

# COLLECTIES DUURZAAM REGISTREREN IN EEN REKENBLAD

Hoewel databanken stilaan ook bij heemkundige kringen ingeburgerd geraken, is voor veel heemkundigen Excel nog steeds applicatie nummer één om erfgoedcollecties te inventariseren en te beschrijven.

Eigenlijk spreken we beter over een rekenblad, maar de overgrote meerderheid gebruikt één specifiek computerprogramma, namelijk Microsoft Excel. Hoewel, de voorbije jaren zijn er met LibreOffice Calc en Google Spreadsheet twee alternatieven opgedoken met een groeiende groep gebruikers.

De reden voor het succes van rekenbladen is eenvoudig: Je vindt het programma op elke pc of laptop terug. Gegevens in rekenbladen zijn makkelijk doorzoekbaar, sorteerbaar en filterbaar. Een kopie van je collectie-inventaris is snel gemaakt en je kan hem meteen delen met je collega's en andere geïnteresseerden.

Maar inventarissen in rekenbladen zorgen ook voor heel wat hoofdbreken, vooral bij mensen die ze willen hergebruiken voor hun eigen onderzoek, of bij mensen die ze moeten omtoveren tot een mooie website of databank. Gebruikers raken moeilijk wijs uit de precieze betekenis van gegevens in je rekenblad. En op die website lijken willekeurige karakters verwisseld te zijn met de vreemdste tekens.

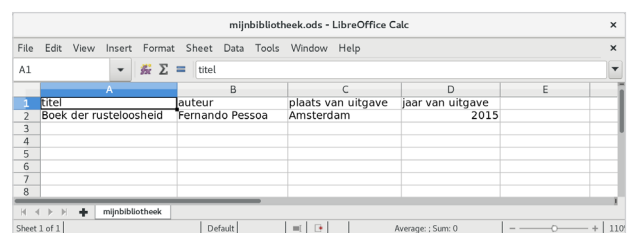
In veel gevallen is de oorzaak dat we rekenbladen te complex maken en er vanuit gaan dat gebruikers over heel wat impliciete kennis beschikken om het rekenblad te interpreteren - kennis die voor de maker evident is, maar voor de gebruiker natuurlijk niet altijd. Maar vaak is de oorzaak ook dat we rekenbladen opslaan in bestandsformaten die enkel het softwareprogramma waarmee ze gemaakt zijn correct kan uitlezen.

Dit artikel bevat een aantal tips hoe je voor onderzoekers en ontwikkelaars die met jouw rekenblad aan de slag gaan het leven een stuk eenvoudiger kan maken. Daarbij volgen we één leidend principe: je rekenblad dient 'zelfverklarend' te zijn. Dat betekent dat je de gegevens in je rekenblad zo organiseert dat onderzoekers of ontwikkelaars ze intuïtief juist kunnen interpreteren.

## PLATTE DATA

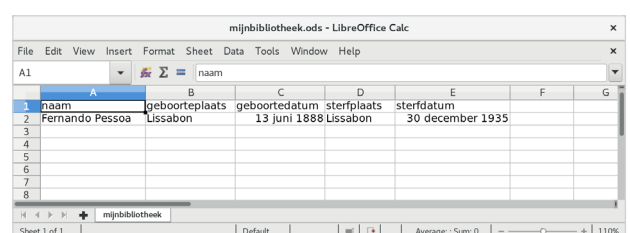
Regel één om je rekenblad 'zelfverklarend' te maken: maak je data zo 'plat' mogelijk. Dat wil zeggen dat je van een erfgoedobject enkel de eigenschappen documenteert die direct van toepassing zijn op het object.

Vermijd om die eigenschappen op hun beurt weer omstandig te documenteren. Een voorbeeld: Een boek heeft als eigenschappen een titel, een auteur, een plaats en een jaar van uitgave.



