BINARY ENCODED FILES

TEXT DATA
ASCII/UTF-8 ENCODED FILES

EXTENSIONS

PICTURES (JPG, PNG…)

MOVIES (MP4, MOV…)

AUDIO (MP3, WAV, AIFF…)

OFFICE FILES (DOCX, PPTX…)

ZIP

PDF

…

PROPRIETARY AND/OR PATENTED ENCODING/DECODING TECHNIQUES

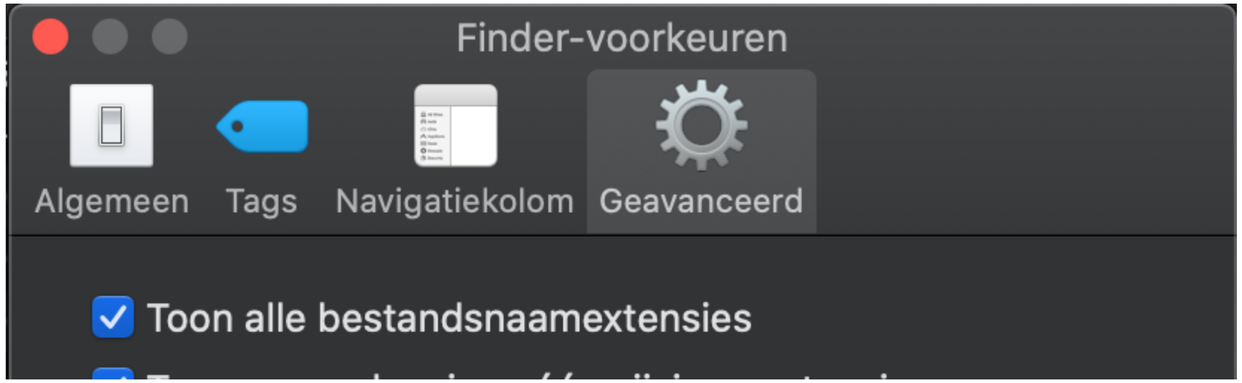
NO REAL "EXTENSION"
(You can use anything you want)

NO PROPRIETARY ENCODING/DECODING

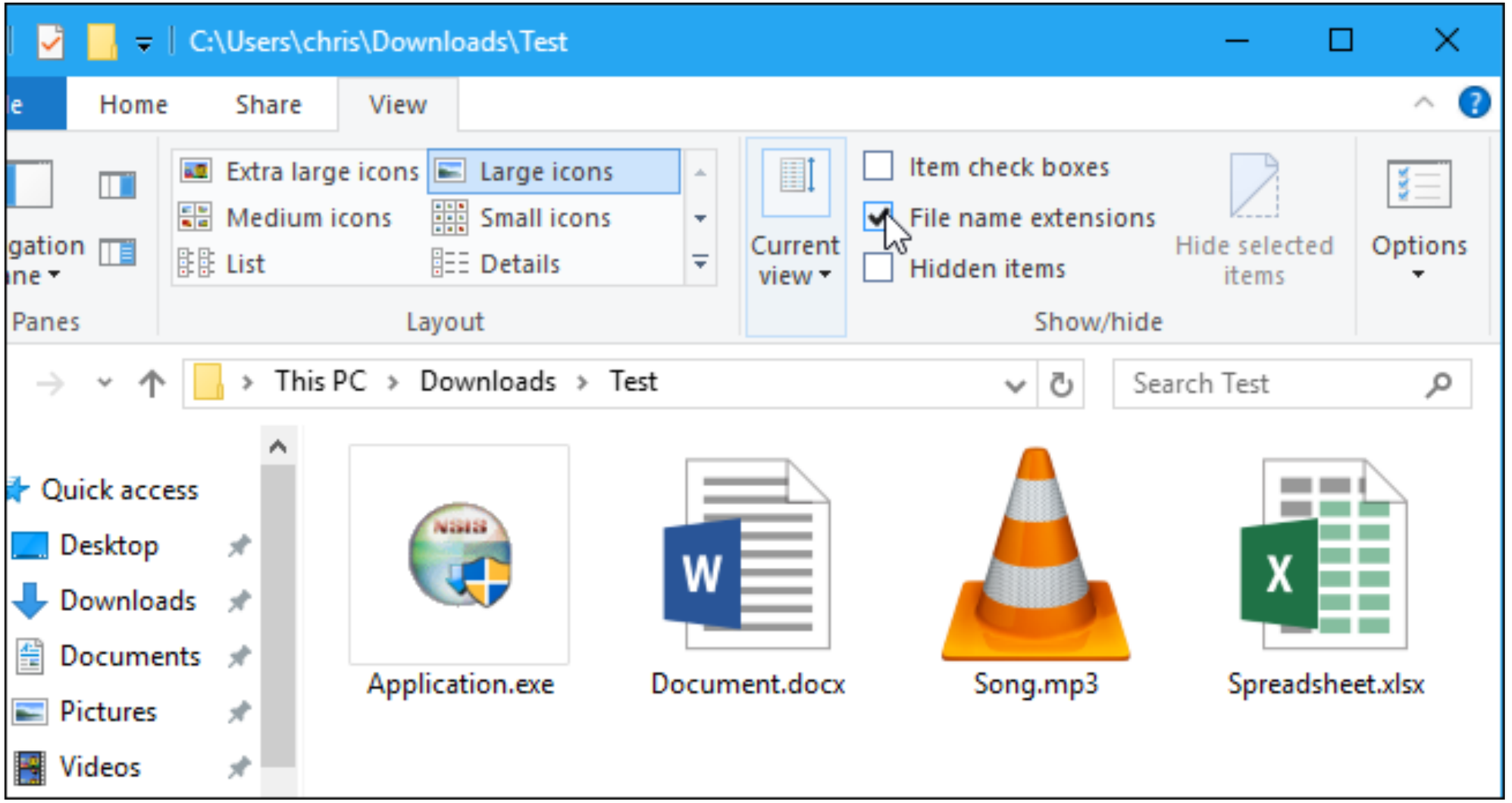
Very important for a data specialist:
make the extensions visible on your computer

Mac OS X


Finder
Archief
Wijzig
Over Finder
Voorkeuren... ⌘,
Leeg prullenmand... ⇧⌘⌫
Voorzieningen ▶
Verberg Finder ⌘H
Verberg andere ⌥⌘H
Toon alles
Documenten
Schermafbeeldin


Finder-voorkeuren
Algemeen
Tags
Navigatiekolom
Geavanceerd
Toon alle bestandsnaamextensies

Windows


C:\Users\chris\Downloads\Test
Home
Share
View
Extra large icons
Large icons
Medium icons
Small icons
List
Details
Panes
Layout
Current view
Item check boxes
File name extensions
Hidden items
Hide selected items
Show/hide
Options
This PC
Downloads
Test
Search Test
Quick access
Desktop
Downloads
Documents
Pictures
Videos
Application.exe
Document.docx
Song.mp3
Spreadsheet.xlsx

Saving this data as ASCII / UTF8
would make the files too large.

TEXT DATA
ASCII/UTF-8 ENCODED FILES

We dive into "text files''
ASCII/UTF-8 ENCODED FILES

You can open these types of files in a
text editor such as Notepad or Brackets.

