

Datapreparatie

In de komende oefening wordt er gekeken naar verschillende veel voorkomende problemen met data ten behoeve van vectoranalyses. Verschillende vormen van datapreparatie komen aan bod.

De oefening bestaat uit allerlei korte losse onderdelen en geven een goed overzicht in verschillende manieren om met data om te gaan en tegen welke problemen u aan kunt lopen bij het uitvoeren van een analyse.

Alle data die we in deze oefening gaan gebruiken voor de analyse is te vinden in het mapje: C:\EsriTraining\VAAP. Download eventueel de oefendata uit het de leeromgeving en pak het zipbestand uit en kopieer het naar de hier boven genoemde locatie.

De oefening bestaat uit de volgende stappen:

| | | |
|---------|---|----|
| Stap 1: | Instellingen ArcGIS Pro project (environments)..... | 1 |
| Stap 2: | Coördinaten aanpassen | 4 |
| Stap 3: | Check Geometry / Repair Geometry..... | 8 |
| Stap 4: | Compact Geodatabase | 9 |
| Stap 5: | Cells uit CAD data halen | 10 |

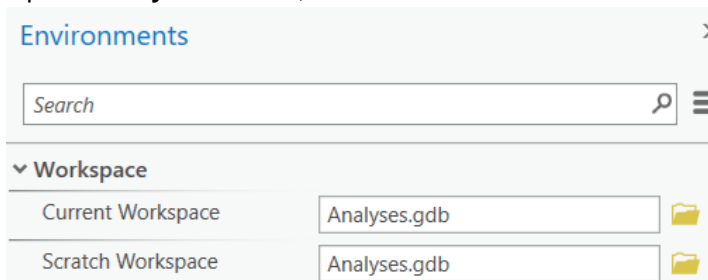
Stap 1: Instellingen ArcGIS Pro project (environments)

Bij het werken met analyses (maar ook in andere gevallen) is het goed om van te voren de juiste instellingen te verzorgen in het ArcGIS Pro Project. Op deze manier kunt u efficiënt te werk gaan en zorgt u er van tevoren voor dat alle data op juiste plek wordt opgeslagen, het coördinatensysteem goed staat ingesteld etc. In deze oefening worden een aantal van deze instellingen doorlopen en toegelicht.

- ☐ Indien nodig, download de oefendata, aanwezig in de leeromgeving bij de eerste oefening, en pak deze uit in de folder **C:\EsriTraining\VAAP**.
- ☐ In de folder **C:\EsriTraining\VAAP\Analyses**, open het ArcGIS Pro project **Analyses.aprx**.
- ☐ Maak (indien nodig) een folderconnectie in het catalog venster met de locatie waar de cursusdata is opgeslagen. **Folders > Add Folder Connection > C:\EsriTraining\VAAP**
- ☐ Voor deze oefening staat de data in verschillende folders en geodatabases. Het is echter goed om de data die wordt gemaakt tijdens de analyses op te slaan op één plek. *Default* wordt er al een geodatabase aangemaakt bij het aanmaken van het ArcGIS Pro project met dezelfde naam als het Pro Project

U gaat nu de instellingen van uw ArcGIS Pro Project bekijken in de **Environment Settings**.

- Op het **Analysis** tabblad, klik **Environments**.

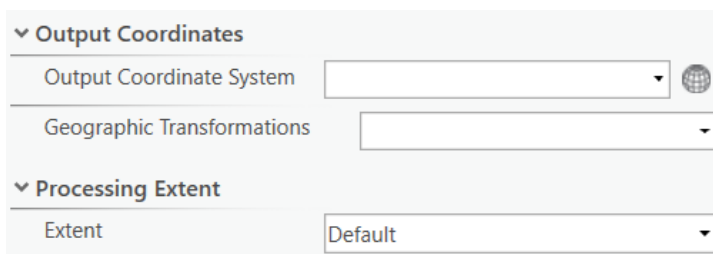


- Alle output van de geoprocessing tools wordt opgeslagen in de **Analyses** geodatabase. Dit laten we voor nu zo staan. Het is echter goed om te weten dat u de default database hier kunt wijzigen, mocht het voor andere projecten relevant zijn.