	A	B	C	D	E
1	titel	auteur	plaats van uitgave	jaar van uitgave	
2	Boek der rusteloosheid	Fernando Pessoa	Amsterdam	2015	
3					
4					
5					
6					
7					
8					

Maar over één van die eigenschappen, bijvoorbeeld de auteur, valt op zich ook heel wat te vertellen. Bijvoorbeeld waar en wanneer hij geboren en gestorven is:



	A	B	C	D	E	F	G
1	naam	geboorteplaats	geboortedatum	sterfplaats	sterfdatum		
2	Fernando Pessoa	Lissabon	13 juni 1888	Lissabon	30 december 1935		
3							
4							
5							
6							
7							
8							

Beide tabellen kan je ook samenvoegen tot één tabel.

titel	auteur	geboorteplaats	geboortedatum	sterfplaats	sterfdatum	plaats van uitgave	jaar van uitgave
1 Boek der rusteloosheid	Fernando Pessoa	Lissabon	13 juni 1888	Lissabon	30 december 1935	Amsterdam	2015

Maar dan meng je eigenschappen over twee verschillende zaken (het boek en zijn auteur) in één tabel. Dat is verwarrend voor gebruikers die niet met de inventaris vertrouwd zijn. Voor wie later met je data aan de slag wil is het handiger als de eigenschappen van auteurs in een aparte tabel zitten.

Voor ontwikkelaars is het eenvoudiger om die twee tabellen apart te importeren in een databank. En het is ook een stuk efficiënter om een geboorteplaats in een aparte lijst te documenteren: je hoeft dan voor elke auteur immers maar één keer de geboorteplaats vastleggen.

Een andere type informatie dat zich lastig in een rekenblad laat vatten zijn archiefinventarissen waarbij je eerst het volledige fonds beschrijft, vervolgens de grote series waaruit het archief bestaat, dan de afzonderlijk dossiers, enzovoort.

```

└─ Parochiearchief Sint-Laurentius Betekom (Begijnendijk) – fonds
  └─ Parochieregisters. Akten – Reeks
    └─ Parochieregisters. Doopakten – Reeks
      └─ 20 april 1611 - 1 november 1622 - Deelreeks
        └─ 8 maart 1616 - 12 februari 1642 - Deelreeks
      └─ Parochieregisters. Huwelijksakten - Reeks
    └─ Parochieregisters. Overlijdensakten -Reeks
  
```

Daarbij geldt de regel dat informatie die op een hoger niveau gedocumenteerd is, ook van toepassing is op lagere niveaus. Er zijn allerlei slimme trucjes te bedenken om dat soort van hiërarchische relaties tussen fonds, series en dossiers te documenteren in een platte tabel, maar die vereisen heel wat impliciete kennis om de tabel te interpreteren. En dat dien je te vermijden. Ook hier heb je de mogelijkheid om fonds, series en dossiers in aparte tabellen te documenteren die naar elkaar verwijzen.

Wanneer het voor jou belangrijk is om uitgebreid bepaalde aspecten van een object of hiërarchische

relaties te beschrijven, neem je beter een volwaardig databanksysteem in gebruik. Je bespaart jezelf dan heel wat tijd en frustratie.

## KOMMAGESCEIDEN BESTAND

Regel twee om je rekenblad zo eenvoudig mogelijk leesbaar te maken: sla je data op als een csv-bestand. Csv staat voor ‘comma-separated-values’ of ‘kommagescheiden waarden’. Een csv-bestand is een eenvoudig tekstbestand waarbij de horizontale en verticale lijnen van je rekenblad gecodeerd zijn met leestekens. Een verticale lijn van een kolom wordt typisch gecodeerd als een komma. Een horizontale lijn van een rij wordt typisch gecodeerd door een nieuwe regel.

inventarisnr	titel	auteur	uitgever	plaats van uitgave	jaar van uitgave
1	Boek der rusteloosheid	Fernando Pessoa	De Arbeiderspers	Amsterdam	2015
2	Wereldschandkroniek	Jorge Luis Borges	De Bezige Bij	Amsterdam	1970
3	Dode Zielen	Nicolai Vasil Gogol	Atheneum-Polak & Van Gemrep	Amsterdam	2011

```

"inventarisnr","titel","auteur","uitgever","plaats van uitgave","jaar van uitgave"
1,"Boek der rusteloosheid","Fernando Pessoa","De Arbeiderspers","Amsterdam",2015
2,"Wereldschandkroniek","Jorge Luis Borges","De Bezige Bij","Amsterdam",1970
3,"Dode Zielen","Nicolai Vasil Gogol","Atheneum-Polak & Van Gemrep","Amsterdam",2011

```

Microsoft Excel en LibreOffice Calc gebruiken hun eigen bestandsformaat om tabelgegevens op te slaan, met name .xls en .ods. Maar daarbij worden, naast de tabelgegevens, ook heel wat andere gegevens opgeslagen, zoals opmaakinformatie en formules. Om die tabellen correct te lezen, gebruik je bij voorkeur dezelfde applicatie als waarmee ze gemaakt zijn. Die afhankelijkheid van een specifieke applicatie vermijdt je best.