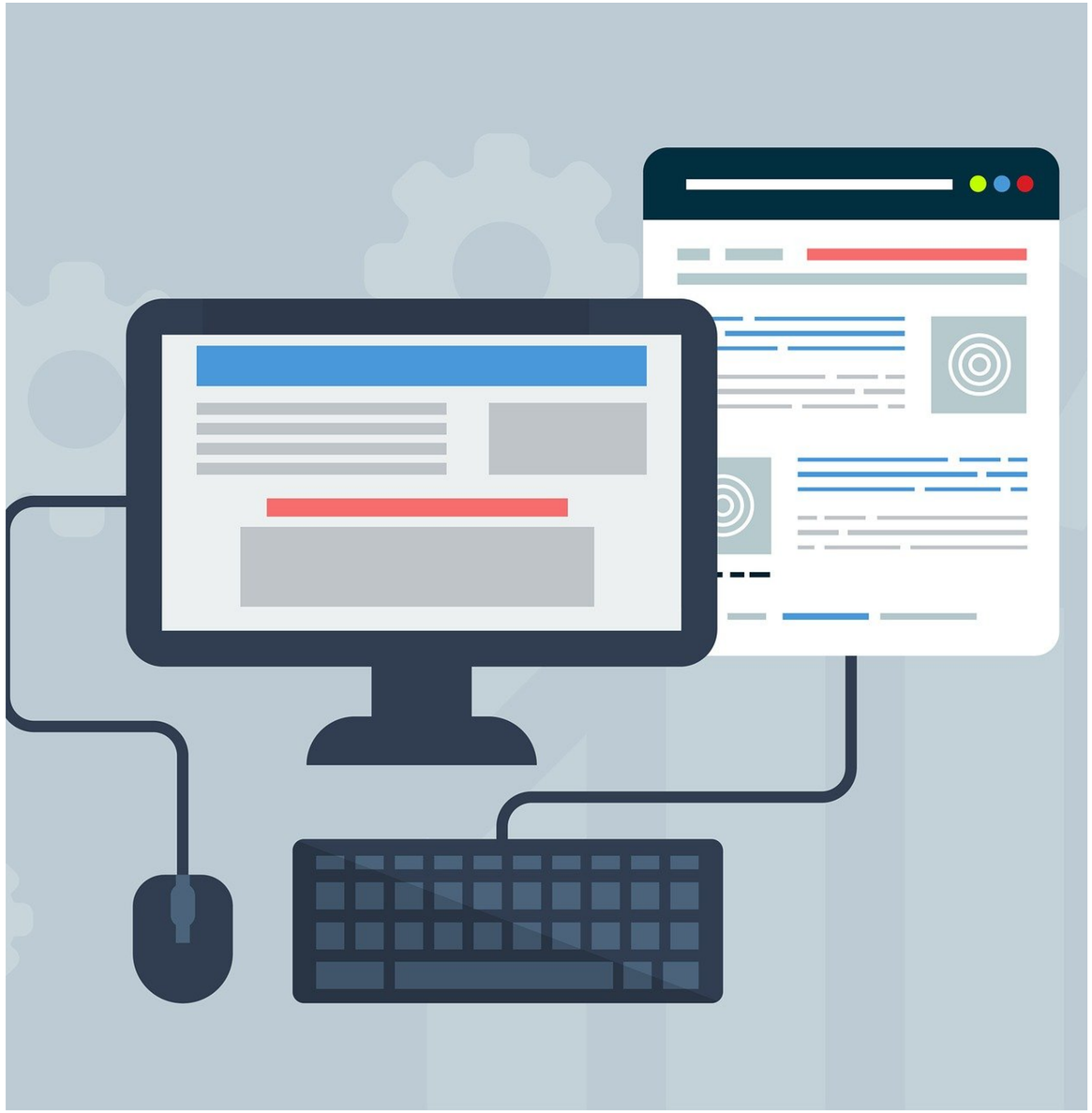
You don't need any special "software"
to "**decrypt**" or **decode** these types of files.

“Flat” file

Just text content or numbers

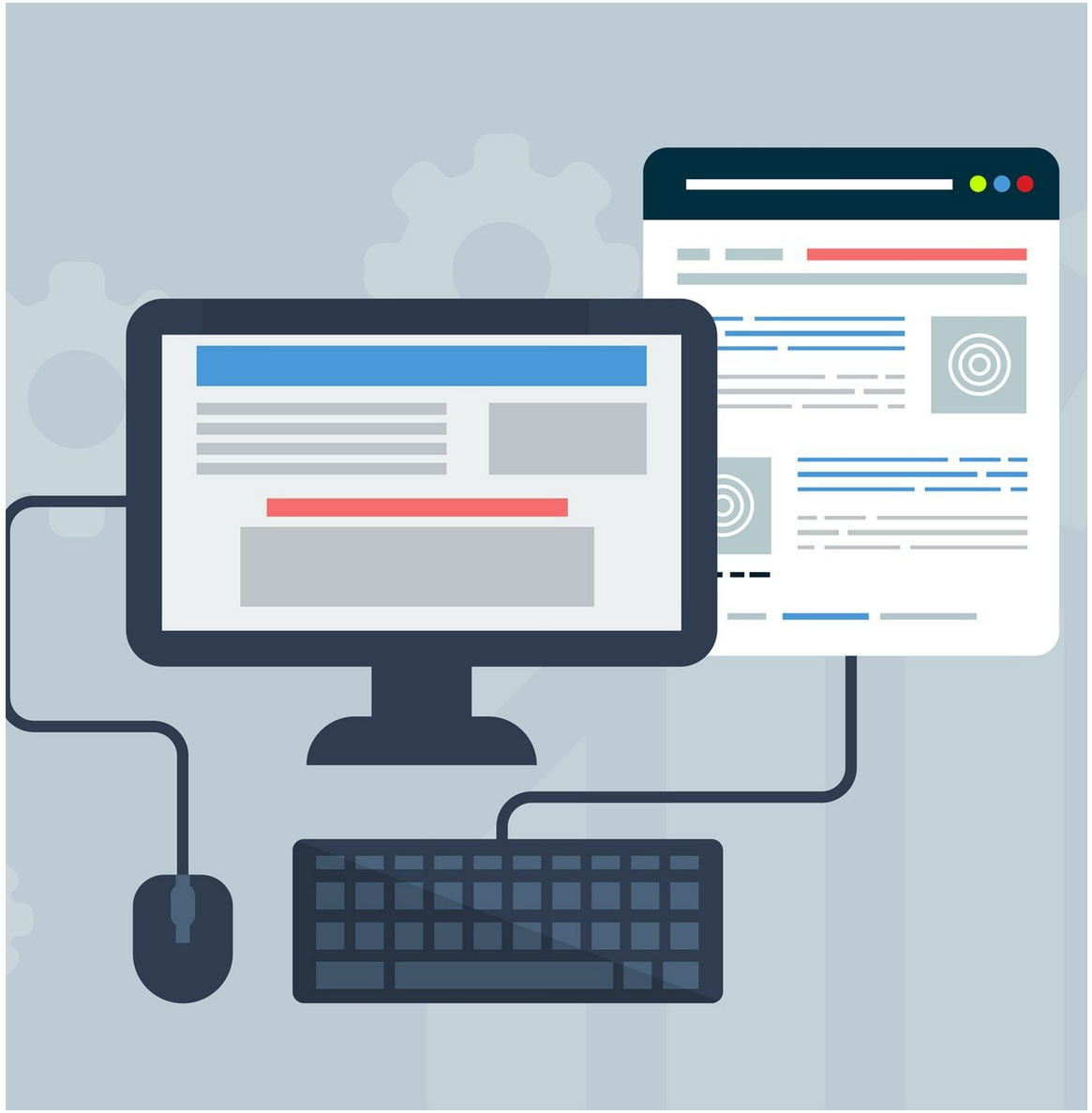
No visible layout

No WYSIWYG



STRUCTURE

If you want the computer to "use" the text data for some task, you need to **structure** it.



