

Data

Data

1. Ooh data... kennen we al
2. Data in het wild - formats for data transfer

Data

1. **Ooh data... kennen we al**
2. Data in het wild - formats for data transfer

Data - waarom data?

Wij verzamelen data, slaan data op, beheren data...

Omdat zij iets zegt over de **wereld** waarin wij leven



Data - kennen wij al?

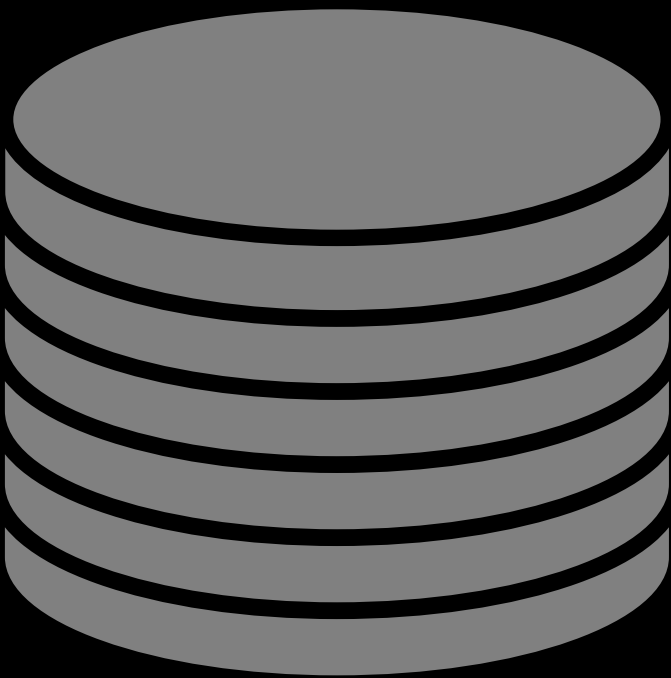
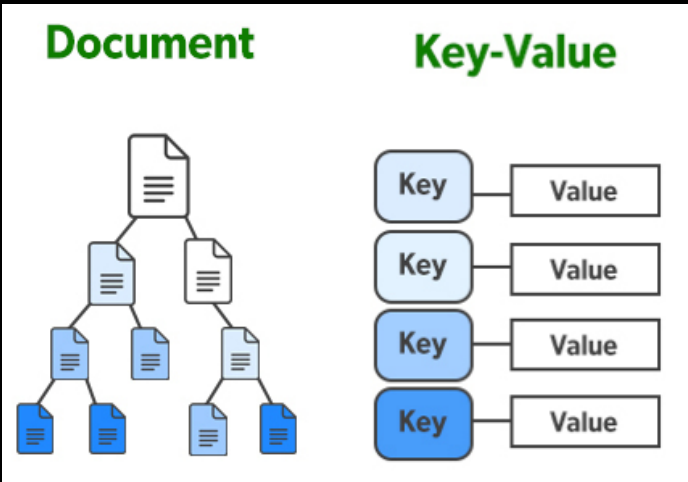
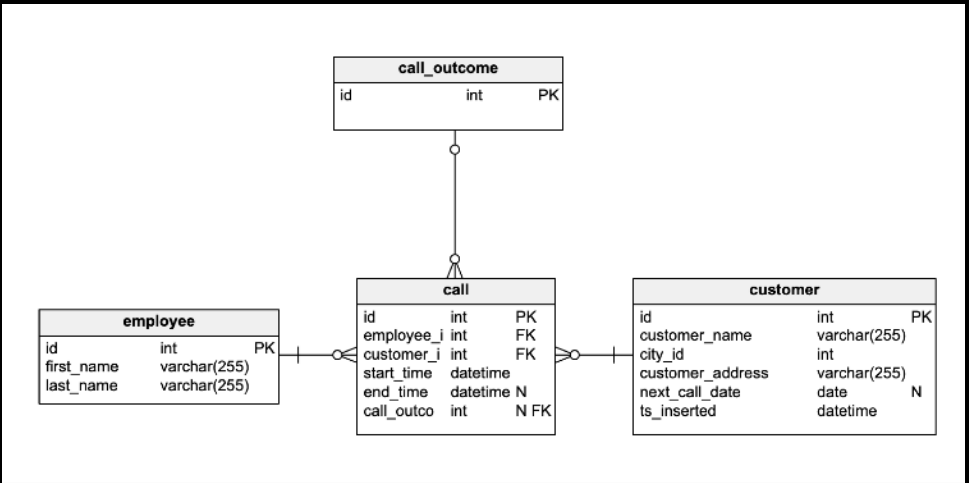
Kleine dataverzamelingen