Voor je inventaris heb je die opmaakinformatie en formules in principe niet nodig. Een csv-bestand brengt het rekenblad dat je in Excel of Calc gemaakt hebt terug tot de essentie: een tabel met louter gegevens in rijen en kolommen.

Een belangrijke keerzijde bij het gebruik van csv-bestanden is echter dat csv geen formeel gestandaardiseerd bestandsformaat is. Er is dus geen universeel afgesproken standaard die vastlegt hoe je tabellen in tekst codeert. Je bent dus vrij om in plaats van komma's en nieuwe regels andere karakters als scheidingstekens te gebruiken voor rijen en

kolommen. Dat uiteenlopend gebruik van scheidingstekens is een van de belangrijkste bronnen van verwarring bij het gebruik van csv-bestanden.

Er is echter wel de ‘memo’ Common Format and MIME Type for Comma-Separated Values (CSV) Files van de Internet Engineering Taskforce (IETF), de gemeenschap van ingenieurs en technici die de architectuur en de goede werking van het internet bewaakt, dat een poging doet om formeel te beschrijven hoe je een csv-bestand samenstelt<sup>1</sup>. In de rest van dit artikel gebruiken we die memo als leidraad.

## EEN TABEL SAMENSTELLEN

Je opent je favoriete rekenbladprogramma. Voor je enthousiast de cellen van je tabel begint te vullen, stel je jezelf twee vragen:

- Wat zijn de namen van de kolommen in mijn rekenblad?
- Hoe stop ik meerdere waarden in één cel van mijn rekenblad?

### Kolomnamen

Kolomnamen staan in de bovenste rij van je tabel en geven de gebruiker een idee van de inhoud van elke kolom in je rekenblad. Het is de enige manier waarop je de gebruiker een soort van handleiding kan meegeven voor hoe hij het rekenblad moet lezen.

Bij de keuze van kolomnamen is het belangrijk om eenvoudige en heldere labels te kiezen die de inhoud van de kolom zo volledig mogelijk weergeeft. Een kolom met de naam ‘auteur’ bevat enkel en alleen namen van schrijvers, geen uitgevers, vertalers of illustratoren. Een kolom met de naam ‘hoogte’ bevat enkel de verticale dimensie van objecten. Maar heeft je kolom de naam ‘afmetingen’, dan ben je vrij eender welke dimensie van het object te documenteren.

Vind je het moeilijk om de juiste kolomnaam te bepalen? Of om te bepalen hoeveel kolommen je nu precies nodig hebt om je object te beschrijven? Heel wat mensen voor jou hebben die oefening al gemaakt. Hun oplossingen vind je terug in een aantal standaarden waaraan je kolomnamen kan ontlenen. De gangbare standaarden in de erfgoedsector vind je terug op CEST.<sup>2</sup>

Hieronder vind je een selectie standaarden die je makkelijk kunt gebruiken als bron voor je kolomnamen:

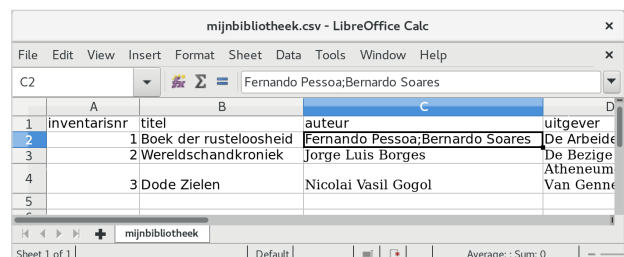
- Dublin Core<sup>2</sup> voor allerlei soorten erfgoedobjecten;
- MODS<sup>3</sup> voor publicaties;
- ISAD(G)<sup>4</sup> voor archiefbeschrijvingen
- Basisregistratie<sup>5</sup> voor museumobjecten

## MEERDERE WAARDEN IN EEN CEL

De tweede belangrijke vraag die je je moet stellen is hoe je meerdere waarden in een cel documenteert.