PATTERNS

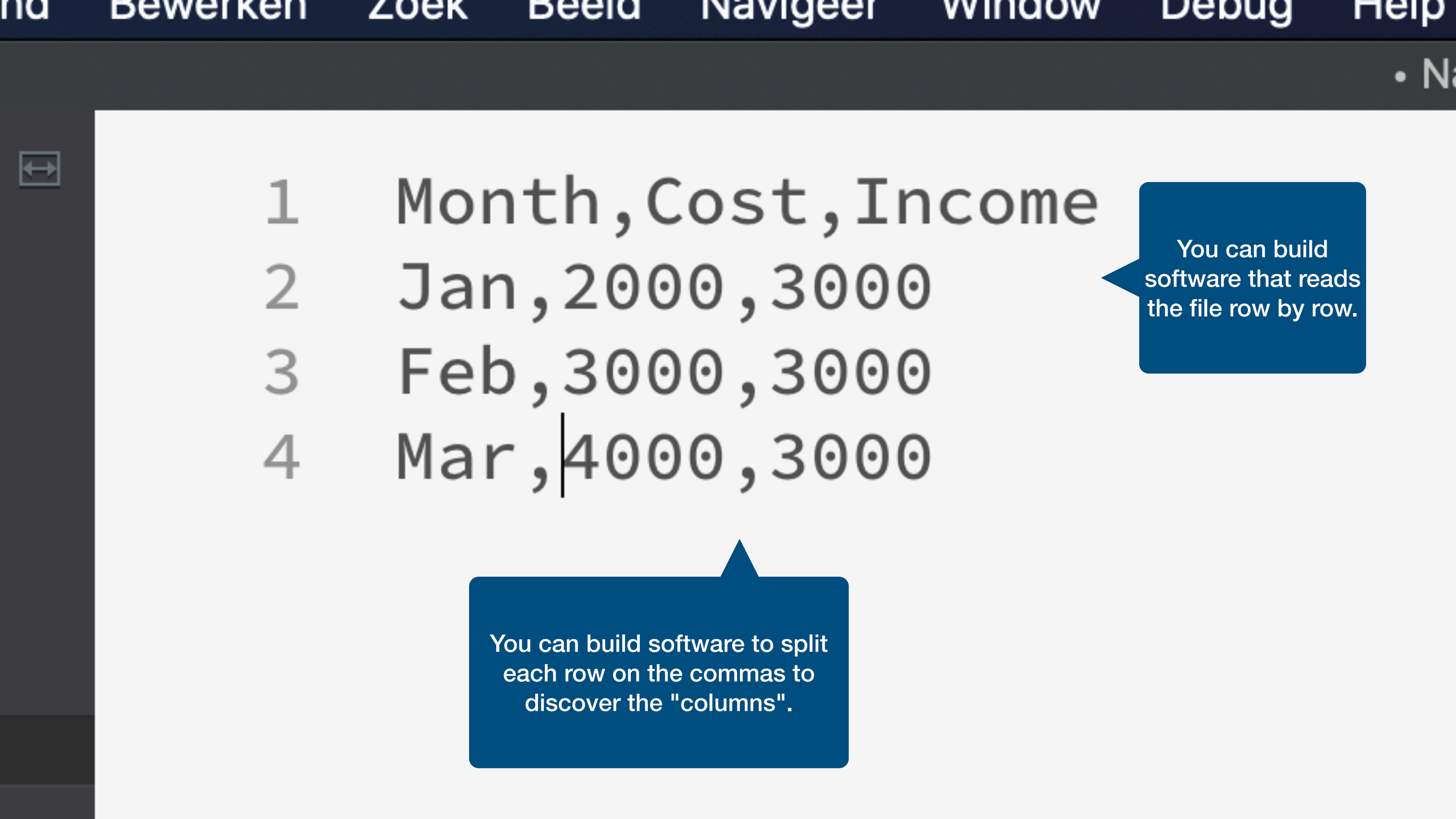
by applying structure,
the software can
recognize patterns

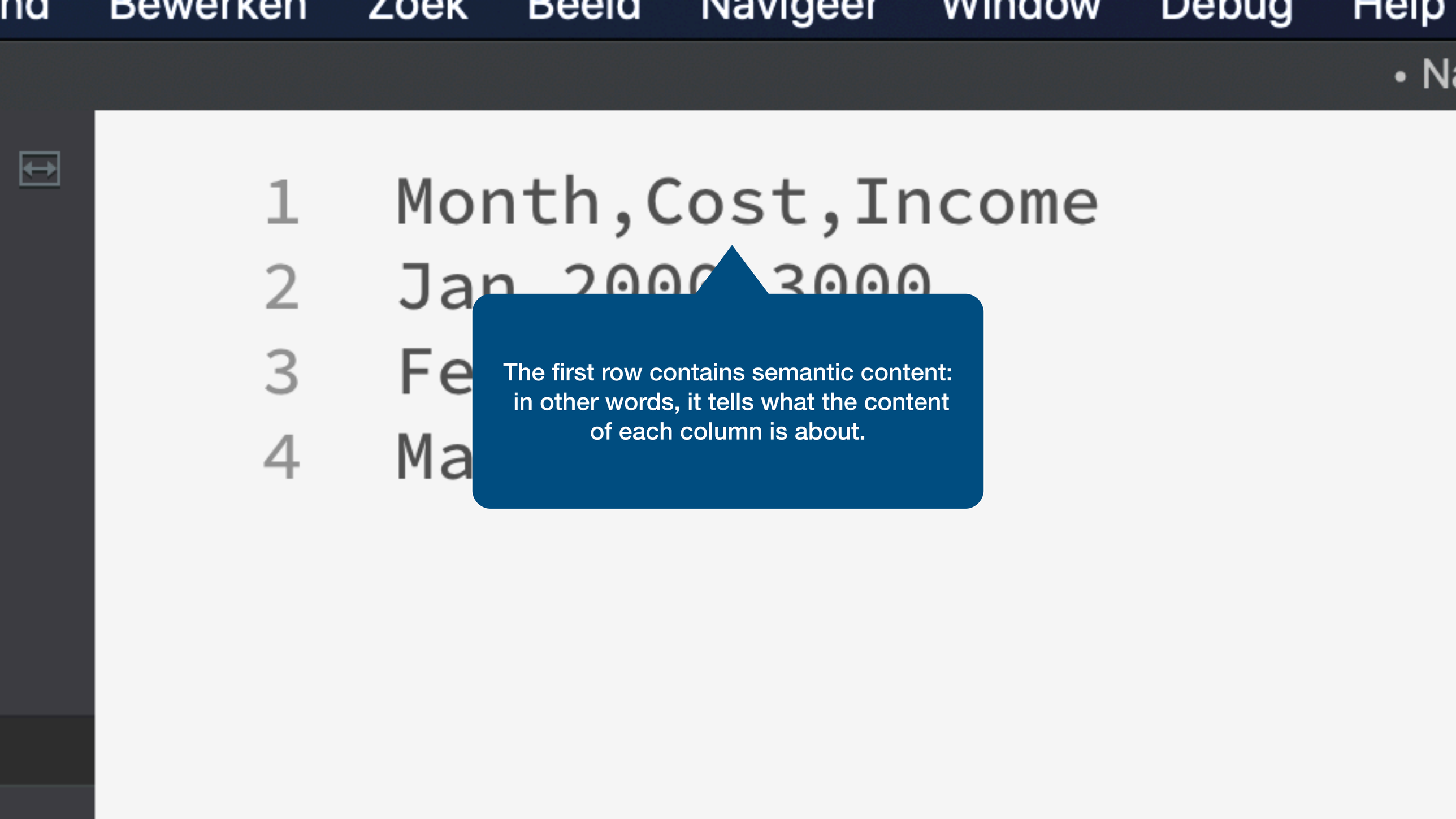


```
1   2000,3000
2   3000,3000
3   4000,3000
```

```
1   Month,Cost,Income
2   Jan,2000,3000
3   Feb,3000,3000
4   Mar,4000,3000
```