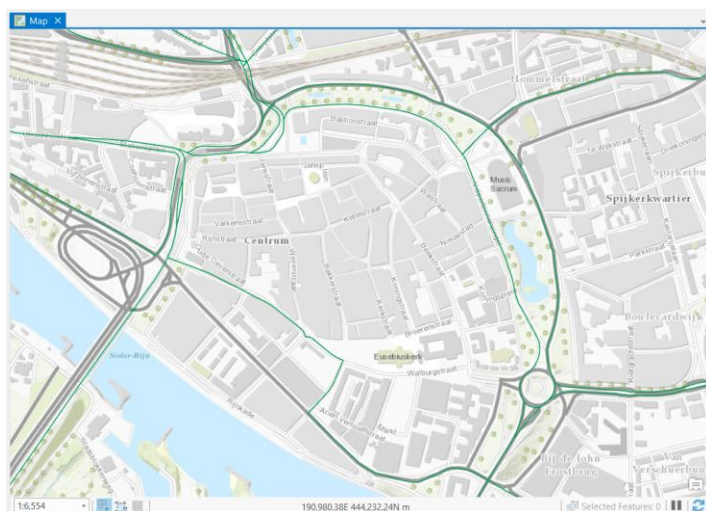
De **Scratch Workspace** kan verschillen van de huidige opslagplaats voor data, bijvoorbeeld voor data die u niet wilt bewaren. Deze optie is met name functioneel wanneer er gebruik gemaakt wordt van Python scripts of de ModelBuilder. Deze onderdelen komen in deze training niet aan bod.

- In het **Environments** venster zijn ook zaken vast te leggen over de **Output Coordinates** en de **Processing Extent**. Zo kunt u de output coördinaten van alle data al van te voren vastleggen. De **Processing Extent** definieert welke featureklassen of rasters (niet van toepassing in deze training) ArcGIS Pro meeneemt in de tools. Dit is vooral nuttig bij grote datasets waar u maar een klein deel van wilt gebruiken. Alleen de features die binnen de *processing extent* vallen worden meegenomen in de tool en gebruikt voor het creëren van de output.



We gaan nu kijken wat voor effect de processing extent heeft in onze analyse door een buffer analyse twee keer uit te voeren. Eenmaal met en eenmaal zonder processing extent.

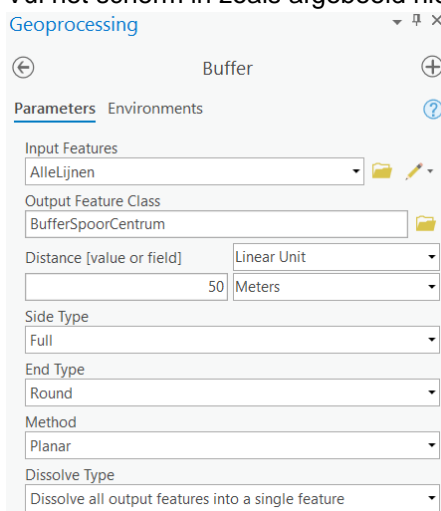
- Sluit het **Environments** venster.
- Open een nieuwe map in uw project.
- Uit de **Arnhem.gdb (C:\EsriTraining\VAAP\Analyses)** voeg de featureklasse **Alle Lijnen** toe aan de kaart (**Arnhem.gdb > Openbaar Vervoer**).
- Zoom in op het centrum van Arnhem (ongeveer schaal 1: 6500).



- ❑ Op het **Analysis** tabblad, selecteer **Environments** en stel de **Processing Extent** in als **Current Display Extent** .
- ❑ Klik op **OK**.

Nu gaan we buffers maken rondom deze spoorlijnen met de zojuist ingestelde opties.

- ❑ Op het **Analysis** tabblad (of in de Toolbox), selecteer de **Buffer** tool.
- ❑ Vul het scherm in zoals afgebeeld hieronder.