Een boek kan bijvoorbeeld meerdere auteurs hebben. Op zich is dat geen probleem: je schrijft gewoon twee namen achter elkaar in een cel. Maar wanneer je die namen achteraf in een databank wil stoppen, is het voor ontwikkelaars vaak moeilijk om die namen terug uit elkaar te halen. Ook hier kun jij alvast een handje helpen door een extra scheidingsteken in te voeren voor meerdere waarden in een cel.

Gebruik bij voorkeur een puntkomma om twee waarden van elkaar te scheiden, best zonder bijkomende spaties om de waarden van het leesteken te scheiden, bijvoorbeeld:



	A	B	C	D
1	inventarisnr	titel	auteur	uitgever
2	1	Boek der rusteloosheid	Fernando Pessoa;Bernardo Soares	De Arbeid
3	2	Wereldschandkroniek	Jorge Luis Borges	De Bezige
4	3	Dode Zielen	Nicolai Vasil Gogol	Atheneum
5				Van Gene

Als die twee vragen beantwoord zijn, kun je aan de slag met je rekenblad. Voeg indien nodig nieuwe kolommen toe, maar denk eraan om je data zo plat en eenvoudig mogelijk te houden.

## EEN CSV-BESTAND MAKEN

De volgende stap is het exporteren van je rekenblad als een csv-bestand. Zoals al vermeld, gebruiken sommige rekenbladen als Microsoft Excel en LibreOffice Calc hun eigen bestandsformaat om rekenbladen op te slaan. Dus je kan niet zomaar op de *Save/Opslaan-knop* drukken. Wanneer je een rekenblad opslaat zijn er vier zaken die je moet instellen of op zijn minst controleren voor je op *Save/Opslaan* drukt:

1. Verander het bestandsformaat naar *text/csv*. Hierdoor verwijder je alle opmaak en formules en hou je enkel de tabelgegevens over.
2. Bepaal of controleer de scheidingstekens voor kolommen en rijen. Stel deze in volgens de regels van de IETF Memo: komma's voor kolommen, nieuwe regel voor rijen.
3. Zet dubbele aanhalingstekens rond tekstwaarden. Op die manier vermijd je dat een applicatie in de war raakt wanneer je bijvoorbeeld komma's in een volzin gebruikt. Alle komma's tussen de dubbele aanhalingstekens worden dan genegeerd.
4. Stel de karaktercodering in op UTF-8. UTF-8 is een standaard codering waarmee karakters (letters, cijfers, leestekens) worden omgezet in bytes. UTF-8 zorgt ervoor dat jouw tabelgegevens in elke applicatie terug omgezet worden naar de juiste karakters.

Het is echter niet zo eenvoudig om deze vier zaken met behulp van je favoriete rekenbladprogramma toe te passen. De instellingen worden door elk programma anders benoemd en zitten soms heel diep verscholen in de gebruikersinterface.

Hieronder vind je stap-voor-stap instructies hoe je de vier bovenstaande zaken concreet instelt in Microsoft Excel, LibreOffice Calc en Google Spreadsheet. Daarbij wordt telkens aangegeven wat je wel of niet kan configureren en welke zaken het programma automatisch uitvoert.

## MICROSOFT EXCEL

De instructie voor Microsoft Excel vergt enige aandacht, onder andere omdat je ook in je besturingsstelsel een instelling moet aanpassen. Maar als je deze instructie één keer hebt uitgevoerd, onthoudt je computer veel van deze instellingen en gaat het de volgende keer een stuk sneller. We gaan ervan uit dat je Excel gebruikt in een recente Windows omgeving (Windows 7 of hoger).