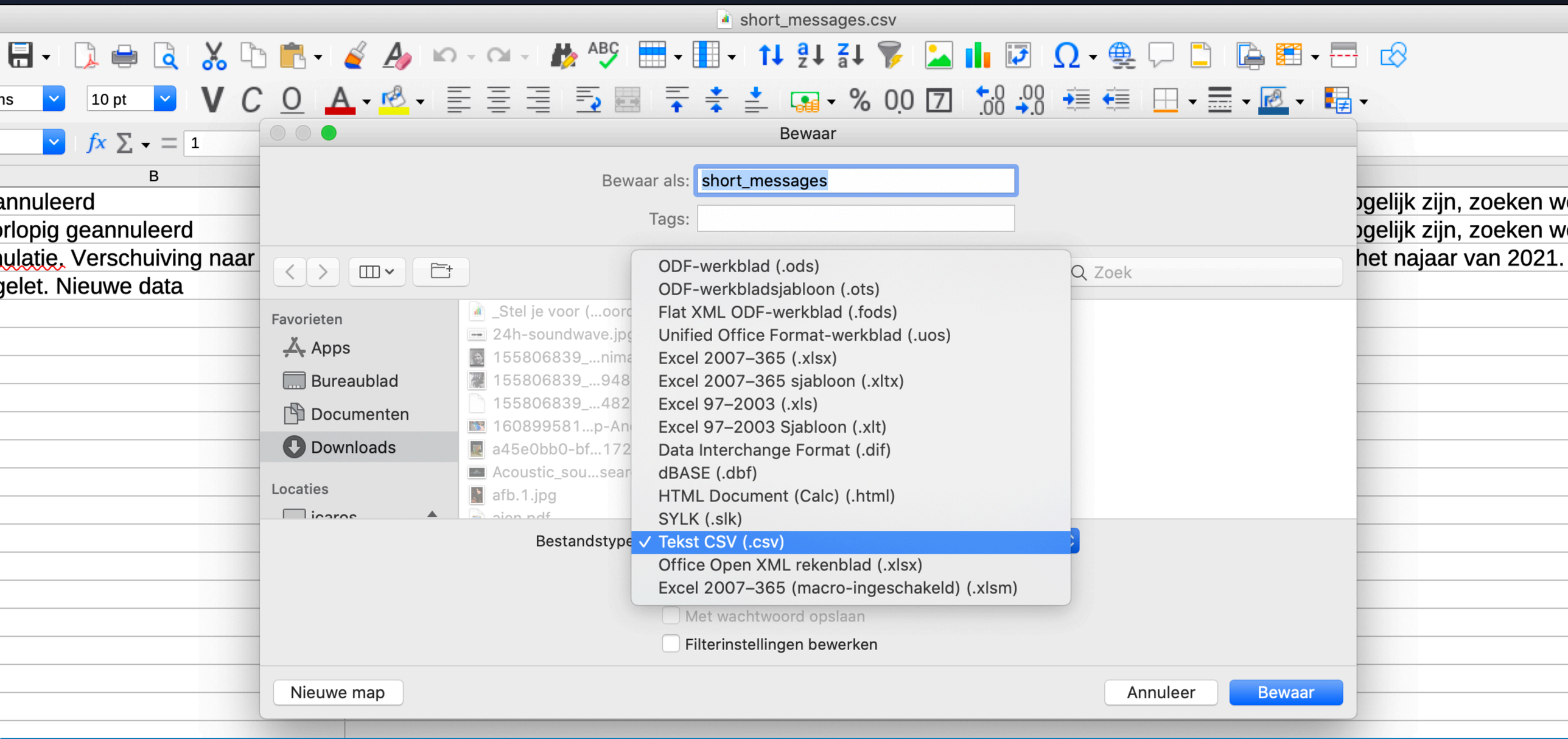

nd Bewerken Zoek Beeld Navigeer Window Debug Help
1 Month,Cost,Income
2 Jan,2000,3000
3 Feb,3000,3000
4 Mar,4000,3000
You can build software that reads the file row by row.


Bewerken Zoek Beeld Navigeer Window Debug Help
1 Month,Cost,Income
2 Jan,2000,3000
3 Feb,3000,3000
4 Mar,4000,3000
You can build software that reads the file row by row.
You can build software to split each row on the commas to discover the "columns".


nd Bewerken Zoek Beeld Navigeer Window Debug Help
1 Month,Cost,Income
2 Jan 2000 3000
3 Fe
4 Ma
The first row contains semantic content: in other words, it tells what the content of each column is about.

```
1   Month,Cost,Income
2   Jan,2000,3000
3   Feb,3000,3000
4   Mar,4000,3000
```

