



Parkeerprocedure

Procedure voor het tijdelijk bewaren van
voorlopig geaccepteerde digitale archieven

| | |
|--|-------------------------------|
| Opgesteld door Regionaal Archief Alkmaar | |
| Datum | 16-4-2019 |
| Versie | 2.2 |
| Auteurs | Twan Mars Anje van der Lek |

Aanleiding

In 2018 heeft het Regionaal Archief Alkmaar een e-depotvoorziening aangeschaft. Digitale informatie kan worden opgenomen in dit e-depot als de aangeboden dataset voldoet aan bepaalde aanleververeisten. Archiefvormers zullen nog niet altijd in staat zijn om datasets in deze vorm aan te bieden. In sommige gevallen zullen wij als Regionaal Archief Alkmaar deze informatie toch in ontvangst nemen. Op korte termijn gaat het om de dataset "Videotulen" van de gemeente Den Helder en de webarchieven van de voormalige gemeenten Schermer en Graft de Rijk.

Ook datasets van particuliere archiefvormers kunnen in aanmerking komen. Aan particuliere archiefvormers kunnen wij niet dezelfde eisen aan een goede, geordende en toegankelijke staat stellen als aan de bij ons aangesloten overheden, omdat zij niet vallen onder archiefwet- en regelgeving.

Doel / afbakening

Deze procedure is geschreven om ervoor te zorgen dat datasets die nog niet opgenomen kunnen worden in het e-depot op een beheersbare en veilige manier in ontvangst worden genomen en worden beheerd. Het gaat dus om een tijdelijk beheer; het zogezegd "parkeren" van de dataset bij het Regionaal Archief Alkmaar tot deze geschikt is gemaakt voor opname in het e-depot. Wij gaan gedurende deze "parkeerperiode" uit van een beheerregime van bitpreserving en een verkorte beschrijving van de dataset. Toegankelijkheid voor het publiek is tijdens de parkeerperiode niet aan de orde.

Samenvatting procedure

- Het RAA besluit of een dataset in aanmerking komt voor de Parkeerprocedure.
- De archiefvormer, ICT leverancier of derde partij maakt een export van de over te dragen dataset.
- De archiefvormer levert de dataset met een ingevuld formulier "Beschrijving voorlopig geaccepteerde dataset" in. Zie het formulier in Bijlage 1.
- Het RAA registreert de aanwinst in het excelbestand "Parkeerschijf".
- De dataset wordt gekopieerd naar 2 lokaties: Bewaarlokatie 1 (NAS-schijf 1), en de Werklokatie (NAS-schijf 2) en met het formulier en eventuele documentatie opgeslagen volgens een vaste mappenstructuur en naamsconventie.
- De originele drager wordt bewaard als 2e bewaarkopie / backup. Indien data niet via een fysieke drager is geleverd, of als de drager niet geschikt is als backup, wordt een 2e bewaarkopie op een externe harde schijf gemaakt.
- De integriteit van de bewaarkopieën en de werkkopie wordt gecontroleerd aan de hand van de bestandseigenschappen / properties.
- Gedurende de bewerking wordt periodiek gecheckt wat de status is van de bewerking. Het is niet wenselijk dat datasets langdurig op deze manier worden bewaard.
- Na een succesvolle opname in het e-depot worden de bewaarkopieën en de werkkopie verwijderd / gewist.

Werkafspraken

Vorbereiding

Met de aanbieder van de dataset worden afspraken gemaakt over op welke wijze de dataset wordt aangeboden. De aanbieder kan een andere persoon / organisatie zijn dan de archiefvormer, als de export bijvoorbeeld door de ICT leverancier of een derde partij wordt gemaakt.

Het formulier "**Beschrijving voorlopig geaccepteerde dataset**" (zie Bijlage 1) wordt ingevuld door de archiefvormer en gecontroleerd en aangevuld door de medewerker e-depot van het RAA. Het RAA vult in elk geval de reden in waarom deze dataset voorlopig wordt geaccepteerd. Eventueel aanvullende beschrijvende informatie over de dataset kan in aparte documenten worden aangeleverd (zoals: datamodellen, overzichten, eigentijdse toegangen etc).

Met de archiefvormer wordt afgesproken of de brondata wordt verwijderd, of dat deze beschikbaar moet blijven gedurende de periode vóór de formele opname in het e-depot.