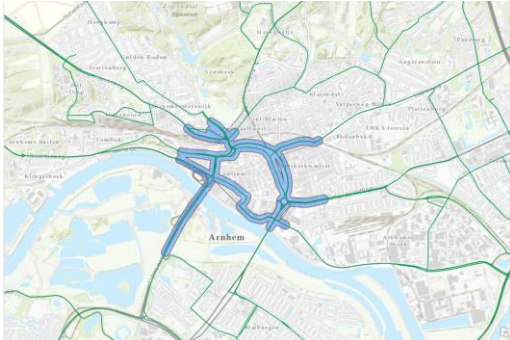
Zoals u kunt zien wordt de nieuwe laag (**BufferSpoorCentrum**) opgeslagen in **Analyses.gdb** (zoals ingesteld in de environment omgeving).

- ❑ Klik op **Run**.
- ❑ Bekijk het resultaat.
Zoals u kunt zien is de **Buffer** tool alleen uitgevoerd in het centrum van Arnhem. Alle lijnstukken die (gedeeltelijk) binnen de extent vallen zijn meegenomen in de analyse. Alle andere lijnstukken niet.
- ❑ Ga terug naar het **Environments** scherm en kies nu bij processing extent voor **Default**.

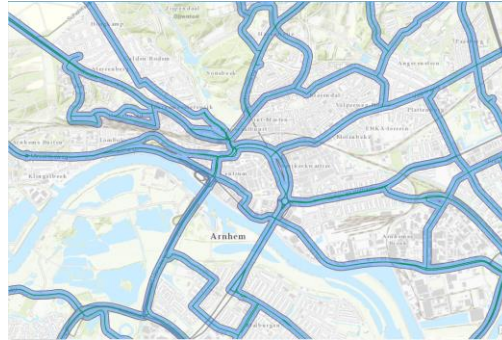
- ❑ Voer opnieuw de **Buffer** tool uit, ditmaal noemt u het resultaat **BufferSpoorArnhem**.

- ❑ Bekijk het resultaat.

De buffers zijn nu rondom alle spoorlijnen in heel Arnhem uitgevoerd.

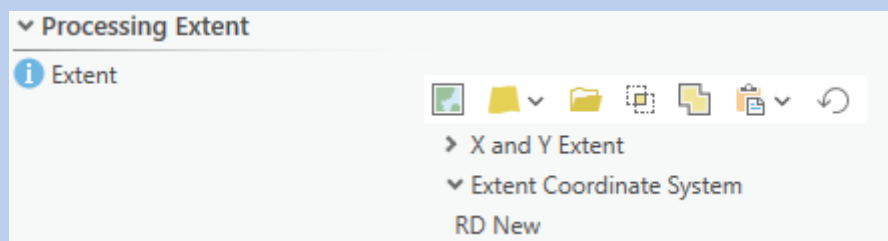


Current Display Extent



Default Extent

Het instellen van de extent kan handig zijn wanneer u met grote datasets werkt en de analyses niet wilt uitvoeren voor een groot gebied, maar een kleiner gedeelte van dit gebied. Naast de Current Display Extent, zijn er nog meer instellingen voor het instellen van de extensie: zo kunt u bijvoorbeeld de extensie instellen gelijk aan een andere kaartlaag, de intersectie van de input, etc.



Zie voor meer informatie: <http://pro.arcgis.com/en/pro-app/tool-reference/environment-settings/output-extent.htm>

Stap 2: Coördinaten aanpassen

Als u in een map werkt met verschillende coördinatensystemen kan ArcGIS Pro dit voor u 'on-the-fly' transformeren. Stel u heeft eerst een dataset waarvan het coördinaatsysteem Rijksdriehoekstelsel (RD_New) is. Vervolgens voegt u een dataset toe die met een GPS is ingewonnen en volgens het coördinaatsysteem WGS1984 (geografische coördinaten in lengte- en breedtegraden) is opgeslagen. Een transformatie is dan nodig om de data 'on-the-fly' op de juiste manier te projecteren.

Bedenk echter dat de originele data nog steeds in RD_New staat. De toegevoegde data staat in WGS1984 en wordt dus 'on-the-fly' getransformeerd. Dit kost altijd wat rekenkracht en dus tijd, en meer naar gelang er vaker wordt gezoomd en gepand. Voor de beste performance is het verstandig om te zorgen dat de kaart en de kaartlagen hetzelfde coördinaatsysteem hebben en dus niet 'on-the-fly' getransformeerd hoeft te worden. Zeker wanneer u aan de slag gaat met analyses is dit van belang voor een zo goed mogelijke performance.