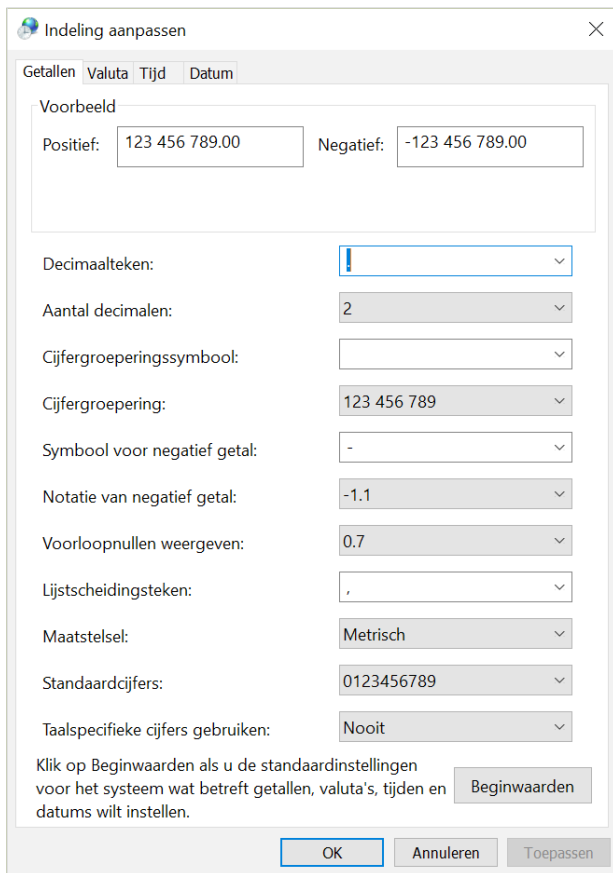
Eerst stel je het scheidingsteken voor kolommen in. Dat doe je niet in de instellingen van Windows Excel, maar via het Configuratiescherm van je Windows omgeving:

1. Klik op de *Start-knop* > type en/of selecteer *Configuratiescherm* > Klik op *Klok, taal en land/regio* > Klik op *De notatie voor datums, tijd of getallen wijzigen* > Klik op de knop *Meer instellingen...*

Er opent nu een schermje waarop je het lijstscheidingsteken kan instellen. Dat is het scheidingsteken voor kolommen in Excel. Bij Nederlandstalige Windows-computers staat dit standaard ingesteld op puntkomma.

2. Verander het *lijstscheidingsteken* in een komma. Opgelet! Controleer ook wat er in het veld *Decimaalteken* staat bovenaan in de lijst. Als daar ook een komma staat, dan zal Excel voor csv-bestanden alsnog een puntkomma gebruiken.

Dus als er bij *Decimaaltekens* ook een komma staat, verander die dan bijvoorbeeld in een punt.



Druk tweemaal op de *OK-knop* en open *Excel* om je rekenblad te maken.

Als je klaar bent met gegevens in te voeren doe je het volgende:

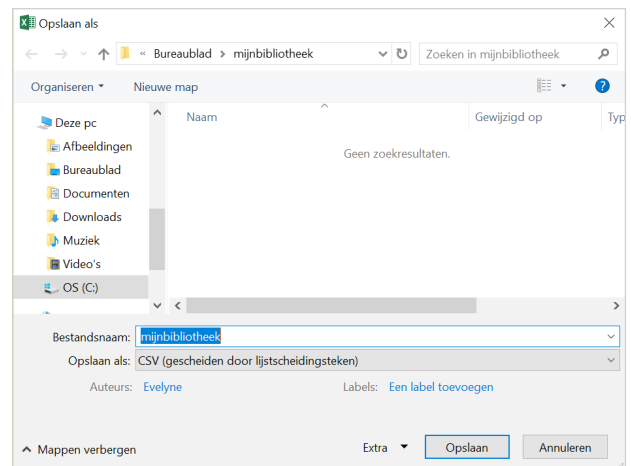
3. Klik *Bestand* > *Opslaan als* > Selecteer de map waar je het csv-bestand wil opslaan.
4. Bepaal in het vakje *Bestandsnaam*: de naam van het csv-bestand. Bijvoorbeeld 'mijncollectie'.
5. Klik op het uitklapmenu *Opslaan als*: en selecteer *CSV (gescheiden door lijstscheidingstekens)*.

Hiermee heb je bepaald dat je tabel als een tekstbestand wordt opgeslagen en dat het *lijstscheidingstekens* - de komma dus - gebruikt wordt om kolommen te coderen.

Merk op dat je de codering van de verticale lijnen niet kan instellen. Dat gebeurt in Microsoft Excel steeds met een nieuwe regel.