Overdracht

Net als in de papieren situatie zijn er digitale archieven in allerlei soorten en maten. De datasets houdende deze digitale archieven zijn afkomstig van overheid en particulieren. Ze kunnen volledig voorbereid en gereed zijn voor opname in het E-depot (dan komen ze niet in aanmerking voor de parkeerprocedure) of ze kunnen, zoals bij papieren archieven, ongestructureerd in een digitale bananendoos worden aangeleverd.

Voor deze procedure onderscheiden we twee manieren waarop een digitaal archief kan worden overgedragen:

1. Via een externe drager (externe harde schijf, USB, CD-ROM, Floppy, e.d.)
2. Via een File Transfer Protocol (FTP) of andere gegevensoverdracht (WeTransfer, Dropbox, e.d.)

Ontvangst

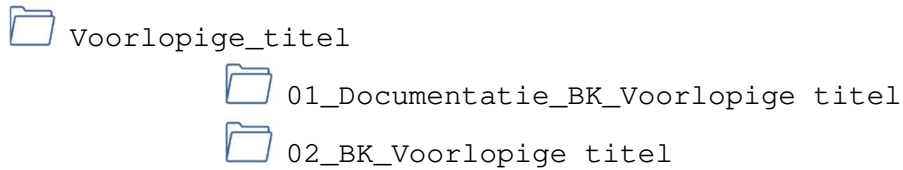
Bij beide vormen van overdracht worden de datasets op drie verschillende lokaties opgeslagen. Dit zijn altijd twee bewaarkopieën van de dataset aangeleverd door de archiefvormer (BK1 en BK2) en een werkkopie (WK). De werkkopie wordt gebruikt om bewerkingen uit te voeren en de dataset klaar te maken voor opname in het e-depot.

Het kopiëren van drager naar NAS-schijven kan het beste uitgevoerd worden op een werkstation / laptop die op dat moment niet voor andere taken wordt gebruikt. Het kopiëren van grote datasets kan lange tijd duren.

De medewerker e-depot van het RAA voert na ontvangst de volgende stappen uit:

1. De dataset wordt geregistreerd in het formulier "Parkeerschijf", opgeslagen op NAS2_Werklokatie.
2. *Optioneel:* Als het meerwaarde geeft, worden er foto's gemaakt van de individuele drager(s) en de wijze waarop de data wordt aangeleverd (de spreekwoordelijke bananendoos). Bijvoorbeeld als het op veel verschillende dragers wordt aangeleverd, er etiketten op de dragers staan, etc.

3. Op NAS1_BEWAARLOKATIE wordt een mappenstructuur aangemaakt volgens de volgende naamsconventie:



Bij voorlopige titel wordt de in het formulier opgegeven voorlopige titel gebruikt (eventueel in afgekorte vorm; bijvoorbeeld: VideotulenDH)

4. Van de data op de originele drager wordt een kopie gemaakt (exacte kopie van de data in de structuur zoals deze op de drager staan) en deze wordt opgeslagen in de map "02_BK_Voorlopige titel". Zie "instructie kopiëren via opdrachtprompt" in Bijlage 2. Indien de data op verschillende dragers wordt aangeleverd, wordt per drager een kopie gemaakt, die in aparte submappen onder "02_BK_Voorlopige titel" worden opgeslagen.
5. Aan de hand van de "eigenschappen" wordt de integriteit van de kopie gecontroleerd. (Dit gebeurt dus niet aan de hand van een checksum, zie kader). Het aantal bytes, aantal bestanden en aantal mappen van de kopie moet precies overeenkomen met die eigenschappen op de originele drager. Deze controle kan worden herhaald. De eigenschappen worden ook ingevuld op het formulier "**Beschrijving voorlopig geaccepteerde dataset**", op elk moment kan de integriteit op deze manier worden vastgesteld.

Waarom wij niet hebben gekozen voor het maken van een checksum

In het e-depot is een MD5 checksum het middel om de integriteit van bestanden te controleren. De checksum controle is preciezer dan de controle op basis van eigenschappen. Het zou kunnen dat het aantal bytes niet veranderd als er kleine veranderingen plaatsvinden, de checksum zal altijd veranderen als er ook maar één byte omvalt.