Middelgrote dataverzamelingen

Grote dataverzamelingen

	A	B	C	D	E	F	G
1	Donation ID	Donor ID	Type	Method	Status	Amount	Date
2	D153	S067	Donation	Credit card	Completed	\$100,00	02.05
3	D154	S123	Shirt	Credit card	Abandoned	\$25,00	03.05
4	D155	S345	Shirt	Paypal	Completed	\$25,00	07.05
5	D156	S367	Donation	Cash	Completed	\$50,00	07.05
6	D157	S121	Shirt	Paypal	Failed	\$25,00	07.05
7	D158	S112	Donation	Credit card	Completed	\$75,00	10.05
8	D159	S055	Donation	Credit card	Completed	\$100,00	12.05
9	D160	S089	Donation	Paypal	Completed	\$60,00	13.05
10	D161	S523	Shirt	Credit card	Failed	\$25,00	15.05
11	D162	S123	Shirt	Cash	Completed	\$25,00	17.05
12	D163	S165	Donation	Paypal	Abandoned	\$35,00	17.05
13	D164	S094	Donation	Paypal	Completed	\$75,00	21.05
14	D165	S054	Donation	Credit card	Completed	\$50,00	22.05
15	D166	S031	Shirt	Credit card	Completed	\$25,00	24.05
16	D167	S451	Donation	Cash	Completed	\$100,00	24.05
17	D168	S112	Shirt	Credit card	Abandoned	\$25,00	27.05
18	D169	S002	Donation	Paypal	Failed	\$50,00	27.05
19	D170	S112	Shirt	Credit card	Completed	\$25,00	30.05

```
1 {
2   "type": "FeatureCollection",
3   "name": "HISTORISCHE_BEBOUWING",
4   "features": [
5     {
6       "id": 3,
7       "type": "Feature",
8       "geometry": {
9         "type": "Point",
10        "coordinates": [
11          4.897985,
12          52.3743134
13        ]
14      },
15      "properties": {
16        "Naam": "Oude Kerk 1306",
17        "Hoofdgroep": "Religie",
18        "Functie": "Kerk",
19        "Foto": "Oude kerk.jpg",
20        "Website": "",
21        "Opmerking": "Bron: Cornelis Anthonisz 1544",
22        "Begin": 1306,
23        "Eind": 0,
24        "SELECTIE": "RELIGIE"
25      }
26    },
27    {
28      "id": 4,
29      "type": "Feature",
30      "geometry": {
31        "type": "Point",
32        "coordinates": [
33          4.8967309,
34          52.3770523
35        ]
36      },
37      "properties": {
38        "Naam": "Windmolenzijdepoorthuis 1300-1380",
39        "Hoofdgroep": "Militair",
40        "Functie": "Poort",
41        "Foto": "",
42        "Website": "",
43        "Opmerking": "",
44        "Begin": 1300,
45        "Eind": 1380,
46        "SELECTIE": "MILITAIR"
47      }
48    },
49    {
50      "id": 5,
51      "type": "Feature",
52      "geometry": {
53        "type": "Point",
54        "coordinates": [
55          4.8946872,
56          52.3789139
57        ]
58      },
59      "properties": {
60        "Naam": "Haarlemmerpoort 1425-1579",
61        "Hoofdgroep": "Militair",
62        "Functie": "Poort",
63        "Foto": "Haarlemmerpoort.jpg",
64        "Website": "",
65        "Opmerking": "Bron: Cornelis Anthonisz 1544",
66        "Begin": 1425,
67        "Eind": 1579.
```



Wat is het verschil?

