

---

*datum* 1 september 2009  
*auteur* Bart Severi ([bart.severi@bz.vlaanderen.be](mailto:bart.severi@bz.vlaanderen.be))  
*onderwerp* Duurzame archiveringsstrategie voor digitale beeldbestanden  
*versienummer* 1

---

## Duurzame archiveringsstrategie voor digitale beeldbestanden

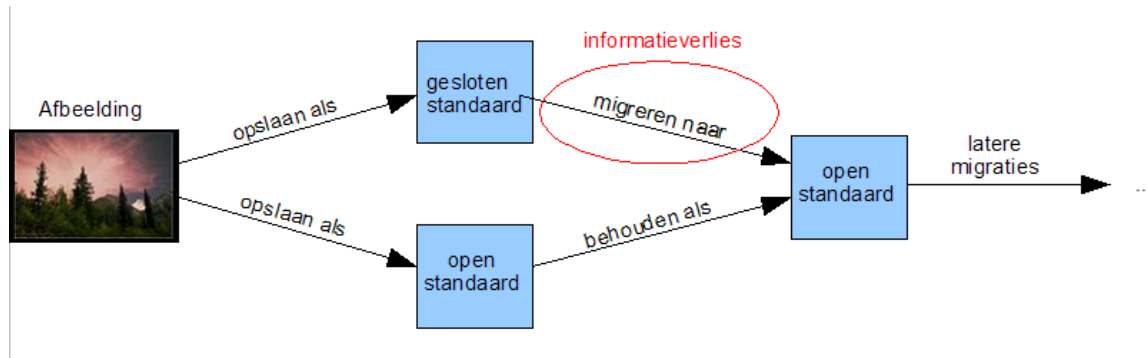
1. Samenvatting.....	1
2. Bereik.....	2
3. Gebruik de juiste bestandsformaten bij de creatie.....	3
3.1 Gebruikscontext van de beeldbestanden.....	3
3.2 Categorieën van beeldbestanden.....	3
3.3 Vastgelegde bewaartermijnen.....	4
3.4 Digitaal duurzaam.....	4
4. Geschikte bestandsformaten bij de creatie.....	5
4.1 Rasterafbeeldingen.....	5
4.1.1 tiff.....	5
4.1.2 png.....	6
4.1.3 jpeg 2000.....	6
4.1.4 Andere formaten: dng en Openraw.....	7
4.2 Vectorafbeeldingen.....	7
4.2.1 svg.....	7
4.3 cad-afbeeldingen.....	7
4.3.1 Goede praktijken voor de creatie van cad-documenten.....	8
4.4 Compositie beeldbestanden.....	8
4.4.1 Webcgm.....	8
4.4.2 svg.....	8
5. Selectieve migratie van bestaande beeldbestanden.....	8
5.1 Beslissingstabel.....	8
5.2 Migratie van vectorafbeeldingen.....	9
5.3 Archivering van cad-afbeeldingen.....	9
6. Gestandaardiseerde metadata.....	11

### 1. Samenvatting

De doelgroep van deze tekst is de informatiebeheerders van de entiteiten en de beleidsdomeinen, samen met informatici die bekommerd zijn om de bewaring van digitale informatie gedurende de tijd dat de organisatie dit verwacht.

Wereldwijd bestaat er consensus over de wijze waarop digitale duurzaamheid nagestreefd moet worden. Door informatie op te slaan in open standaarden is een toekomstvaste strategie mogelijk. Voor beeldbestanden is dit in hoge mate mogelijk, aangezien open standaarden sterk ingeburgerd zijn. Er is één belangrijke uitzondering: CAD-bestanden. Om de langetermijntoegankelijkheid van deze beeldbestanden te bewaken, zal specifieke actie nodig zijn (zie 4.3).

De problematiek van gesloten standaarden wordt hieronder geïllustreerd:



**Figuur 1: open en gesloten standaarden**

De volgende strategie, die start vanaf de creatie, wordt voorgesteld:

1. *beheersregime*: maak een onderscheid tussen te bewaren moederkopieën en afgeleide gebruikerskopieën (zie 3.1)
  - CAD-documenten vereisen bijkomende maatregelen (zie 4.3.1)
2. *creatie*: sla moederkopieën vanaf de creatie op in een geschikt bestandsformaat (zie 4)
  - moederkopieën zijn de originele documenten en zijn onderworpen aan regelgeving
  - een geschikt formaat houdt rekening met:
    - de gebruikscontext van het beeldbestand (zie 3.1)
    - de categorie van het beeldbestand (zie 3.2)
    - de vastgelegde bewaartermijn (zie 3.3)
    - de digitale duurzaamheid (zie 3.4)
3. *metadata*: gebruik een gestandaardiseerde metadataset (zie 6)
4. *archivering*: migreer ongeschikte bestandsformaten naar geschikte formaten (zie 5.1)
  - CAD-documenten vereisen bijkomende maatregelen (zie 5.3)

## 2. Bereik

In de administraties van de Vlaamse overheid worden afbeeldingen gebruikt voor uiteenlopende doeleinden: de foto's die als bijlage bij een proces-verbaal van een inspecteur gevoegd zijn, een beeldbank van luchtfoto's of schilderijen, het logo van een entiteit, CAD-tekeningen voor topografie of wegenstudie, enzovoort. Net zoals alle andere procesgebonden informatie (zoals e-mails, databanken, presentaties of rekenbladen), zijn deze afbeeldingen onderworpen aan een bepaalde bewaartermijn.

Deze strategie beoogt de archivering van digitale beeldbestanden op een duurzame wijze, met maximale kansen op raadpleegbaarheid en toegankelijkheid op korte, middellange en lange termijn.