Toch is voor deze procedure bewust gekozen om controle op basis van een checksum achterwege te laten. Hier hebben we voor gekozen omdat het niet gemakkelijk is om een checksum te maken voor een dataset bestaand uit vele bestanden (files) en mappen (folders). Een checksum kan alleen worden berekend van een enkelvoudig digitaal bestand (file), niet op een map (folder).

Door een dataset bestaande uit verschillende bestanden/ mappen in een containerbestand te slaan (bijvoorbeeld een zip-file of bagit) zou een checksum over de hele dataset kunnen worden gemaakt. Het zippen van de dataset in één containerbestand brengt echter risico's met zich mee op schade aan de bestanden. Bovendien is het niet praktisch werkbaar: het zippen van grotere datasets duurt heel erg lang, met kans op fouten / vastlopen.

Het controleren op basis van bestandseigenschappen van de gekopieerde mappen is een veel snellere methode, die voor het doel van deze procedure geschikter is.

Indien de aanbieder al een of meerdere checksums heeft gemaakt, kunnen deze natuurlijk gebruikt worden voor integriteitscontrole.

Wij onderzoeken een methode om met een script van meerdere bestanden tegelijk checksums aan te maken. Als dit een werkbare methode oplevert, zal dit als extra check in de procedure op worden genomen.

6. Het ingevulde formulier "**Beschrijving voorlopig geaccepteerde dataset**" wordt als PDF/A opgeslagen in de map "01_Documentatie_Voorlopige titel" volgens de volgende naamsconventie:

FORM_Voorlopige titel.pdf

7. Eventuele andere documentatie en foto's van de dragers worden opgeslagen in de map "01_Documentatie_Voorlopige titel"

Aanmaken Werkkopie

Vanaf NAS1 of de originele drager wordt een kopie overgezet naar onze tweede NAS omgeving. Deze omgeving is de enige lokatie waar de dataset kan worden bewerkt met het doel een voor het E-depot geschikte versie van de dataset te maken. Bij fouten in deze versie van een dataset wordt de gehele dataset verwijderd en opnieuw van de NAS1_BEWAARLOKATIE gehaald. De fysieke drager wordt tot de completering van de voor het E-depot geschikte dataset bewaard in onze kluis.

8. Op NAS2_WERKLOKATIE wordt een mappenstructuur aangemaakt volgens de volgende naamsconventie:

```
└─ Voorlopige_titel
   └─ 01_Documentatie_Voorlopige titel
      └─ 02_WK_Voorlopige titel
```

9. De bij punt 4 gemaakte kopie wordt opnieuw gekopieerd en opgeslagen in de map 02_WK_Voorlopige titel
10. Aan de hand van de "eigenschappen" wordt de integriteit van de 2^e kopie gecontroleerd: het aantal bytes, aantal bestanden en aantal mappen van de werkkopie moet precies overeenkomen met die eigenschappen op de originele drager en de bewaarkopie. Deze controle is eenmalig. Als er bewerkingen op de werkkopie zijn uitgevoerd, zullen de eigenschappen veranderen.
11. Het ingevulde formulier "**Beschrijving voorlopig geaccepteerde dataset**" (FORM_Voorlopige titel.pdf) wordt opgeslagen in de map "01_WK_Documentatie_Voorlopige titel".
12. Eventuele andere documentatie en foto's over de dataset worden opgeslagen in de map "01_WK_Documentatie_Voorlopige titel".

Aanmaken 2^e bewaarkopie

De tweede bewaarkopie dient als back-up, en wordt op een andere plaats bewaard dan de ruimte waar de NAS-schijven staan (serverruimte). De back-up hoeft niet via het netwerk benaderbaar te zijn.

13. De originele drager kan dienen als 2^e bewaarkopie. In de volgende gevallen kan dit niet en wordt een 2^e bewaarkopie op een externe harde schijf gemaakt:
 - Als de originele drager niet meer gelezen kan worden met de huidige standaardapparatuur (Bijvoorbeeld floppy disc, CD ROM) of van slechte kwaliteit is.
 - Als de originele drager wordt gebruikt om de dataset uit te wisselen met een derde partij die de dataset zal analyseren of bewerken.
 - Als de dataset niet op een fysieke drager is overgedragen, maar via een dataverbinding.
14. De 2^e bewaarkopie wordt gelabeld en met een geprinte versie van het formulier **“Beschrijving voorlopig geaccepteerde dataset”** opgeborgen. De fysieke vindplaats wordt geregistreerd in het bestand “Parkeerschijf”.