Er zijn twee soorten tools van belang bij het aanpassen en toekennen van coördinatensystemen.

Define Projection tool - voegt coördinatensysteem toe aan een featureklasse.

Met deze tool wordt er geen nieuwe featureklasse aangemaakt maar een ontbrekend of incorrect coördinatensysteem gewijzigd.

Project tool – kent nieuw coördinaatsysteem toe.

Maakt nieuwe featureklasse aan (originele featureklasse & coördinatensysteem blijft bewaard). De tool verandert de coördinaten van de features en dit kan belangrijk zijn bij editen en analyses (voor het verbeteren van de performance).

In deze oefening zullen we werken met de **Project tool**.

Voordat de oefening verder start, moet u de volgende applicatie-instelling controleren betreffende coördinatensystemen.

- ❑ Open een nieuwe map.
- ❑ Selecteer **Project > Options > Map and Scene**.
- ❑ Controleer of de **Basemap** ingesteld staat op **Topo RD**.



Standaard is de optie **Default basemap of your organization** actief. Dit wil zeggen dat de basemap die uw ArcGIS Online portaalbeheerder heeft ingesteld, ook gebruikt wordt in ArcGIS Pro. Wanneer u gebruik maakt van het Opleidingen ArcGIS Online portaal (en inlogt in ArcGIS Pro met een van de student-accounts), dan is de default basemap **Topo RD**. U hoeft in dat geval de opties in dit scherm niet aan te passen.

U gaat nu een nieuwe map aanmaken met data uit twee verschillende coördinatensystemen in dit project.

- ❑ In het Catalog paneel, blader naar **DataOpOrde.gdb (C:\EsriTraining\VAAP\Data_Preparatie)** en voeg de featureklasse **FietsRouteLijnWGS84** toe.
- ❑ Indien een waarschuwing over het 'on the fly' transformeren van de data verschijnt, klik op **OK**.

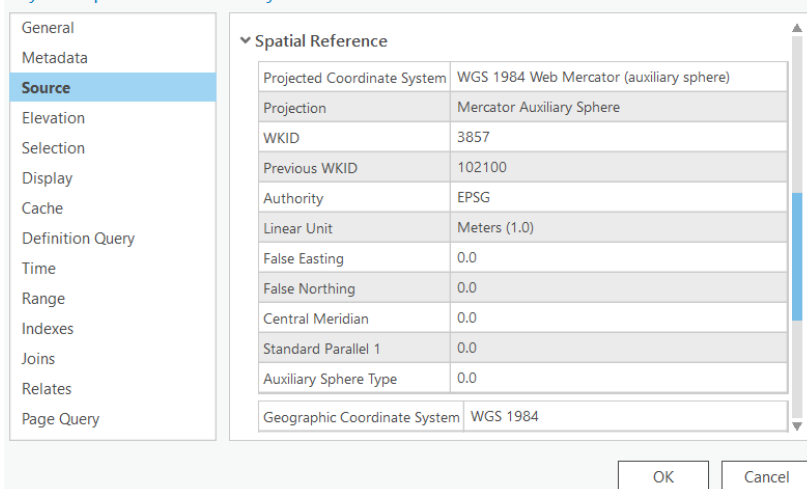


Default verschijnt deze melding niet in ArcGIS Pro. U kunt dit echter wel instellen bij **Project > Options > Map and Scene > onder Spatial Reference: Warn if transformation between coordinate system is required to align data sources correctly**.

- ❑ We gaan nu het coördinatensysteem van de kaartlagen bekijken.
- ❑ In het **Content** paneel, rechts-klik op de laag **FietsRouteLijnWGS84** en selecteer **Properties**.
- ❑ Klik op het **Source** tabblad.

- Blader naar beneden en klap het **Spatial Reference** menu open.

