuren om het archiefsysteem aan te sluiten op het e-depot?
- In welke volgorde moet dat gebeuren?
- Welke randvoorwaarden zijn van toepassing?
- Welke mensen (rollen/functies/kennis) zijn hierbij nodig?
- Hoeveel tijd is hiervoor nodig?
- Wat zijn de kosten voor de archiefvormer en de archiefdienst?

De praktijk

- Demo e-Depot (werkt alleen binnen netwerk Nationaal Archief)
- [Deltaprogramma op GahetNA](#)
- [Website Minister van Boxtel](#)
- Sneak preview:
 - Deltaprogramma “nieuwe stijl” (werkt alleen binnen netwerk Nationaal Archief)
 - Schepenbank “nieuwe stijl” (werkt alleen binnen netwerk Nationaal Archief)
- [Open Data Indexen](#)
- [Open Data Inventarissen](#)

En nu verder?

- Informatieobjecten met een documentvorm die gegroepeerd zijn in dossiers kunnen we goed opnemen en ontsluiten
- Maar... informatie heeft steeds vaker geen klassieke “document-in-dossier”-vorm meer, bijvoorbeeld:
 - Databases
 - Websites
 - Websites die deels gevuld worden met de inhoud van een database
 - Bijvoorbeeld: [Energielabel](#) (is geen document maar een website)
 - E-mailarchieven