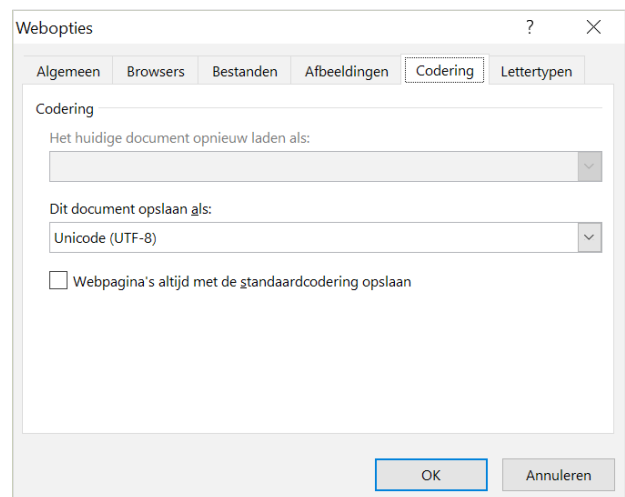
Merk ook op dat je geen dubbele aanhalingstekens

kan instellen voor tekstwaarden. Excel voegt automatisch dubbele aanhalingstekens toe aan waarden waar een komma of een dubbel aanhalingsteken instaat.



6. Onderaan naast de *Opslaan-knop* bevindt zich nog een uitklapmenu *Extra*. Klik op dit menu en selecteer *Webopties...* Er opent een nieuw scherm.

7. Ga naar het tabblad *Codering* en stel het veld *Dit document opslaan als*: in op *Unicode (UTF-8)*.

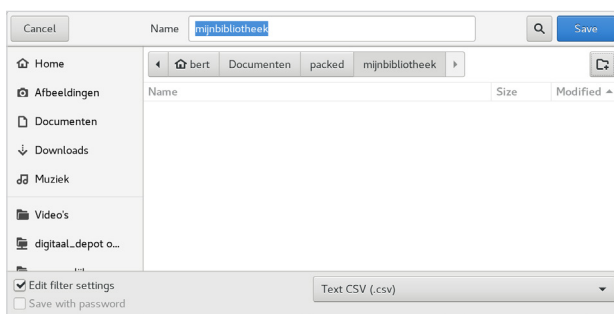


8. Druk op de *Ok-knop* en vervolgens op *Opslaan*. Het csv-bestand wordt in de map van jouw keuze aangemaakt.

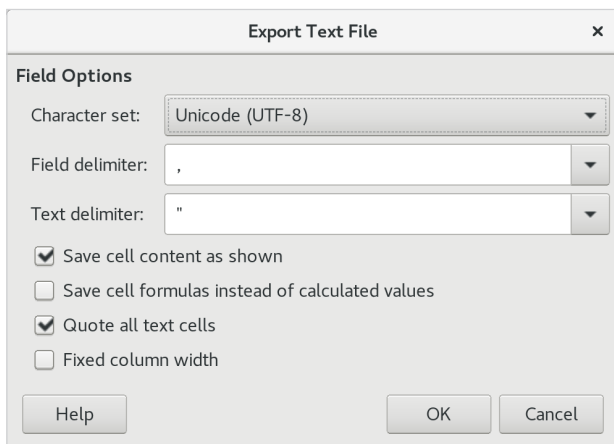
## LIBREOFFICE CALC

Dit is het meest gebruikte open source rekenblad. De grote sterkte van *Calc* in vergelijking met de andere rekenbladen is dat je vrijwel alle parameters om je csv-bestand vorm te geven van configureren.

1. Klik File > Save As... > Selecteer de map waar je het csv-bestand wil opslaan.
2. Bepaal in het vakje *Name*: bovenaan het scherm de naam van het csv-bestand. Bijvoorbeeld 'mijncollectie'.
3. Klik rechtsonder op het uitklapmenu waar je het formaat van je bestand kan bepalen en selecteer *Text CSV (.csv)*.
4. Linksonder staan er twee aanvinkvakjes. Vink het vakje *Edit Filter Settings* aan, want daarmee ga je de scheidingstekens en de karaktercodering kunnen instellen.