Layer Properties: FietsroutelijnWGS84



| Spatial Reference | |
|------------------------------|--|
| Projected Coordinate System | WGS 1984 Web Mercator (auxiliary sphere) |
| Projection | Mercator Auxiliary Sphere |
| WKID | 3857 |
| Previous WKID | 102100 |
| Authority | EPSG |
| Linear Unit | Meters (1.0) |
| False Easting | 0.0 |
| False Northing | 0.0 |
| Central Meridian | 0.0 |
| Standard Parallel 1 | 0.0 |
| Auxiliary Sphere Type | 0.0 |
| Geographic Coordinate System | WGS 1984 |

Zoals te zien, zijn de lijnen geprojecteerd in het WGS 1984 Web Mercator systeem. Dit bestand is van oorsprong een GPX-bestand (data verzameld aan de hand van een GPS-tracker) en bevat een fietsroute door Zuid-Limburg en België.

De basemap (zie **Map Properties > Coordinate Systems**) staat echter ingesteld op het Nederlandse RD_New (o.a. te zien aan de harde grenzen van Nederland) en de rest van de data in de projecten ook. We willen dus het coördinatensysteem van de data aanpassen en niet alleen 'on-the-fly' transformeren (zoals nu het geval is).

- Klik op **OK**.



- Open de **Project** tool.



Tools op het **Analyse** tabblad.

- ❑ Vul het scherm in zoals hieronder afgebeeld. Kies bij **Output Coordinate System** voor **Current Map**. Dan verschijnt automatisch **RD_New** en de juiste transformatie.

Geoprocessing

Project

Parameters Environments

Input Dataset or Feature Class
 FietsrouteLijnWGS84

Input Coordinate System:
 WGS_1984_Web_Mercator_Auxiliary_Sphere

Output Dataset or Feature Class
 FietsrouteLijn_RD

Output Coordinate System
 RD_New

Geographic Transformation
 Amersfoort_To_WGS_1984_NTv2

☐ Preserve Shape

We kiezen er nu dus voor om een nieuwe featureklasse aan te maken, in het Nederlandse coördinatensysteem **RD_New** met de transformatie **Amersfoort_To_WGS_1984_NTv2**.

In plaats van deze transformatie 'on-the-fly' te laten toepassen elke keer als u zoomt in de kaart, wordt de transformatie toegepast op een nieuwe featureklasse.

- ❑ Klik op **Run**.
- ❑ Bekijk opnieuw het coördinatensysteem van de zojuist aangemaakte kaartlaag.



Dit staat onder **Properties**.

Layer Properties: FietsrouteLijn_RD

General
 Metadata
Source
 Elevation
 Selection
 Display
 Cache
 Definition Query
 Time
 Range
 Indexes
 Joins
 Relates
 Page Query

Extent

Spatial Reference

| | |
|------------------------------|----------------------|
| Projected Coordinate System | RD New |
| Projection | Double Stereographic |
| WKID | 28992 |
| Authority | EPSG |
| Linear Unit | Meters (1.0) |
| False Easting | 155000.0 |
| False Northing | 463000.0 |
| Central Meridian | 5.38763888888889 |
| Scale Factor | 0.9999079 |
| Latitude Of Origin | 52.15616055555555 |
| Geographic Coordinate System | Amersfoort |

OK Cancel

Het veranderen van de projectie is gelukt en de data is nu opgeslagen in de database als **RD_New** data, maar ook de originele data is nog bewaard gebleven. De transformatie die is gebruikt, is dezelfde voor het transformeren terug naar WGS 1984 data (en ook dezelfde die wordt gebruikt om 'on-the-fly' te projecteren). Deze tool is goed om toe te passen wanneer u aan de slag gaat met analyses in verschillende projectiesystemen.

Stap 3: Check Geometry / Repair Geometry

Esri heeft twee tools die kunnen controleren of de geometrie correct is opgeslagen en dit zo nodig repareren. Incorrecte geometrie komt wel eens voor wanneer de data is aangemaakt in bijvoorbeeld CAD-systemen of wanneer de data met zelfgemaakte converters naar een GIS-omgeving wordt omgezet.



Onder niet 'correct opgeslagen' kunnen een aantal zaken worden verstaan. In zwart staan de fouten die het vaakst zullen voorkomen.

Short segment—Segmenten die korter zijn dan de toegestane systeemeenheden (xy tolerantie) van de gebruikte geodatabase.

Null geometry—De feature heeft geen geometrie of geen informatie in de SHAPE kolom.