Onder de noemer 'beeldbestanden' wordt digitale afbeeldingen verstaan. Hieronder vallen zowel rasterafbeeldingen als vectorafbeeldingen. Ook CAD-afbeeldingen worden inbegrepen. Bewegende beelden (video) en geografische informatiesystemen (GIS) vereisen om technische redenen een eigen aanpak en worden in deze strategie niet behandeld.<sup>1</sup>

Tijdens de opmaak van deze strategie is informeel overleg gepleegd met (sub)entiteiten die intensief met beeldbestanden werken, zoals Team Vlaams Bouwmeester (BZ), Fotoarchief Vlaanderen (LNE), Afdeling Planning en Coördinatie en Afdeling West-Vlaanderen (Agentschap Wegen en Verkeer), Algemene Technische Ondersteuning (Departement MOW), het Agentschap voor Facilitair Management (BZ), de Dienst Natuurlijke Rijkdommen (Departement LNE), Luchthaven Antwerpen en Luchthaven Oostende (allen MOW) en het Vlaamse Instituut voor Onroerend Erfgoed (RWO).

<sup>1</sup> Wat in andere bestandsformaten (zoals PDF, MSG of DOC) ingebedde afbeeldingen betreft, wordt doorverwezen naar de respectieve bewaarstrategieën voor deze formaten.

### 3. Gebruik de juiste bestandsformaten bij de creatie

Indien de archiefvormende organisatie ervoor kiest om zijn informatie vanaf de creatie in geschikte bestandsformaten op te slaan, minimaliseert men het risico op een gefnuikte raadpleegbaarheid. In dat geval is immers geen onmiddellijke migratie naar een andere, open standaard nodig, waardoor ook de kans op informatieverlies verkleint.<sup>2</sup>

Een geschikt bestandsformaat voor beeldbestanden is:

- afgestemd op de gebruikscontext
- afgestemd op de beeldcategorie
- afgestemd op de bewaartermijn

Een geschikt bestandsformaat voor moederkopieën van beeldbestanden is bovendien digitaal duurzaam.

De tabel achteraan dit document toont in detail welke bestandsformaten en coderingen in aanmerking komen voor de verschillende categorieën van beeldbestanden (zie blz. 14).

#### 3.1 Gebruikscontext van de beeldbestanden

Het is zinvol om een onderscheid te maken tussen **de moederkopieën en de gebruikskopieën** van beeldbestanden. De moederkopie is de *master* die gedurende een in het archiefbeheersplan vastgelegde bewaartermijn bewaard moet blijven en ze wordt opgeslagen in een archiveringsformaat. De gebruikskopie wordt afgeleid van de moederkopie voor bepaalde doeleinden, zoals *thumbnails* in een beeldbank of lageresolutieversies van foto's voor gebruik op een website. Een ongecomprimeerd TIFF-bestand als moederkopie is immers doorgaans vele malen groter dan een gecomprimeerd JPEG-bestand.

Het spreekt voor zich dat de eisen die gesteld worden aan de moederkopieën strenger zijn dan voor gebruikskopieën. Eerstgenoemde klasse van beeldbestanden vormt dan ook het onderwerp van deze strategie.

Het werken met **verschillende beheersregimes** voor moederkopieën en gebruikskopieën heeft als bijkomend voordeel dat de moederkopieën op een gecontroleerde manier beheerd kunnen worden en dat visuele artefacten (informatieverlies) bij gebruikskopieën door opeenvolgende bewerkingen hersteld kunnen worden door terug te grijpen naar de moederkopieën.

#### 3.2 Categorieën van beeldbestanden

Idealiter kiest men het juiste bestandsformaat voor de juiste afbeelding. De keuze van het bestandsformaat is sterk afhankelijk van de soort grafische informatie. Voor rasterafbeeldingen zijn 4 categorieën te onderscheiden: creatieve, fotografische, typografische en geografische afbeeldingen. Voor vectorafbeeldingen maakt men een onderscheid tussen creatieve afbeeldingen en CAD-afbeeldingen. Tenslotte vormen de compositiete beeldbestanden (een combinatie van raster en vector) een aparte categorie.

#### 3.3 Vastgelegde bewaartermijnen

Vanuit het oogpunt van digitale duurzaamheid en leveranciersafhankelijkheid moet men ernaar streven om al zijn informatie vast te leggen in open standaarden. Vandaar dat ook deze strategie ervoor pleit om beeldbestanden vanaf de creatie in een geschikt archiveringsformaat op te slaan, dat meteen als moederkopie fungeert.

Beeldbestanden zoals de foto's in een beeldbank dragen hun eigen bewaartermijn. De meeste beeldbestanden vormen echter meestal samen met andere documenten een dossier, zoals de foto's die bij het bij een proces-verbaal van een inspecteur horen. Uiteraard nemen de beeldbestanden de bewaartermijn over van het dossier.

<sup>2</sup> De gekozen voorkeursformaten zijn gebaseerd op de bewaarstrategieën van eDAVID, de National Archives of Australia, de Koninklijke Bibliotheek (Den Haag) en het Florida Digital Archive. F. Boudrez en H. Dekeyser, *DAVID. Digitaal archiveren in de praktijk. Handboek* (Antwerpen-Leuven, 2003) 3.3. Zie <http://www.edavid.be/davidhandboek/index.html>. Zie ook R. Gillesse, J. Rog en A. Verheusen, *Alternative File Formats for Storing Master Images of Digitisation Projects* (2008) 41. Zie [http://www.kb.nl/hrd/dd/dd\\_links\\_en\\_publicaties/publicaties/Alternative%20File%20Formats%20for%20Storing%20Masters%20201.pdf](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/Alternative%20File%20Formats%20for%20Storing%20Masters%20201.pdf). Voorts <http://www.naa.gov.au/records-management/secure-and-store/e-preservation/at-NAA/software.aspx> en tenslotte <http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf>.