Kleine dataverzamelingen (Excell):

+ lokaal, overzichtelijk

- 'foto', data vaak niet up-to-date

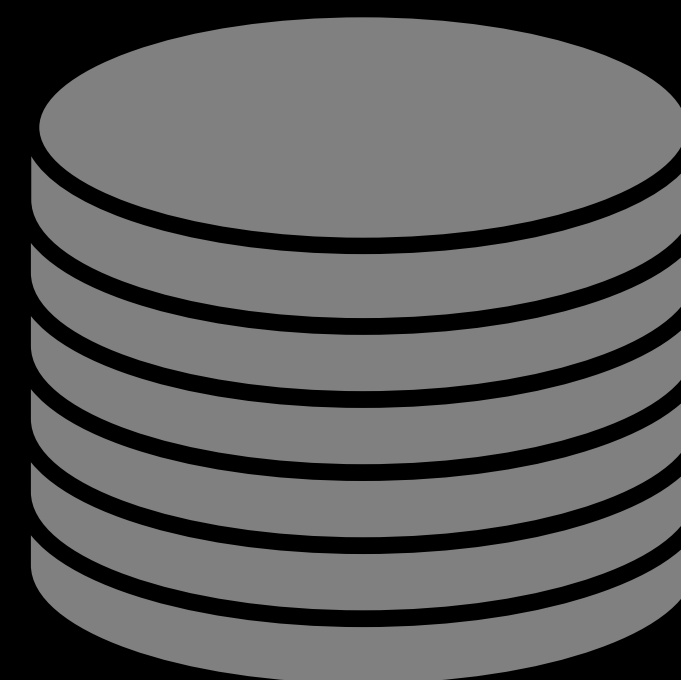
	A	B	C	D	E	F	G
1	Donation ID	Donor ID	Type	Method	Status	Amount	Date
2	D153	S067	Donation	Credit card	Completed	\$100,00	02.05
3	D154	S123	Shirt	Credit card	Abandoned	\$25,00	03.05
4	D155	S345	Shirt	Paypal	Completed	\$25,00	07.05
5	D156	S367	Donation	Cash	Completed	\$50,00	07.05
6	D157	S121	Shirt	Paypal	Failed	\$25,00	07.05
7	D158	S112	Donation	Credit card	Completed	\$75,00	10.05
8	D159	S055	Donation	Credit card	Completed	\$100,00	12.05
9	D160	S089	Donation	Paypal	Completed	\$60,00	13.05
10	D161	S523	Shirt	Credit card	Failed	\$25,00	15.05
11	D162	S123	Shirt	Cash	Completed	\$25,00	17.05
12	D163	S165	Donation	Paypal	Abandoned	\$35,00	17.05
13	D164	S094	Donation	Paypal	Completed	\$75,00	21.05
14	D165	S054	Donation	Credit card	Completed	\$50,00	22.05
15	D166	S031	Shirt	Credit card	Completed	\$25,00	24.05
16	D167	S451	Donation	Cash	Completed	\$100,00	24.05
17	D168	S112	Shirt	Credit card	Abandoned	\$25,00	27.05
18	D169	S002	Donation	Paypal	Failed	\$50,00	27.05
19	D170	S112	Shirt	Credit card	Completed	\$25,00	30.05

Database

Remote, dynamisch

+ altijd up-to-date

- oncontroleerbaar (oneindig) groot



Interessante
datavisualisaties

Hoe groot is “groot”...

Banktransacties in Nederland

18.000.000 bankrekeningen x

10 transacties/dag

180.000.000 transacties/dag in NL

1 / 10.000 foutpercentage

18.000 fouten/dag in NL

Probleem: de omvang

Grootste zorg in database engineering

Betrouwbaarheid van data garanderen

De code achter een database moet “goed” zijn



Ooh code... kennen toch wij ook al?

CMD'ers maken User Interfaces

Wij kunnen “zien” wat de code doet

Er is ook code waarvan je pas kan “zien” wat hij doet, als er een groot probleem is

- Medische software: Eurotransplant
- Beslissings Ondersteunende Systemen: Deltaprogramma
- Software die vliegtuigen bestuurt ...