Incorrect ring ordering—Bij een polygoon met een gat (een donut) moet de ordening van de vertices correct zijn (de buitenring—met de klok mee, binnenring—tegen de klok in).

Incorrect segment orientation—Wanneer een lijn uit meerdere segmenten bestaat moeten alle segmenten de juiste richting op wijzen.

Self intersections—Een polygoon mag zichzelf niet kruisen.

Unclosed rings—De laatste vertex in een ring (polygoon) moet samenvallen met de vertex van het eerste segment. De polygoon moet gesloten zijn.

Empty parts—De geometrie bestaat uit meerdere delen (multi-parts) en één van hen is leeg (has no geometry).

Duplicate vertex—De geometrie heeft twee of meer vertices met dezelfde coördinaten.

Mismatched attributes—De Z- of M-coördinaat van het eindpunt van een lijnsegment valt niet samen met het beginpunt van de volgende Z- of M-coördinaat.

Discontinuous parts—Eén van de geometriedelen bestaat uit 'disconnected' of discontinue delen.

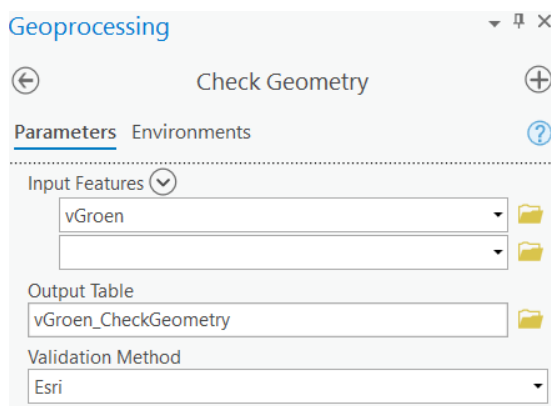
Empty Z values—De geometrie heeft één of meer vertices met een lege Z-waarde (NaN, bijvoorbeeld).

Bad envelope—De extent van de dataset komt niet overeen met de extent horend bij het coördinatensysteem.

Bad dataset extent—De extent van de dataset bevat niet alle features uit deze dataset. M.a.w. er worden buiten de extent elementen opgeslagen.

- ☐ Zoek de **Check Geometry** tool en open deze.
- ☐ Voor **Input Features**, selecteer **vGroen (C:\EsriTraining\VAAP\Data_Preparatie\DataOpOrde.gdb)**, een (oud) bestand afkomstig van Rijkswaterstaat.

- ❑ Voor **Output Table**, typ `vGroen_CheckGeometry` en sla deze op in de **DataOpOrde.gdb**.



- ❑ Klik **Run**.

Afhankelijk van de grootte van de dataset en de hoeveelheid bevindingen duurt dit even.

- ❑ Open de tabel `vGroen_CheckGeometry`.

De tabel toont nu de vlakken waar een 'self intersection' in zit. De kolom **Feature ID** verwijst naar het unieke nummer (het **ObjectID** van groen) van het element met het betreffende probleem. Met deze unieke nummers kunt u de 'self intersections' in **vGroen** eventueel gaan bekijken, al zullen deze mogelijk heel klein zijn.

Nu we weten dat er 'self intersections' voorkomen kunnen we deze ook gaan oplossen.

- ❑ Open de **Repair Geometry** tool.
- ❑ Voeg **vGroen** toe als input.
- ❑ Klik **Run**.

Na het runnen van de tool zijn alle onvolkomenheden opgelost en kan de data weer worden gebruikt worden. De tabel **vGroen_CheckGeometry** kan weer gekoppeld worden (join) aan de oorspronkelijke groenvlakken (**vGroen**). Op deze manier kan men ook inzicht krijgen in waar deze 'self intersections' zich bevinden.