COMMA SEPARATED VALUES
CSV



VERY IMPORTANT EXCHANGE FORMAT

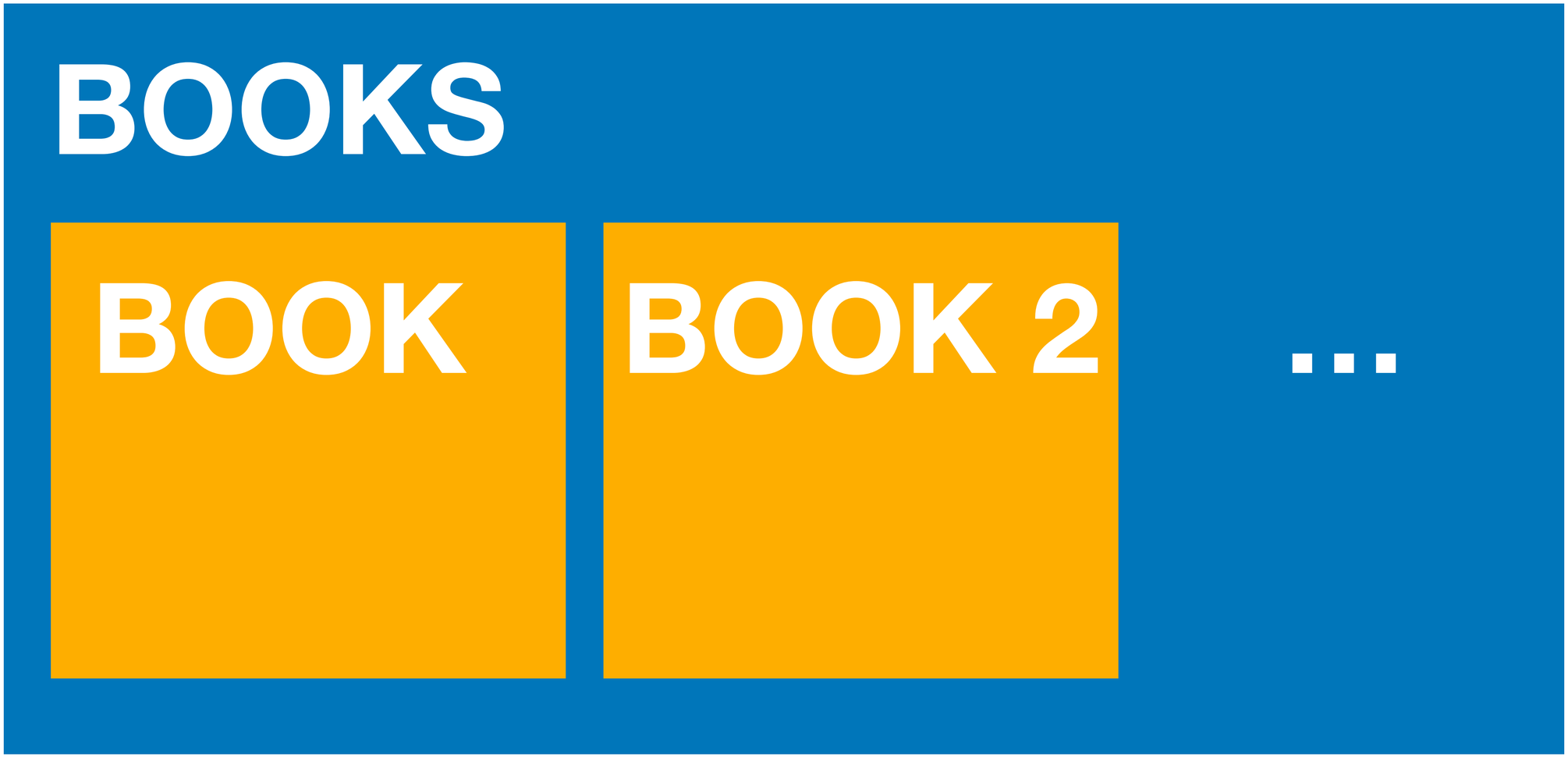
two-dimensional

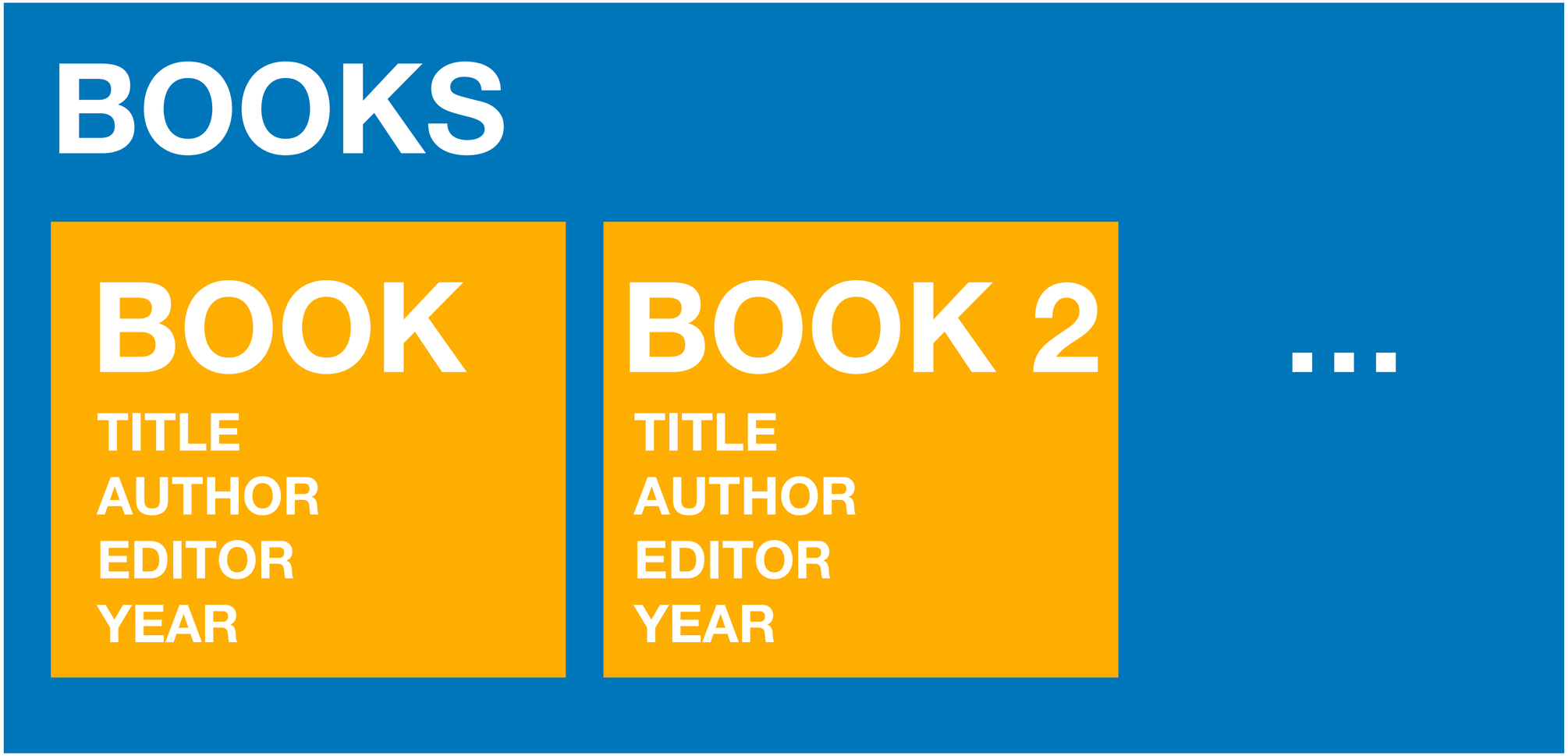
CSV-table

multi-dimensional or tree structure

XML, JSON

BOOKS


BOOKS
BOOK
BOOK 2
...


BOOKS
BOOK
TITLE
AUTHOR
EDITOR
YEAR
BOOK 2
TITLE
AUTHOR
EDITOR
YEAR
...

XML
Extended markup language

```xml
<books>
    <book>
        <title>On the origin of species</title>
        <author>Charles Darwin</author>
        <editor>Allan Lane</editor>
        <year>2014</year>
    </book>
    <book>
        <title>Dream Machines</title>
        <author>Ted Nelson</author>
        <year>1974</year>
    </book>
</books>
```

XML
Extended markup language



<book>value</book>

New lines and tabs are not important.
They only serve to **keep**
the content readable for human users.

JSON
Javascript Object Notation

```
[
    {
        "title": "On the origin of species",
        "author": "Charles Darwin",
        "editor": "Allan Lane",
        "year": "2014"
    },
    {
        "title": "Dream Machines",
        "author": "Ted Nelson",
        "year": "1974"
    }
]
```



"title": "Dream Machines"

JSON
key value pairs

New lines and tabs are not important.
They only serve to **keep**
the content readable for human users.

XML and JSON
Multidimensional

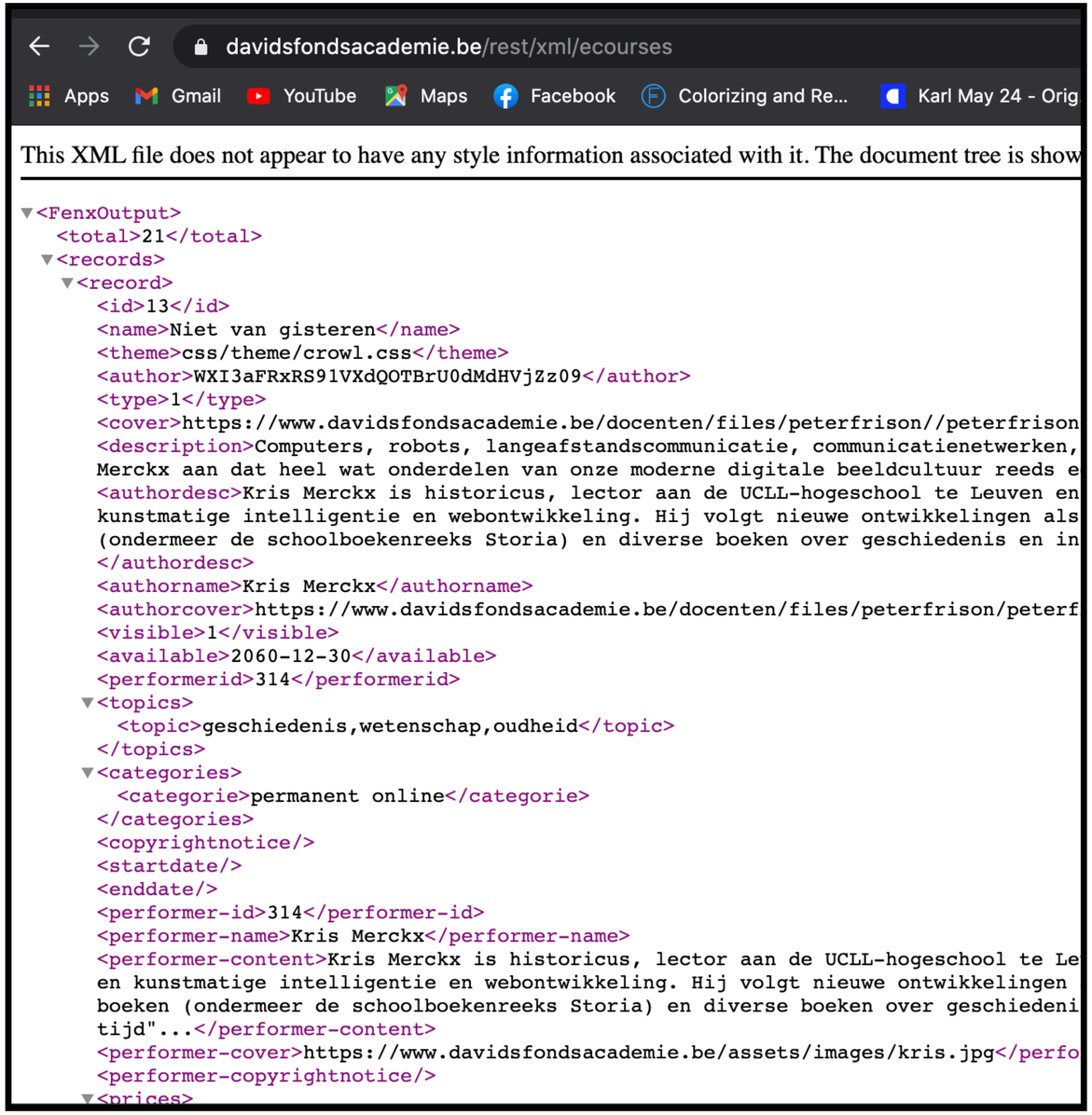
```xml
<books>
    <book>
        <title>On the origin of species</title>
        <author>Charles Darwin</author>
        <editor>Allan Lane</editor>
        <year>2014</year>
    </book>
    <book>
        <title>Dream Machines</title>
        <author>Ted Nelson</author>
        <year>1974</year>
        <identifier>
            <isbn>0-89347-002-3</isbn>
            <isbn-13>0-89347-002-3</isbn-13>
            <oclc>217227165</oclc>
        </identifier>
    </book>
</books>
```

```
[
    {
        "title": "On the origin of species",
        "author": "Charles Darwin",
        "editor": "Allan Lane",
        "year": "2014"
    },
    {
        "title": "Dream Machines",
        "author": "Ted Nelson",
        "year": "1974",
        "identifier": {
            "isbn": "0-89347-002-3",
            "isbn-13": "0-89347-002-3",
            "oclc": "217227165"
        }
    }
]
```

```
[
    {
        "title": "On the origin of species",
        "author": "Charles Darwin",
        "publisher": {
            "name": "Davisfonds",
            "street": "Quinten Metsysplein",
            "postalcode": "3000",
            "city": "Leuven",
            "tel": "+3216234567",
            "countrycode": "BE"
        };
        "year": "2014"
    },
```