Omwille van de technologische vernieuwingscyclus van software en hardware is de bewaring van digitale informatie die een **bewaartermijn van meer dan vijf jaar** heeft, onderworpen aan een bijzonder regime. Beeldbestanden die minder dan een half decennium bewaard moeten worden, hoeven dan ook niet ingepast te worden in deze strategie.

### 3.4 Digitaal duurzaam

Een archiveringsformaat is **digitaal duurzaam**; dat wil zeggen dat het toekomstvast is. Het is een open standaard en dus vrij beschikbaar. Het sluit eveneens externe afhankelijkheden zoveel mogelijk uit, zowel van technische als van juridische aard. Dat betekent dat Windows Bitmap (BMP)<sup>3</sup> wegens platformafhankelijkheid, het Graphics Interchange Format (GIF) wegens licentiebeperkingen<sup>4</sup> en RAW (een verzamelnaam van verschillende bedrijfseigen, gesloten formaten)<sup>5</sup> niet in aanmerking komen als geschikt formaat voor moederkopieën.

Afbeeldingsformaten die niet breed ondersteund zijn, zijn eveneens ongeschikt als archiveringsformaat, zelfs indien zij open standaarden zijn. Een voorbeeld is PC Paintbrush Bitmap (PCX), waarvan de specificaties beschikbaar zijn en dat platformonafhankelijk is, maar waarvan de **marktondersteuning** sinds midden jaren 1990 taant.<sup>6</sup>

Het gebruik van **comprimering moet zoveel mogelijk vermeden worden**, niet alleen omdat bepaalde compressiealgoritmes onherstelbaar gegevensverlies met zich meebrengen, maar ook omdat dit een extra bewerkingsslag inhoudt, wat indruist tegen het principe van digitale duurzaamheid dat dergelijke externe afhankelijkheden zoveel mogelijk wil uitsluiten en idealiter kiest voor eenvoud.

Dat wil echter niet zeggen dat bestandsformaten die comprimering ondersteunen, niet toegepast kunnen worden, zolang het compressiealgoritme aan dezelfde toekomstvaste voorwaarden voldoet als het bestandsformaat zelf en zolang de comprimering geen onherstelbaar gegevensverlies met zich meebrengt. Vandaar dat enkel gekozen wordt voor comprimering die open en breed ondersteund is, en tegelijk verliesloos (*lossless*) is. Een voorbeeld: TIFF\_ZIP is een acceptabele versie met comprimeringsmogelijkheden. Daarentegen is TIFF\_LZW, waarbij de afbeelding gecomprimeerd wordt via een algoritme dat aan sterke licentiebeperkingen onderhevig is, niet aan te bevelen.

Indien toch gekozen wordt voor verlieslatende comprimering (in bijvoorbeeld JPEG 2000), kan men via PSNR (Peak Signal to Noise Ratio) kwantificeren welk verschil er mag optreden tussen het origineel en de gecomprimeerde versie. Het voordeel is dat de comprimering wordt aangepast aan de visuele inhoud van het beeldbestand. Het werkt echter niet goed bij beeldbestanden met relatief weinig onderling kleurverschil tussen de pixels (zoals zwart-witfoto's).<sup>7</sup>

Soms kunnen archiefvormers verlieslatende comprimering die *visually lossless* maar in de praktijk *lossy* is, geen probleem vinden. Men ziet immers weinig verschil met het blote oog. Maar voor de bewaring op middellange en lange termijn is zowel verlieslatende als verliesloze comprimering sterk af te raden wegens latere, onvermijdelijke migraties naar andere formaten, waarbij gegevensverlies potentieel uitvergroet kan worden.

## 4. Geschikte bestandsformaten bij de creatie

### 4.1 Rasterafbeeldingen

Rasterafbeeldingen stellen een afbeelding voor door middel van beeldpunten (pixels). In haar meest eenvoudige vorm is een rasterafbeelding een verzameling pixels die horizontaal en verticaal geordend zijn. De wijze waarop deze beeldpunten opgeslagen worden, verschilt van formaat tot formaat. Een rasterafbeelding bestaat dus uit een vaste hoeveelheid pixels, uitgedrukt in breedte en hoogte (resolutie), die een bepaalde kleurdiepte hebben. De kleurdiepte is het aantal kleuren

<sup>3</sup> F. Boudrez, *Standaarden voor bestandsformaten* (Digitaal archiveren: Richtlijn en advies 4) (2006) 31. Zie <http://www.edavid.be/davidproject/teksten/Richtlijn4.pdf>.

<sup>4</sup> Voor de creatie van GIF-bestanden dient men in principe te beschikken over een Unisyslicentie. Zie Boudrez, *Standaarden*, 33 (zie noot 3).

<sup>5</sup> <http://www.openraw.org/info/>

<sup>6</sup> <http://www.qzx.com/pc-gpe/pcx.txt>

<sup>7</sup>B. Comstock, 'Using PSNR thresholds to modulate the degree of lossy compression in JPEG2000 files', DPC/BL Joint JPEG 2000 Workshop, 25 juni 2007. Zie <http://www.dpconline.org/docs/events/0706jpeg2000comstock.pdf>.

dat één pixel kan bevatten. Hoe groter de resolutie of de kleurdiepte, hoe groter de omvang van het (ongecomprimeerde) beeldbestand.