5. Klik nu op de *Save* knop rechtsboven. Er opent een nieuw schermje:



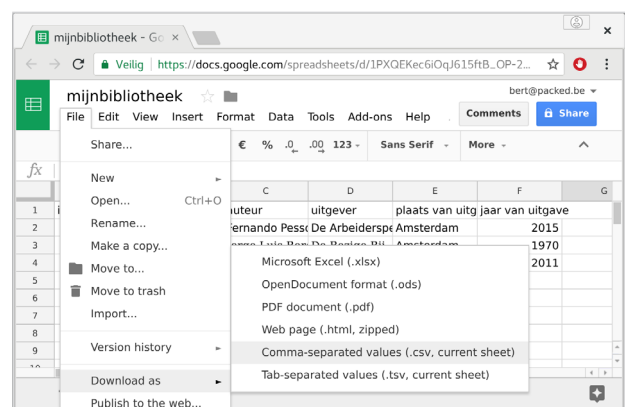
Op dit schermje kan je de andere parameters instellen:

6. Stel het veld *Character set*: in op *Unicode (UTF-8)*.
  7. Stel het veld *Field delimiter*: in op [,] om de verticale lijnen van je tabel te coderen met een komma.
  8. Stel het veld *Text delimiter*: in op [“] en vink het vakje *Quote all text cells* aan om alle tekstwaarden tussen dubbele aanhalingstekens te zetten.
- Merk op dat je de codering van de verticale lijnen niet kan instellen. Dat gebeurt in *LibreOffice Calc* steeds met een nieuwe regel.
9. Druk op de *OK-knop* onderaan en je csv-bestand wordt gemaakt.

## GOOGLE SPREADSHEET

Google Spreadsheet is geen computerprogramma dat je installeert om je eigen pc of laptop. Het is een webdienst waarvan je gebruik kan maken via je webbrowser. In vergelijking met de andere rekenbladen kan je hier vrijwel niets configureren, maar Google Spreadsheet exporteert je rekenblad wel volledig volgens de regels van de IETF Memo.

1. Klik File > Download as > Comma-separated values (.csv, current sheet)



2. Je csv-bestand wordt aangemaakt in de downloadmap van je browser.

Wanneer je het csv-bestand opent in een eenvoudige tekstverwerker zal je merken dat Google Spreadsheet csv-bestanden automatisch aanmaakt volgens de regels van de IETF memo: komma's voor

tabellen, nieuwe regels voor rijen. De applicatie detecteert automatisch tekstwaarden en zet ze tussen automatische aanhalingstekens.

## BESLUIT

Tot slot nog eens de zeven deugden van een 'zelf-verklarend' csv-bestand op een rij:

1. Maak je data zo plat mogelijk.
2. Sla je rekenblad op als een CSV-bestand.
3. Gebruik eenduidige, liefst standaard kolomnamen.
4. Gebruik puntkomma's om meerdere waarden in een cel van elkaar te scheiden.
5. Gebruik de UTF-8 karaktercodering.
6. Gebruik komma's om kolommen en nieuwe regels om rijen van elkaar te scheiden.
7. Gebruik dubbele aanhalingstekens rond tekstwaarden.

*Microsoft Excel, LibreOffice Calc en Google Spreadsheet* zijn de meest gebruikte rekenbladen. Alle drie kunnen ze een csv-bestand produceren dat aan deze zeven deugden beantwoordt. Ze verschillen echter in de mate dat parameters handmatig kan instellen en de mate waarin het programma automatisch scheidingstekens toevoegt.

Voor de bewaring van gegevens gedurende een lange periode is het interessant om nu en dan een csv-bestand te maken van de rekenbladen die men maakt.

**Bert Lemmens**  
**Packed vzw**

## EINDNOTEN

1. Voor meer info, zie <https://tools.ietf.org/html/rfc4180>
2. <http://projectcest.be>
3. [https://www.projectcest.be/wiki/Standaard:Dublin\\_Core](https://www.projectcest.be/wiki/Standaard:Dublin_Core)
4. <https://www.projectcest.be/wiki/Standaard:MODS>
5. [https://www.projectcest.be/wiki/Standaard:ISAD\(G\)](https://www.projectcest.be/wiki/Standaard:ISAD(G))
6. <https://www.projectcest.be/wiki/Standaard:Basisregistratiekaart>