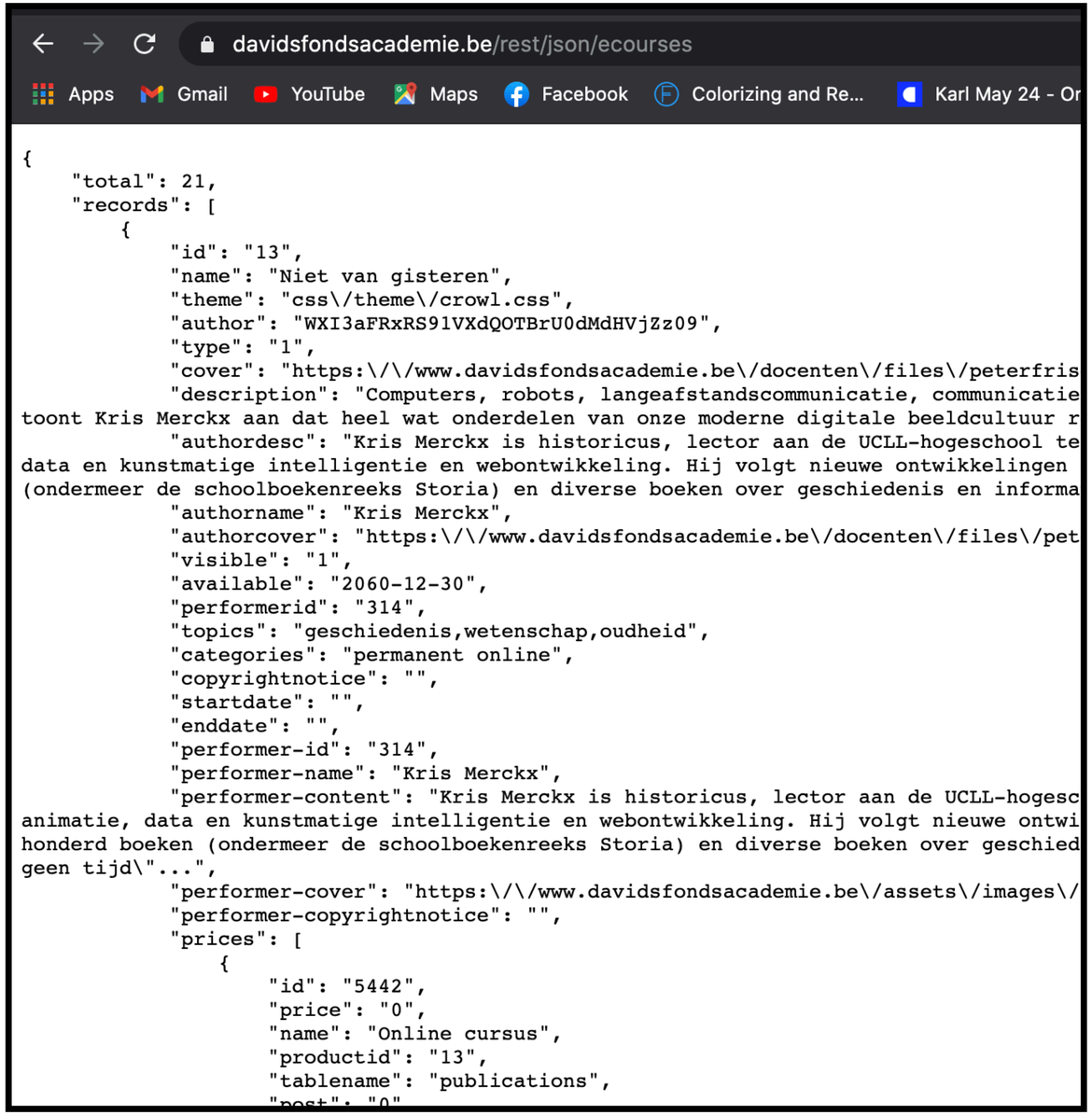
Example

XML OUTPUT



Example

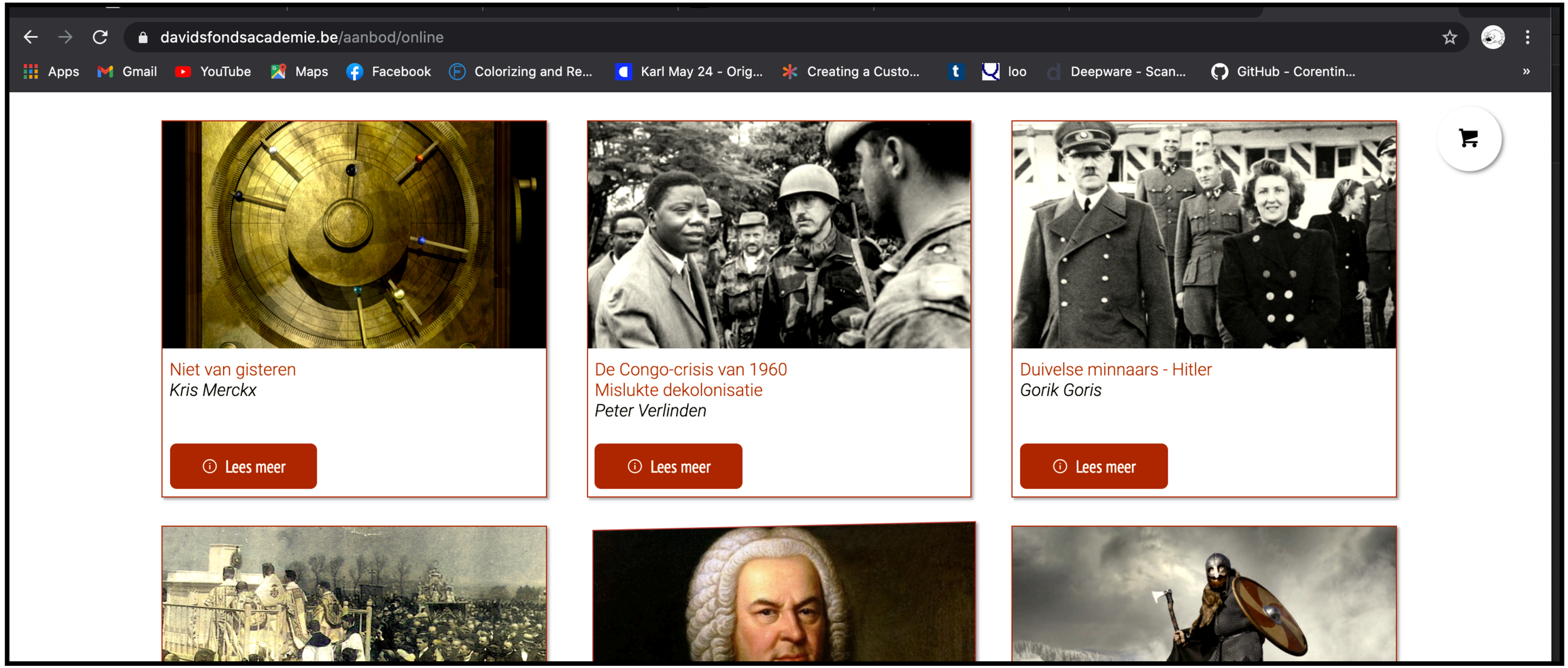
JSON OUTPUT




```
{
    "total": 21,
    "records": [
        {
            "id": "13",
            "name": "Niet van gisteren",
            "theme": "css\/theme\/crowl.css",
            "author": "WXI3aFRxRS91VXdQOTBrU0dMdHVjZz09",
            "type": "1",
            "cover": "https:\/\/www.davidsfondsacademie.be\/docenten\/files\/peterfris
            "description": "Computers, robots, langeafstandscommunicatie, communicatie
toont Kris Merckx aan dat heel wat onderdelen van onze moderne digitale beeldcultuur r
            "authordesc": "Kris Merckx is historicus, lector aan de UCLL-hogeschool te
data en kunstmatige intelligentie en webontwikkeling. Hij volgt nieuwe ontwikkelingen
(ondermeer de schoolboekenreeks Storia) en diverse boeken over geschiedenis en informa
            "authorname": "Kris Merckx",
            "authorcover": "https:\/\/www.davidsfondsacademie.be\/docenten\/files\/pet
            "visible": "1",
            "available": "2060-12-30",
            "performerid": "314",
            "topics": "geschiedenis,wetenschap,oudheid",
            "categories": "permanent online",
            "copyrightnotice": "",
            "startdate": "",
            "enddate": "",
            "performer-id": "314",
            "performer-name": "Kris Merckx",
            "performer-content": "Kris Merckx is historicus, lector aan de UCLL-hogesc
animatie, data en kunstmatige intelligentie en webontwikkeling. Hij volgt nieuwe ontwi
honderd boeken (ondermeer de schoolboekenreeks Storia) en diverse boeken over geschied
geen tijd\"...",
            "performer-cover": "https:\/\/www.davidsfondsacademie.be\/assets\/images\/
            "performer-copyrightnotice": "",
            "prices": [
                {
                    "id": "5442",
                    "price": "0",
                    "name": "Online cursus",
                    "productid": "13",
                    "tablename": "publications",
                    "post": "0"
```