#### 4.1.1 TIFF

*Status: voorkeursformaat*

Het Tagged Image File Format is platformonafhankelijk en wordt breed ondersteund. Het kent een wildgroei aan varianten omdat het relatief eenvoudig uitbreidbaar is.<sup>8</sup> Deze genieten echter niet allemaal even brede ondersteuning. Gebruik daarom de gestandaardiseerde baseline TIFF 6.0, die ongecomprimeerd is (TIFF\_UNC). Indien compressie noodzakelijk is wegens plaatsgebrek, verdient TIFF\_ZIP, dat een vrij beschikbaar compressiealgoritme is, de voorkeur op TIFF\_LZW, waarvan de rechten nog steeds opgeëist worden.<sup>9</sup> Het Group 4-comprimeringsalgoritme (TIFF\_G4) is geschikt voor beeldbestanden met voornamelijk tekstuele informatie, zoals ingescande documenten.

Fotografisch-technische gegevens, zoals het cameramodel, de sluitertijd of de diafragmaopening, kunnen opgeslagen worden als EXIF-metadata, waarvoor TIFF\_UNC\_EXIF een geschikt formaat is.

Als gebruikskopie is TIFF minder geschikt omwille van de grootte. Gebruik daarvoor een formaat zoals GIF of JPEG, dat omwille van de tijdelijke waarde niet moet voldoen aan de criteria voor geschikte archiveringsstandaarden.

Eenvoudige cartografische gegevens kunnen als metadata ingebed worden in een beeldbestand. Baseline GeoTIFF is hierbij een geschikt formaat, in tegenstelling tot het bedrijfseigen, gesloten MrSid. GeoTIFF ondersteunt echter niet de interactieve functionaliteiten (zoom en multiresolutie) van MrSid, terwijl JPEG 2000 wel de interactiviteit biedt maar niet de geografische metadata.

#### 4.1.2 PNG

*Status: alternatief formaat*

Het formaat Portable Network Graphics (ISO-15948:2004) werd ontwikkeld als alternatief voor GIF (Graphics Interchange Format), dat onderhevig was (en is) aan licentieproblemen. Het unieke voordeel van PNG is dat het volledig vrij is van licentiebeperkingen.

Het geniet een brede ondersteuning en biedt enkele verbeteringen ten opzichte van GIF. In tegenstelling tot GIF kan het echter geen animatie bevatten.<sup>10</sup> Het is niettemin een zeer geschikt archiveringsformaat en het wordt momenteel ook gebruikt voor de duurzame archivering van meerdere categorieën van beeldbestanden.<sup>11</sup>

#### 4.1.3 JPEG 2000

*Status: alternatief formaat*

JPEG2000 (ISO-15444-1:2000) is een open standaard die de eigenschappen van JPEG en TIFF wil combineren. Ze ondersteunt een comprimeringsalgoritme dat *lossless* is, waarbij de omvang van het bestand verkleint maar zonder kwaliteitsverlies (J2K\_C\_LL). De bekende voorganger JPEG ondersteunt eveneens comprimering maar met onherstelbaar kwaliteitsverlies van de afbeelding. Voor moederkopieën moet dit absoluut vermeden worden, aangezien dit verlies meegenomen en potentieel vergroot wordt tijdens de bewaartermijn. Het voordeel van JPEG2000 is dat bij *lossless* comprimering de bestandsomvang tot de helft kleiner kan zijn dan een ongecomprimeerde TIFF.<sup>12</sup>

Het bestandsformaat JPEG2000 leent zich tot flexibel gebruik, waardoor het tegelijkertijd als moederkopie en gebruikerskopie gebruikt kan worden. Het kan multiresolutieafbeeldingen creëren uit één JPEG2000-bitstream van dezelfde afbeelding, waardoor voorbeeldafbeeldingen (*thumbnails*) automatisch gecreëerd kunnen worden. Het biedt ook meer functionaliteiten dan andere bestandsformaten zoals interactieve zoom en panoramische beweging (*panning*).<sup>13</sup>

<sup>8</sup> Boudrez, *Standaarden*, 25-26 (zie noot 3).

<sup>9</sup> Hoewel het oorspronkelijke LZW-copyright in 2004 verstreken is, claimt Unisys nog de rechten te bezitten op uitbreidingen van het Lempel-Ziv-Welch-algoritme. R. Gillese, J. Rog en A. Verheusen, *Alternative File Formats for Storing Master Images of Digitisation Projects* (2008) 35-37. Zie [http://www.kb.nl/hrd/dd/dd\\_links\\_en\\_publicaties/publicaties/Alternative%20File%20Formats%20for%20Storing%20Masters%20%201.pdf](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/Alternative%20File%20Formats%20for%20Storing%20Masters%20%201.pdf)

<sup>10</sup> Boudrez, *Standaarden*, 31 (zie noot 3).

<sup>11</sup> Bijvoorbeeld door de National Archives of Australia. Zie <http://www.naa.gov.au/recordkeeping/preservation/digital/summary.html>

<sup>12</sup> R. Gillese, J. Rog en A. Verheusen, *Alternative File Formats for Storing Master Images of Digitisation Projects* (2008) 21. Zie [http://www.kb.nl/hrd/dd/dd\\_links\\_en\\_publicaties/publicaties/Alternative%20File%20Formats%20for%20Storing%20Masters%20%201.pdf](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/Alternative%20File%20Formats%20for%20Storing%20Masters%20%201.pdf)

Het JPEG 2000-bestandsformaat is meer foutbestendig dan TIFF, wat een sterk voordeel is met het oog op langetermijnbewaring.<sup>14</sup> Dat wil zeggen dat een onleesbare bit een veel kleinere impact heeft bij JPEG 2000 dan bij TIFF, hoewel het gebruik van comprimering het risico enorm verhoogt. Bewerkingen met JPEG 2000 vergen echter meer processorcracht dan TIFF, wat hen iets minder handelbaar maakt op dit moment.