Stap 4: Compact Geodatabase

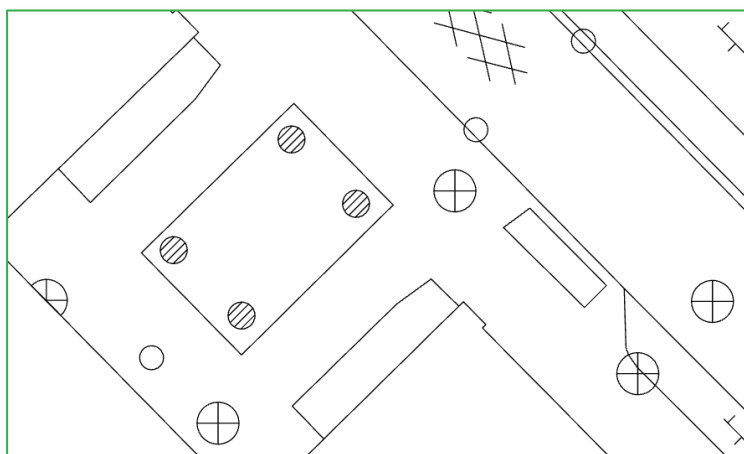
In het dagelijks gebruik van een geodatabase ontstaat veel 'lucht' in de achterliggende tabellen. Er worden tijdens het editen elementen ingetekend, weer weggegooid, soms kolommen toegevoegd en weer verwijderd. Bij deze aanpassingen wordt de database elke keer iets groter (maar niet automatisch kleiner). Om nu alle ruimte in de database weer optimaal te benutten, dient zo nu en dan een **Compact** te worden uitgevoerd.

- ❑ Rechtsklik op een geodatabase (bijvoorbeeld **Analyses.gdb**) en selecteer **Manage**.
- ❑ Een venster opent met een tabblad 'Manage geodatabase functionality'. Klik het vinkje voor **Compact** en vervolgens op **OK**.

Dit kan een of enkele seconden duren (wanneer dit al een paar maanden niet is gebeurd kan het enkele minuten duren). Wanneer het compacten regelmatig gebeurt zal de geodatabase snel blijven werken. Erg handig voorafgaand aan de analyses.

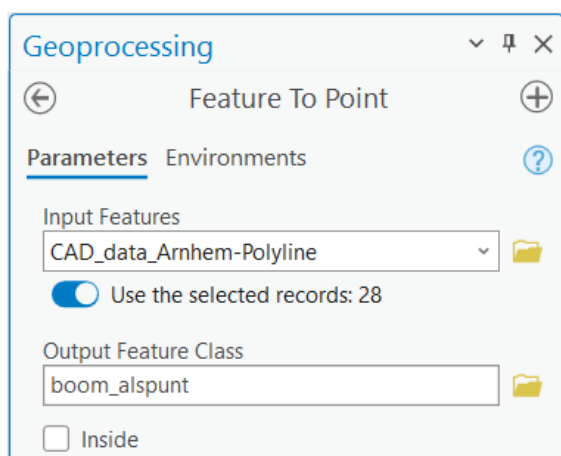
Stap 5: Cells uit CAD data halen

We komen in ons dagelijks werk regelmatig datasets tegen afkomstig uit CAD-applicaties. Dit kunnen applicaties zijn als AutoDesk of bijvoorbeeld MicroStation. Deze datasets zijn prima te gebruiken in onze analyses, echter soms zitten er zogenaamde cells of blocks in en dan is het raadzaam de data eerst te bewerken. De afbeelding hieronder geeft een beeld van Cells en Blocks.



De bomen, kolken en putdeksels zijn hier ingetekend als een Cell. Een Cell is als het ware een kleine tekening van het object. Het object wordt dus weergegeven als een verzameling lijnen. Voor analyses willen we echter liever puntobjecten wanneer het een boom, kolk of putdeksel betreft.

- ❑ Voeg de polyline CAD-data (**CAD_data_Arnhem.dgn > Polyline**) toe aan een nieuwe map.
- ❑ In het Contents paneel, klap de laag van de CAD-laag open en bestudeer de data.
- ❑ Merk op dat er aparte lagen zijn waaronder o.a. een aparte laag met bomen (boom_S_gbk).
- ❑ Maak met behulp van de **Select by Attributes** een selectie van de bomen.
- ❑ In het **Geoprocessing** paneel, open de **Feature To point** tool.



- ❑ Klik **Run**.
- ❑ Bekijk het resultaat.

De bomen kunnen nu veel mooier worden gevisualiseerd nu het punten zijn. Daarnaast zullen ze sneller zijn in analyses.

Einde oefening