Bewerken dataset / opnemen dataset in E-depot

15. De dataset (werkkopie) wordt bewerkt om geschikt te maken voor opname in het e-depot. De bewerking kan bestaan uit het toevoegen / verrijken van metadata, het veranderen van de mappenstructuur, nader onderzoek naar de inhoud, etc. De bewerking zal gebeuren aan de hand van een projectplan of bewerkingsplan. Dit valt buiten de scope van deze procedure.

De bewaarkopieën worden niet bewerkt of gewijzigd, en worden alleen geraadpleegd als het tijdens de bewerking nodig is om werkkopie geheel of gedeeltelijk te herstellen, of tussentijdse kwaliteitscontroles te doen.

16. Periodiek, tenminste 2 keer per jaar, wordt door de coördinator Collectiebeheer geëvalueerd wat de status is van voorlopig geaccepteerde datasets en of er voortgang is in de bewerking / opname in het e-depot.

Verwijderen BK en WK

17. Nadat de dataset succesvol is opgenomen in het e-depot, worden de bewaarkopieën en de werkkopie verwijderd, en kan de originele drager worden gewist of vernietigd. Verwijdering of vernietiging kan plaatsvinden nadat de ingest in het e-depot is geëvalueerd en de coördinator Collectiebeheer akkoord heeft gegeven.

Bijlage 1: Beschrijving voorlopig geaccepteerde dataset (Parkeerprocedure)

Voorlopige Titel

Korte omschrijving, te gebruiken als werktitel. Bijv: "website Graft de Rijk", "Videotulen Gemeente Den Helder", "Digitale aanvulling op archief X"

Archiefvormer

Door wie is de informatie opgemaakt/ontvangen? Gemeente, bedrijf, organisatie, familie of persoon

Periode

Datum creatie oudste informatieobject
dd-mm-jjjj

T/m Datum creatie jongste informatieobject
dd-mm-jjjj

Informatie betreft verwerving

Contactperso(o)n(en)

Contactpersoon + contactgegevens van persoon die namens de archiefvormer of andere instantie die de dataset overdraagt. Indien van toepassing kunnen meerdere contactpersonen worden genoemd, van archiefvormer, leverancier bronstelsel, extern adviseur etc

Datum ontvangst voorlopig geaccepteerde dataset

dd-mm-jjjj

Goede en geordende staat?

Zeg hier iets over de staat van de dataset. Kunnen we fouten / corrupte bestanden verwachten? Komt alles in aanmerking voor opname in het e-depot of moet er nog worden geselecteerd?

Reden voorlopige acceptatie

In te vullen door RAA

Waarom wordt de dataset geaccepteerd zonder dat deze aan de aansluitvoorwaarden / aanleververeisten voldoet/ nog niet meteen in e-depot kan worden opgenomen?

Bijv: "Het betreft de dataset Videotulen, waarvoor de huidige procedures nog niet voldoen. Het RAA werkt in project E-depot '18 aan oplossingen en zal deze set zelf bewerken." Of: "Het betreft een particulier archief dat goed in onze collectie past. Aan particuliere archieven kunnen geen harde aanleververeisten worden gesteld.

Geschiedenis dataset

Korte omschrijving van relevante gebeurtenissen vóór de verwerving. Hoe (in welk proces, systeem, ter uitvoering van welke taak) is informatie tot stand gekomen? Betreft het gedigitaliseerde informatie? Hebben er migraties of conversies plaatsgevonden? Bestaat de archiefvormende organisatie nog, of is deze opgeheven / opgenomen in andere organisatie?

Bijv: Gemeente X is 2010 gefuseerd met gemeenten Y en Z. De bouwdoossiers van de voormalige gemeente X zijn in 2014 gedigitaliseerd in kader van project Digitaal Werken. In 2015 zijn de scans geleverd door digitaliseringsbedrijf X en opgenomen in het DMS Y. DMS Y wordt vanaf 2018 uitgefaseerd. De papieren dossiers zijn overgebracht in 2016.