Twee factoren spelen echter in het nadeel. Het wordt voorlopig slechts in beperkte mate ondersteund<sup>15</sup> en het formaat verkeert in een onzekere juridische situatie wat de geldende patenten betreft.<sup>16</sup> Het wordt momenteel met mondjesmaat in gebruik genomen door archiefinstellingen.<sup>17</sup> Bijgevolg wordt het in deze strategie slechts als alternatief gezien voor TIFF en PNG, hoewel dit in de toekomst uiteraard kan veranderen.

#### 4.1.4 Andere formaten: DNG en OpenRAW

*Status: niet in aanmerking*

Het formaat Digital Negative (DNG) is door Adobe ontwikkeld als open alternatief voor de vele bedrijfseigen varianten van het RAW-formaat, dat gebruikt wordt voor fotografische afbeeldingen. Op dit moment voldoet het niet aan de eisen van een open standaard.<sup>18</sup> In afwachting van een alternatief ijvert OpenRAW voor de publicatie van de verschillende RAW-specificaties.<sup>19</sup>

Wat foto's in RAW-formaat betreft, is het aan te bevelen om ze zo snel mogelijk naar een geschikt formaat te converteren.

## 4.2 Vectorafbeeldingen

Vectorafbeeldingen stellen een afbeelding voor als een geheel van punten waarvan de relatieve coördinaten vastgelegd worden en die verbonden worden door middel van geometrische vormen. Visuele objecten worden dus niet gecreëerd door pixels maar door meetkundige figuren. Dat maakt de afbeelding schaalbaar en kleiner in omvang dan een rasterafbeelding.

### 4.2.1 SVG

*Status: voorkeursformaat*

Het bestandsformaat SVG (Scalable Vector Graphics) is een geschikt formaat voor eenvoudige vectorafbeeldingen, omdat het platformonafhankelijk is en breed ondersteund wordt. Er moet wel naar gestreefd worden dat externe afhankelijkheden vermeden worden, zoals *Java binding*.<sup>20</sup>

## 4.3 CAD-afbeeldingen

CAD-afbeeldingen zijn vectoriële afbeeldingen in twee- of driedimensionele vorm, die gebruikt worden voor technische tekeningen, zoals toegepast in de architectuur, de archeologie of de bouw.

Wat de garantie op digitale duurzaamheid van CAD-bestanden betreft, is de situatie **zonder meer onrustwekkend** te noemen om twee redenen:

- *vendor lock-in*: er is geen volwaardig, open alternatief voor de bedrijfseigen bestandsformaten

<sup>13</sup> R. Buckley, *JPEG 2000 - a Practical Digital Preservation Standard?*, Digital Preservation Coalition (DPC Technology Watch Series Report 08-01) (2008), 2.

<sup>14</sup> V. Heydegger, 'JP2's Preservation Capabilities within PLANETS, DPC/BL Joint JPEG2000 Workshop, Londen, 25 juni 2007. Zie <http://www.dpconline.org/graphics/events/0706jpeg2000wkshop.html> Zie ook P. Buonora en F. Liberati, 'A Format for Digital Preservation of Images: A Study on JPEG 2000 File Robustness', *D-Lib Magazine*, 14 (2008/7-8). Zie <http://www.dlib.org/dlib/july08/buonora/07buonora.html>.

<sup>15</sup> R. Gillesse, J. Rog en A. Verheusen, *Alternative File Formats for Storing Master Images of Digitisation Projects* (2008) 22. Zie [http://www.kb.nl/hrd/dd/dd\\_links\\_en\\_publicaties/publicaties/Alternative%20File%20Formats%20for%20Storing%20Masters%202%201.pdf](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/Alternative%20File%20Formats%20for%20Storing%20Masters%202%201.pdf)

<sup>16</sup> R. Clark, 'JPEG and JPEG 2000: Past, Present, and Future', DPC/BL Joint JPEG 2000 Workshop, Londen, 25 juni 2007. Zie <http://www.dpconline.org/graphics/events/0706jpeg2000wkshop.html>. Zie ook <http://www.jpeg.org/jpeg2000/CDs15444.html>.

<sup>17</sup> Slechts 16,7% van de wetenschappelijke bibliotheken in de VS gebruikt het als archiveringsformaat. D. Lowe en M.J. Bennett, 'Digital Project Staff Survey of JPEG 2000 Implementation in Libraries', University of Connecticut. Zie [http://digitalcommons.uconn.edu/libr\\_pubs/16/](http://digitalcommons.uconn.edu/libr_pubs/16/).

<sup>18</sup> <http://www.adobe.com/products/dng/>

<sup>19</sup> <http://www.openraw.org>.

<sup>20</sup> <http://www.w3.org/TR/SVG/java.html>



- *lange bewaring*: de afbeeldingen zijn meestal gekoppeld aan dossiers met zeer lange bewaartermijnen

Door de combinatie van deze twee factoren ontstaat een situatie die zeer problematisch is. Bedrijfskritische informatie is opgeslagen in een gesloten bestandsformaat, dat gekoppeld is aan een bepaalde applicatie. De wijze waarop geometrische en parametrische informatie opgeslagen wordt in het formaat, verschilt sterk van andere formaten en is niet eenvoudig uitwisselbaar. Wegens de lange ontwerptijd en de langlopende aansprakelijkheid moeten deze bestanden bovendien vaak meerdere decennia bewaard blijven.

Door deze beperkingen is de enige zekere archiveringsstrategie conversie: bestanden worden steeds opgeslagen in een nieuwere versie van het bestandsformaat via een nieuwe, hogere versie van de applicatie. Daarom is men verplicht om geregeld updates of nieuwe versies aan te kopen.