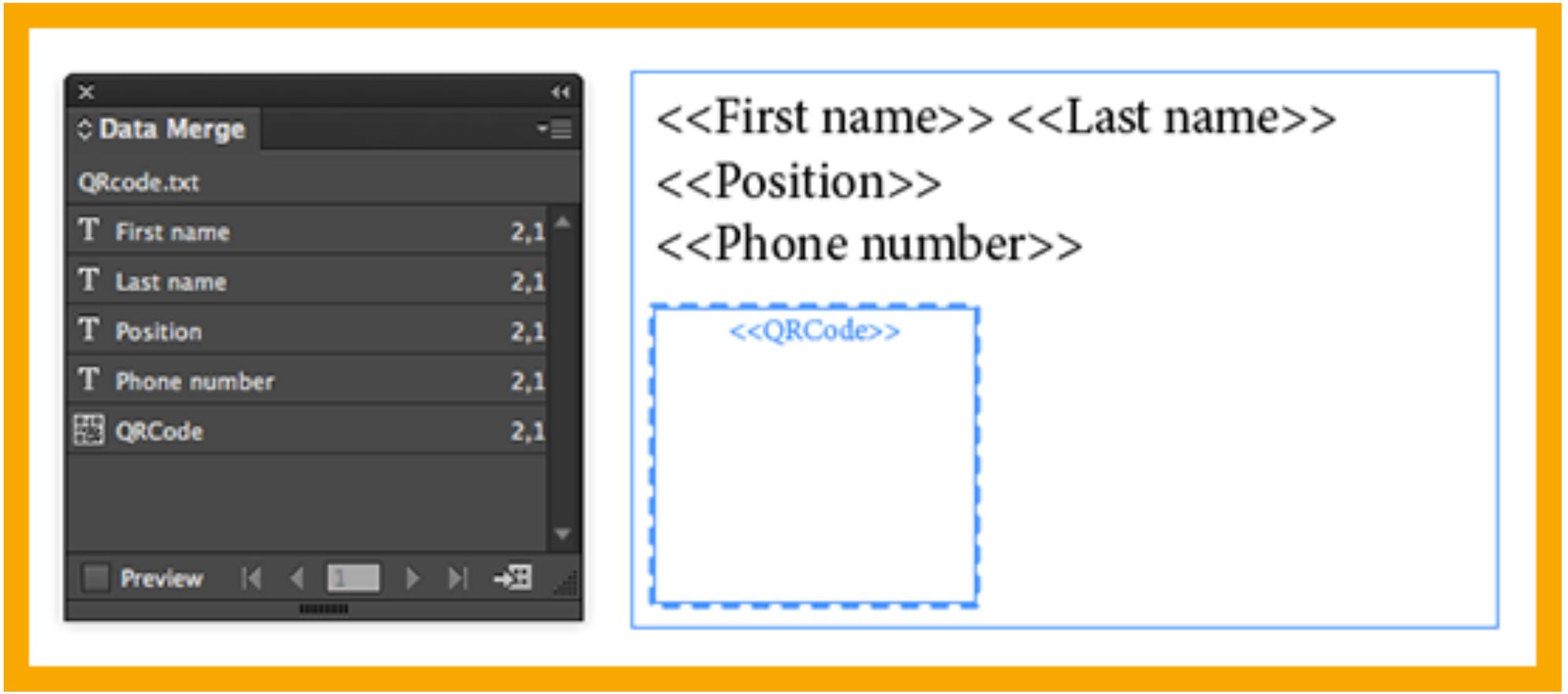
Example

“VIEW” FOR THE END USER

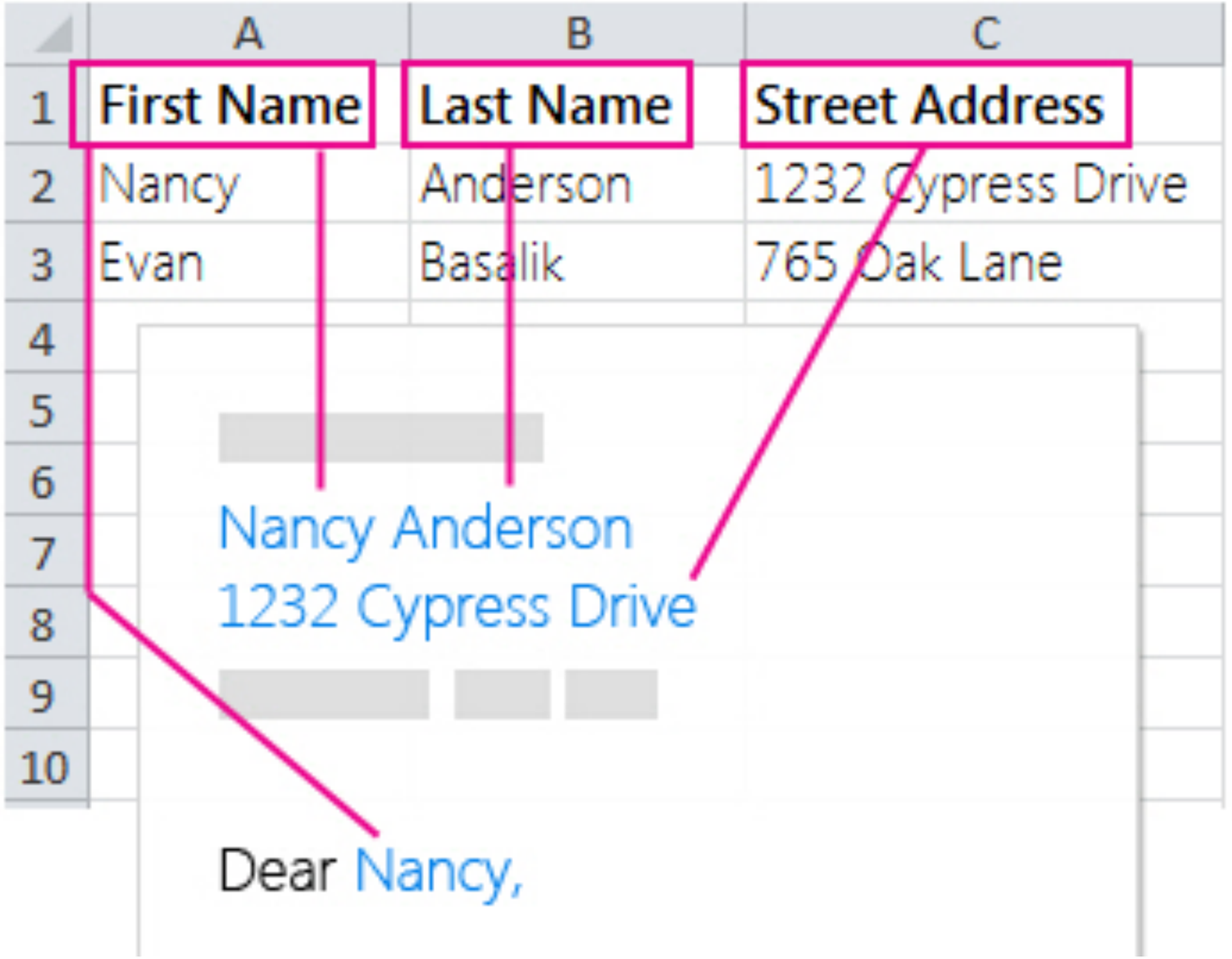



JSON DATA SOURCE
LAYOUT TEMPLATE
“PIECE OF SOFTWARE”
VIEW

Adobe Indesign merging data from a CSV file

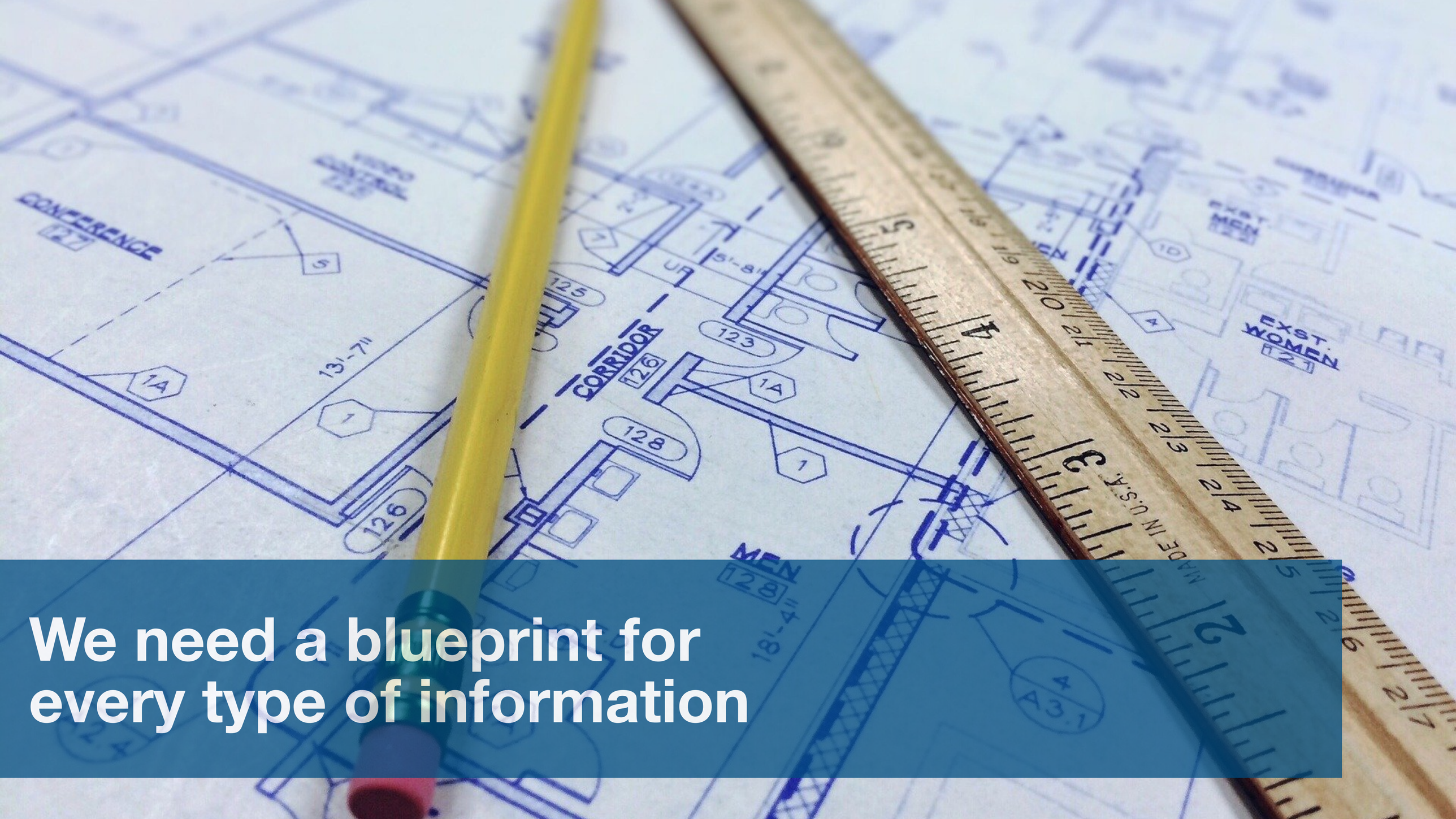


MS Word, mail merge with XLS-data



Data exchange

the need for schemes and (international) agreements



We need a blueprint for every type of information

schema.org

JSON-LD, RDF, MICROTAGS...

JSON LD

JavaScript Object Notation for Linked Data

An example of JSON LD in the source code of a website

```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "ClothingStore",
  "name": "Uit de Naad",
  "url": "https://www.uitdenaad.be",
  "logo": "https://www.uitdenaad.be/img/udnlogo.jpg",
  "image": "https://www.uitdenaad.be/img/uitdenaad_uniek.jpg",
  "address": {
    "@type": "PostalAddress",
    "streetAddress": "LEO DARTELAAN 9",
    "addressLocality": "LEUVEN",
    "postalCode": "3001",
    "addressCountry": "België"
  },
  "priceRange":"40-5000",
  "contactPoint": [{
    "@type": "ContactPoint",
    "telephone": "+32496419646",
    "contactType": "customer service"
  }],
  "telephone": "+32496419646"
}
</script>
```

Google understands what the content of the page is about thanks to JSON LD

naad.be

Naad

waren op maat maken. Ervaren ledermaker **Uit de Naad** maakt en herstelt
n, riemen, accessoires, schorten, op maat. Regio Leuven, Tienen ...

esgeven naar lederwaren

ant: HLN. Handtassen op
nieuwe pop-up ...

Maatwerk

Maatwerk. Ik maak handtassen op
maat, één voor één, un par un ...