Technische kenmerken van de dataset

| | |
|---|---|
| Fysieke drager of elektronische overdracht? | <i>OP welke manier wordt de data overgedragen? Bij fysieke drager(s): een beschrijving van de drager (bijv "Merk X externe harde schijf 1 TB ; 2 afzonderlijke USB sticks). Bij elektronische overdracht: welke wijze van overdracht? (Bijv FTP server, Dropbox, WeTransfer etc.)</i> |
| Omvang /datavolume in MB | <i>Deze kenmerken zijn te achterhalen door de eigenschappen te raadplegen van de hoogstliggende map van de dataset (via rechtermuisknop)</i> |
| Aantal mappen | <i>idem</i> |
| Aantal bestanden | <i>idem</i> |
| Afwijkende bestandsformaten? (Zo ja, welke) | JA / NEE Afwijkende formaten zijn Bestandsformaten die niet worden genoemd in de Handreiking voorkeursformaten Nationaal Archief |
| Databases / gestructureerde gegevens? | JA / NEE Bevat de dataset gestructureerde gegevens, Bijv een datadump uit een systeem in XML. Csv, json of ander formaat voor data uitwisseling? Zo ja, licht dit dan toe bij "Omschrijving van de structuur van het archief" |

Omschrijving van de structuur van het archief

Geef hier de informatie over de structuur van het archief die nodig is om de dataset te kunnen begrijpen / interpreteren. Geef structuren zo veel mogelijk schematisch weer. Er kan ook worden verwezen naar ordeningsplannen / informatiemodellen / xml schema's etc die van toepassing zijn. Voeg deze dan als bijlage toe.

Bijv:

De structuur van de door de leverancier geleverde datadump is beschreven in het document "Ontwerp aan te leveren bestandVideotulen" onder de kop "IST situatie"

Voorwaarden / Risico's voor openbaarheid en toegankelijkheid

Bevat de dataset (nu nog) vertrouwelijke gegevens?

Bevat de dataset persoonsgegevens van nog levende personen? Zo ja, betreft dit ook bijzondere persoonsgegevens?

Bevat de dataset gegevens waar anderen rechten aan kunnen ontlenen? (zoals auteursrecht, portretrecht, databankrecht)

Bevat de dataset encryptie / versleutelde bestanden?

Bijlage 2: Instructie kopiëren via opdrachtprompt

Voor het kopiëren van data naar de bewaar- en werklocaties wordt gebruik gemaakt van de opdrachtprompt, een standaardapplicatie die op elke PC beschikbaar is. Het voordeel van deze methode is, dat precies wordt vastgelegd hoe de kopieeractie verloopt. Mocht de kopieeractie vastlopen, is terug te halen op welk punt de fout is opgetreden.

Kopieeracties via de grafische interface van de verkenners (het “slepen” van bestanden naar mappen) of kopiëren via de sneltoetsen Ctrl+C en Ctrl+V kennen deze rapportagefunctionaliteit niet.

Bij het kopiëren van data naar de bewaar- en werklocaties worden de volgende acties doorlopen:

1. Achterhaal het pad van de map en het station (bijvoorbeeld Q:\) waar de te kopiëren bestanden staan.
 2. Bepaal het pad van de map en het station bijvoorbeeld (O:\) waar de map moet komen te staan. De actie kan niet de map die in stap 1 gemaakt is kopiëren, dus maak zo nodig een doelmap met een identieke naam als de map waar de bestanden vandaan worden gekopieerd.
 3. Open de opdrachtprompt, te vinden op een Windows PC via START>System Tools>Opdrachtprompt en ga naar het station waar vandaan je gaat kopiëren (bijvoorbeeld O:) en druk op Enter.
 4. Voer in de Opdrachtprompt de volgende opdracht in:
robocopy [pad van de te kopiëren map] [pad van de doellocatie] /E
18. Bijvoorbeeld:
robocopy O:\Videotulen Den Helder Q:\ Videotulen Den Helder /E
19. En druk op Enter
5. De zoekactie start nu en laat zien hoe de kopieeractie verloopt.
 6. Na het eindigen van een kopieer actie wordt een verslag weergegeven van de actie.

Het verslag kan worden gekopieerd en worden bewaard in een bestand om de integriteit van de kopieeractie te waarborgen.