**Deze *vendor lock-in* is in geen enkel opzicht een toekomstvaste strategie.**

De dominante CAD-applicatie binnen de Vlaamse overheid is AutoCAD van Autodesk, dat in de productcatalogus van EDS-Telindus opgenomen is. Daarnaast wordt ook MicroStation van Bentley Systems gebruikt, zij het in mindere mate. DWG is het gesloten bestandsformaat van Autodesk-applicaties; DGN is Bentleys formaat. Beiden zijn geen geschikte archiveringsformaten. Bovendien is DWG niet volledig achterwaarts compatibel: oude versies kunnen niet altijd perfect ingelezen worden in jongere versies van AutoCAD.<sup>21</sup>

Net zoals bij de RAW-formaten het geval is, wordt ook voor CAD-bestanden gewerkt aan een open standaard. Het initiatief van de Open Design Alliance heeft echter nog geen concrete standaard opgeleverd.<sup>22</sup>

Een **krachtige oproep vanuit de Vlaamse overheid** naar de ontwikkelaars van CAD-software om een open standaard te ontwikkelen en te ondersteunen zal leiden tot een oplossing van dit probleem. De recente geschiedenis van de standaarden voor kantoordocumenten toont aan dat dit verre van onrealistisch is.

#### 4.3.1 Goede praktijken voor de creatie van CAD-documenten

Door het volgen van een aantal goede praktijken wordt het duurzaam archiveren van CAD-bestanden gedeeltelijk vereenvoudigd.<sup>23</sup> Daarom is het nodig om voor elk project duidelijke richtlijnen te hanteren:

- *koppelingen*: leg relatieve koppelingen met externe bronnen zoals afbeeldingen of databanken i.p.v. absolute
- *versiecontrole*: leg vast welke bestandsformaten en welke versies gebruikt moeten worden
- *documentatie*: documenteer de richtlijnen die zijn gehanteerd

#### 4.4 Composiete beeldbestanden

Composietbeeldbestanden bevatten vectoriële afbeeldingen met ingebedde rasterafbeeldingen. Ze worden vooral gebruikt als formaat voor tweedimensionele, technische handleidingen en ontwerpen.

##### 4.4.1 WebCGM

*Status: voorkeursformaat (bij object linking); alternatief formaat (zonder object linking)*

Het WebCGM-formaat, gebaseerd op het formaat Computer Graphics Metafile (ISO-8632-1/4:1992), is interoperabel en wordt vooral gebruikt indien *object linking* vereist is. *Object linking* is het leggen van verbanden tussen interne en externe grafische of tekstuele componenten, zoals bijvoorbeeld een URL die gekoppeld wordt aan een onderdeel van de afbeelding.<sup>24</sup>

Indien *object linking* geen voorwaarde is, kan men opteren voor het eerder vermelde SVG (zie 2.3.1). Hoewel het aan te bevelen is om comprimering zoveel mogelijk te vermijden, kunnen de in-

<sup>21</sup> K. Westcott, *Preservation Handbook – Computer Aided Design (CAD)* (Arts and Humanities Data Service, 2005) 6. Zie <http://ahds.ac.uk/preservation/cad-preservation-handbook.pdf>.

<sup>22</sup> M. Coyne e.a., *The Significant Properties of Vector Images* (2007), ii. Zie [http://www.jisc.ac.uk/media/documents/programmes/preservation/vector\\_images.pdf](http://www.jisc.ac.uk/media/documents/programmes/preservation/vector_images.pdf). Zie ook <http://www.opendwg.org/>. Er is echter vertraging in de ondersteuning t.o.v. de ontwikkeling van AutoDesk; momenteel wordt tot en met AutoCAD 2002 ondersteund door DWGdirect.

<sup>23</sup> <http://ads.ahds.ac.uk/project/goodguides/cad/>

<sup>24</sup> <http://www.cgmopen.org/technical/cgm-svg.htm#objLink>

gebedde rasterafbeeldingen indien noodzakelijk gecomprimeerd worden volgens de PNG-methode *deflate* (WebCGM\_PNG) of via het Group 4-algoritme (WebCGM\_G4).<sup>25</sup>

#### 4.4.2 SVG

*Status: voorkeursformaat (zonder object linking); alternatief formaat (bij object linking)*

In tegenstelling tot WebCGM is de ondersteuning voor *object linking* bij SVG minder gesofisticeerd. Rasterafbeeldingen kunnen echter wel ingebed worden, zij het dat dit de omvang meer vergroot dan bij WebCGM het geval is. Dit laatste is immers een binair formaat; bij SVG wordt de rasterinhoud via base64 gecodeerd, wat het relatief groter maakt.

## 5. Selectieve migratie van bestaande beeldbestanden

### 5.1 Beslissingstabel

Indien beeldbestanden met een bewaartermijn van meer dan 5 jaar **niet zijn opgeslagen in een geschikt bestandsformaat** (zie 4. *Geschikte bestandsformaten*) en dus een geschikt archiveringsformaat, moeten zij gemigreerd worden bij archivering, waarbij men rekening moet houden met **potentieel informatieverlies**. Migratie verloopt volgens deze beslissingstabel:

soort beeldbestand	korte bewaartermijn (1-5 jaar)	lange bewaartermijn (6-... jaar)	Permanente bewaring
<i>Raster, vector &amp; composiet</i>	origineel formaat	voorkeursformaat (zie tabel 2)	voorkeursformaat + origineel
CAD	origineel formaat	<i>(zie 5.3)</i>	

Tabel 1: Beslissingstabel voor selectieve migratie

Enkele voorbeelden:

- JPEG: een gecomprimeerde foto wordt gemigreerd naar TIFF\_UNC\_EXIF (om de fototechnische metadata te bewaren), naar een gecomprimeerde PNG (uit plaatsgebrek) of naar baseline TIFF\_UNC.
- MrSid: een cartografisch beeldbestand in een gesloten formaat wordt gemigreerd naar GeoTIFF om de geografische metadata te bewaren of naar JPEG 2000 om de zoomfunctie te behouden
- VSD: een Visio Drawing-bestand wordt gemigreerd naar SVG

### 5.2 Migratie van vectorafbeeldingen

Enkele nuttige aandachtspunten:

- *van vector naar vector*: alle gelinkte bestanden (lettertypes, rasterafbeeldingen) moeten aanwezig zijn
- *van vector naar raster*: het gedrag (schaalbaarheid) gaat verloren

### 5.3 Archivering van CAD-afbeeldingen

Zoals al besproken bestaat hiervoor momenteel geen eenvoudige oplossing (zie 4.3 *CAD-afbeeldingen*). Hier wordt een bewaarstrategie voorgesteld die gebaseerd is op 3 bestaande, waaronder die van het Massachusetts Institute of Technology (MIT).<sup>26</sup> Uitgangspunt van de strategie is dat de risico's op verlies van informatie en raadpleegbaarheid zo klein mogelijk gemaakt worden.

1. *Schoning*: verwijder ongebruikte *layers*, *blocks* en andere objecten om de bestandsstructuur te vereenvoudigen en de bestandsomvang te verkleinen. Afhankelijk van hun belang kunnen onzichtbare objecten ook verwijderd worden, net als belichtingsgegevens.
2. *Koppelingen*: externe bestanden (andere beeldbestanden, lettertypes, databanken...) moeten aanwezig zijn en via relatieve links gekoppeld zijn aan het CAD-bestand. Bijkomende, ei-

<sup>25</sup> <http://www.w3.org/TR/REC-WebCGM/REC-02-CGM-Concepts.html>

<sup>26</sup> W. Reilly e.a. 'Preserving Brand-New Buildings. FACADE. Future-proofing Architectural Computer-Aided Design', Digital Library Federation Fall Forum 2008, 14 november 2008. Zie [http://facade.mit.edu/presentations/REILLY\\_DLFFall2008\\_MIT\\_FACADE.ppt.htm](http://facade.mit.edu/presentations/REILLY_DLFFall2008_MIT_FACADE.ppt.htm). Zie voorts <http://csanet.org/newsletter/winter02/nlw0205.html> en <http://ahds.ac.uk/preservation/cad-preservation-handbook.pdf>



gen documentatie over de koppelingen en de locatie van de gearchiveerde externe bestanden is sterk aangewezen wegens de wijzigende technieken waarmee CAD-software koppelingen beschrijft.

3. *Externe bestanden*: deze moeten evenzeer duurzaam bewaard worden<sup>27</sup>
4. *Creëer origineel bestand*: sla de geschoonde CAD-afbeelding op in het bestandsformaat en de versie zoals aanbevolen in de goede praktijken (zie 4.3.1)
5. *Creëer plots*: indien *plots* gedefinieerd zijn, creëer rasterafbeeldingen hiervan (zie tabel 2 voor geschikte formaten)
6. *Creëer DXF*: exporteer de afbeelding naar DXF. Pas de precisiegraad aan de eisen van de afbeelding aan, gaande van hoge precisie (met vaak te hoge *bloat* van de DXF) tot lage precisie (met vaak informatieverlies). Dit is afhankelijk van de afmeting van het ontwerp: hoe groter het ontwerp fysiek is, hoe lager de precisie mag zijn. Het Drawing eXchange Format is een door Autodesk ontwikkeld bestandsformaat voor de uitwisseling van CAD-tekeningen tussen verschillende programma's. De specificaties zijn auteursrechtelijk beschermd maar vrijgegeven en het geniet een brede ondersteuning.<sup>28</sup> Indien gekozen wordt voor de ASCII-opslagmethode, is het een geschikte archiveringsstandaard.<sup>29</sup> Problematisch is echter dat de synchronisatie met DWG steeds met vertraging gebeurt, zodat een aantal recent geïntroduceerde functionaliteiten in AutoCAD niet wordt ondersteund.<sup>30</sup> Daarom is het aanbevolen om steeds in de hoogst mogelijk versie van DXF te bewaren, aangezien de versies wegens de vele wijzigingen niet altijd retrocompatibel zijn met oude versies van DXF.<sup>31</sup> Het is dus voornamelijk een geschikt archiveringsformaat voor oudere AutoCAD-bestanden.
7. *Creëer STEP*: het uitwisselingsformaat STEP (ISO-10303) wordt breed ondersteund door CAD-software en is een valabel archiveringsformaat. Het behoudt echter niet alle functionaliteiten van het oorspronkelijke CAD-bestand.<sup>32</sup>
8. *Creëer PDF/E*: dit maakt een eenvoudige, snelle visualisering mogelijk. De functionaliteiten van de CAD-afbeelding en ingebedde informatie over de analyse, de inspectie en het ontwerp gaan hier wel verloren.<sup>33</sup>
9. *Documenteer*: bewaar de informatie over de gebruikte CAD-software, bestandsstructuur, de locatie van extern gekoppelde bestanden, de gekozen bestandsformaten en –versies.
10. *Archiveer*: bewaar het originele CAD-bestand, de gekoppelde externe bestanden, de *plots*, het DXF-bestand, het STEP-bestand, het PDF/E-bestand en de documentatie. Bewaar ook de software zelf, indien de licentie dit toestaat.