tas ambachtelijke ...

oud om te 'leren'... Bent u
naar die lederwaren ...

Herstellingen van lederwaren

Herstellingen van lederwaren
waarvoor je bij de schoenmaker ...

sen

handgemaakte
sen. Made in Belgium ...

Atelier ambachtelijke ...

Nooit te oud om te 'leren'... Bent u
op zoek naar die lederwaren ...

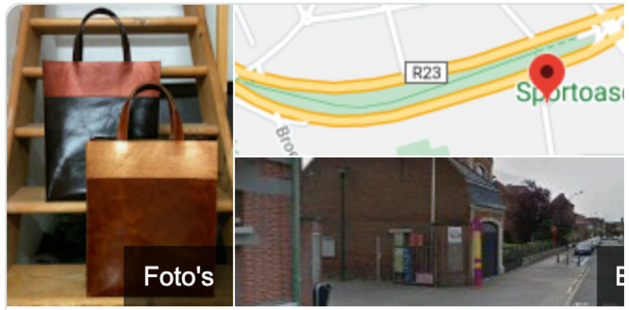
sultaten van uitdenaad.be »

gisch.be › belgische-merken › uit-de-naad

Naad - Ik Koop Belgisch

ad ontwerpt, (her)maakt en herstelt lederwaren. Hiervoor ontving **Uit de Naad** het
n erkend ambacht. Om duurzame redenen legt **Uit de Naad** ...

book.com › uitdenaad Vertaal deze pagina



Uit de naad & Ardeco

Website | Route | Opslaan | Bellen

5,0 ★★★★★ 3 Google-reviews

Kunstorganisatie in Leuven

Adres: Leo Dartelaan 9, 3001 Leuven

Je bent hier geweest in september 2020

Telefoon: 0496 41 96 46

Bewerking voorstellen · Ben je eigenaar van dit be

Ontbrekende gegevens toevoegen

Openingstijden toevoegen

Ken je deze plaats? Deel de meest recente inform

Assignment
XML & JSON

- You now make an assignment in a small group or bubble (2 to 4 students)
- Think of a pattern / structure for a webshop of music. The website that will use your data source wants to be able to display information about the music, the "album", the artist and the record company.
- Your list must contain at least 3 items from real artists / songs / albums.
- Think of your own structure. Do not use a blueprint from schema.org
- Create an XML and a JSON version.
- Upload the assignment on Toledo before the end of the lesson.