Op basis van deze stappen kan in de nabije toekomst een grondige bewaarstrategie uitgebouwd worden als complexere oplossingen zoals LIMMA of RRORfE aan maturiteit winnen, zodat alle relevante geometrische informatie in een open standaard kan worden opgeslagen.<sup>34</sup>

## 6. Gestandaardiseerde metadata

De volgende richtlijnen zijn van belang bij het gebruik van metadata:

- *standaardisering*: gebruik een geformaliseerde metadataset, zoals TIFF, EXIF, XMP of IPTC. Dat vereenvoudigt de overdracht van metadatasets bij migratie van het ene bestandsformaat

<sup>27</sup> Zie de relevante archiveringsstrategieën: <http://www2.vlaanderen.be/proces-en-informatiebeleid/ca/adviezen.html>

<sup>28</sup> IntellicAD, Adobe Illustrator, Corel Draw, Inkscape e.a. ondersteunen het formaat.

<sup>29</sup> De binaire DXF is weliswaar kleiner in omvang maar is minder breed ondersteund. Boudrez, *Standaarden*, 33 (zie noot 3).

<sup>30</sup> [http://en.wikipedia.org/wiki/AutoCAD\\_DXF](http://en.wikipedia.org/wiki/AutoCAD_DXF)

<sup>31</sup> Coyne, *Vector Images*, ii (zie noot 22).

<sup>32</sup> J.D. Mooney, 'STEP Issues', WVU/NETL/ERA Workshop on Digital Preservation of Complex Engineering Data WVU NRCCE (Morgantown, West Virginia, 20-21 april 2009). Zie <http://www.preservationworkshop.wvu.edu/r/download/30112>. STEP is een door de Nederlandse Archiefwet verplichte archiveringsstandaard: Regeling geordende en toegankelijke staat archiefbescheiden (23 februari 2002) art. 6.

<sup>33</sup> W.C. Regli, 'Preservation of Engineering Artifacts'. Atlantic Workshop on Long Term Knowledge Retention, University of Bath, 12-13 februari 2007. <http://www.ukoln.ac.uk/events/ltkr-2007/programme.html>

<sup>34</sup> <http://www.ukoln.ac.uk/projects/grand-challenge/#ball.etal2008tcp>. Zie ook de ontwikkeling van de Britse standaard <http://www.asd-stan.org/Lotar.html>

naar het andere en garandeert – niet onbelangrijk bij langetermijnbewaring – de begrijpelijkheid van de metadatavelden voor andere gebruikers.

- *eenvoud*: beperk het aantal metadatastandaarden tot 1 idealiter
- *opslagwijze*: kies een geschikte opslagwijze voor metadata; ofwel ingekapseld in het beeldbestand (EXIF en IPTC in JPEG 2000 of de standaardmetadataset van baseline TIFF\_UNC<sup>35</sup> en PNG<sup>36</sup>), ofwel extern bijgehouden (in bijvoorbeeld een beeldbank of een XML-schema als MIX<sup>37</sup>), ofwel beide.

---

<sup>35</sup> Federal Agencies Digitization Guidelines Initiative, *Guidelines for TIFF Metadata: Recommended Elements and Format: Version 1.0* (2009). Zie [http://www.digitizationguidelines.gov/stillimages/documents/TIFF\\_Metadata\\_Final.pdf](http://www.digitizationguidelines.gov/stillimages/documents/TIFF_Metadata_Final.pdf)

<sup>36</sup> PNG heeft een beperkt aantal voorgedefinieerde metadatavelden. Zie <http://www.w3.org/TR/PNG/#11textinfo>.

<sup>37</sup> <http://www.loc.gov/standards/mix/>.

Rasterbeeldbestanden	Karakteristieken	Voorbeelden	Voorkeursformaat	Alternatief formaat
Creatieve afbeeldingen	polychromatisch; scherpe contrasten tussen aangelande pixels	diagrammen; grafieken; screenshots van GUI's; gefixeerde vectorafbeeldingen; pictogrammen	TIFF_UNC	TIFF_ZIP; PNG
Fotografische afbeeldingen	bitonaal, grijswaarden of polychromatisch; graduele kleurovergangen	foto's	TIFF_UNC; TIFF_UNC_EXIF	TIFF_ZIP; PNG; J2K_C_LL
Typografische afbeeldingen	bitonaal of grijswaarden; meerdere pagina's in één bestand	ingescande tekst (zonder OCR)	TIFF_UNC; TIFF_G4	J2K_C_LL
Geografische afbeeldingen	bitonaal, grijswaarden of polychromatisch; geografische metadata; georeferreed	landkaarten; orthofoto's	baseline GeoTIFF	geen
Vectorbeeldbestanden	Karakteristieken	Voorbeelden	Voorkeursformaat	Alternatief formaat
Creatieve afbeeldingen	lijntekeningen; eenvoudige kleurpaletten	clipart; animaties; diagrammen; grafieken; pictogrammen	SVG (zonder Java-binding)	geen
CAD-afbeeldingen	2-dimensioneel of 3-dimensioneel; <i>object linking</i>	technische ontwerpen	geen	DXF; STEP; PDF/E
Composietbeeldbestanden	Karakteristieken	Voorbeelden	Voorkeursformaat	Alternatief formaat
Vectorafbeeldingen met ingebedde rasterafbeeldingen	met <i>object linking</i> ; zonder CSS-ondersteuning	technische handleidingen; technische ontwerpen	WebCGM_G4; WebCGM_PNG	SVG
	zonder <i>object linking</i> ; met CSS-ondersteuning	idem	SVG	WebCGM_G4; WebCGM_PNG

**Tabel 2. Geschikte bestandsformaten voor moederkopieën**