Reliability by design

Hoe doe je dat dan?

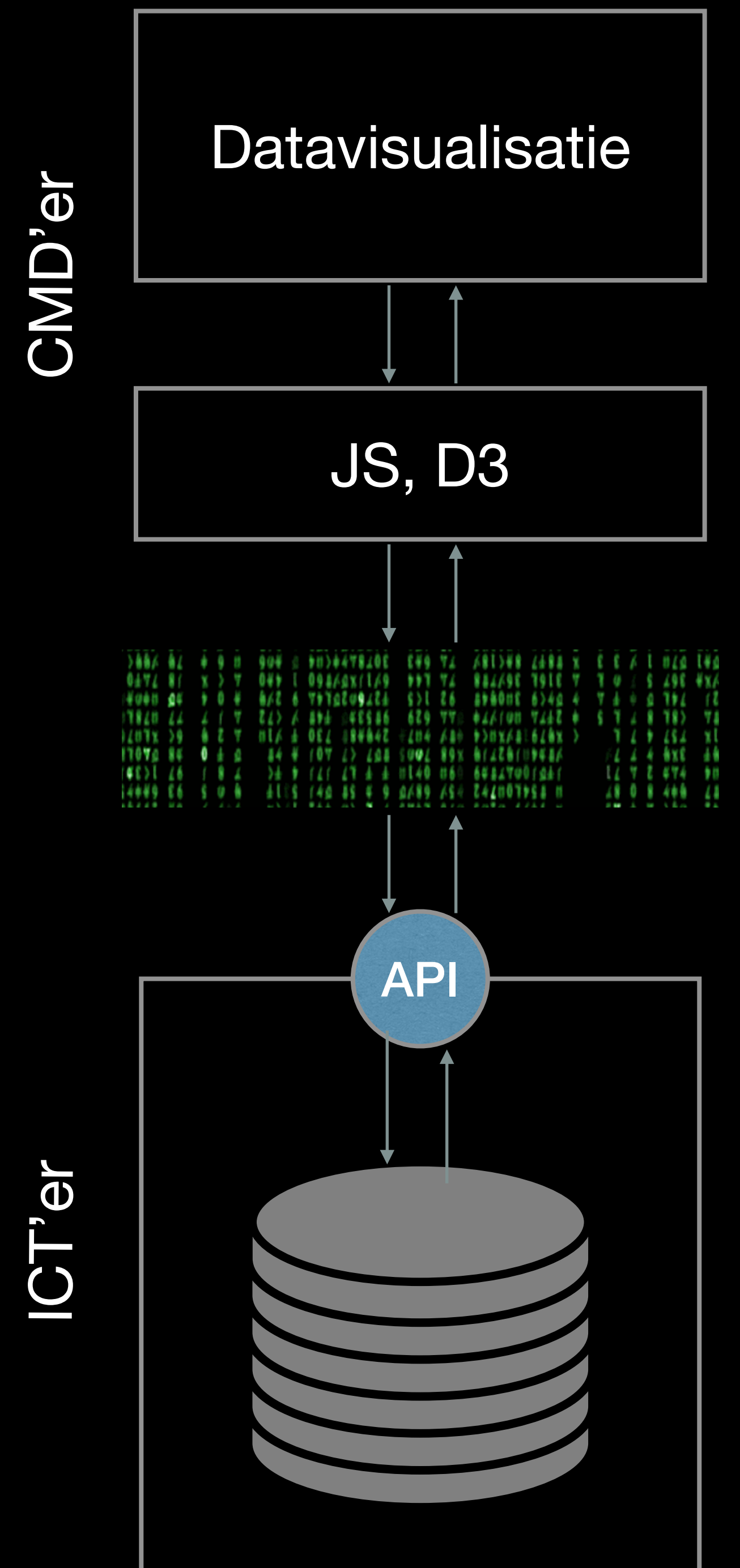
Dat doen wij niet bij CMD

Reliability by design is het vakgebied van de ICT

Zij schrijven API's

Application Programming Interfaces

Die worden gebruikt door collega's als CMD'ers



So what?

Betrouwbaarheid van data is een punt van aandacht

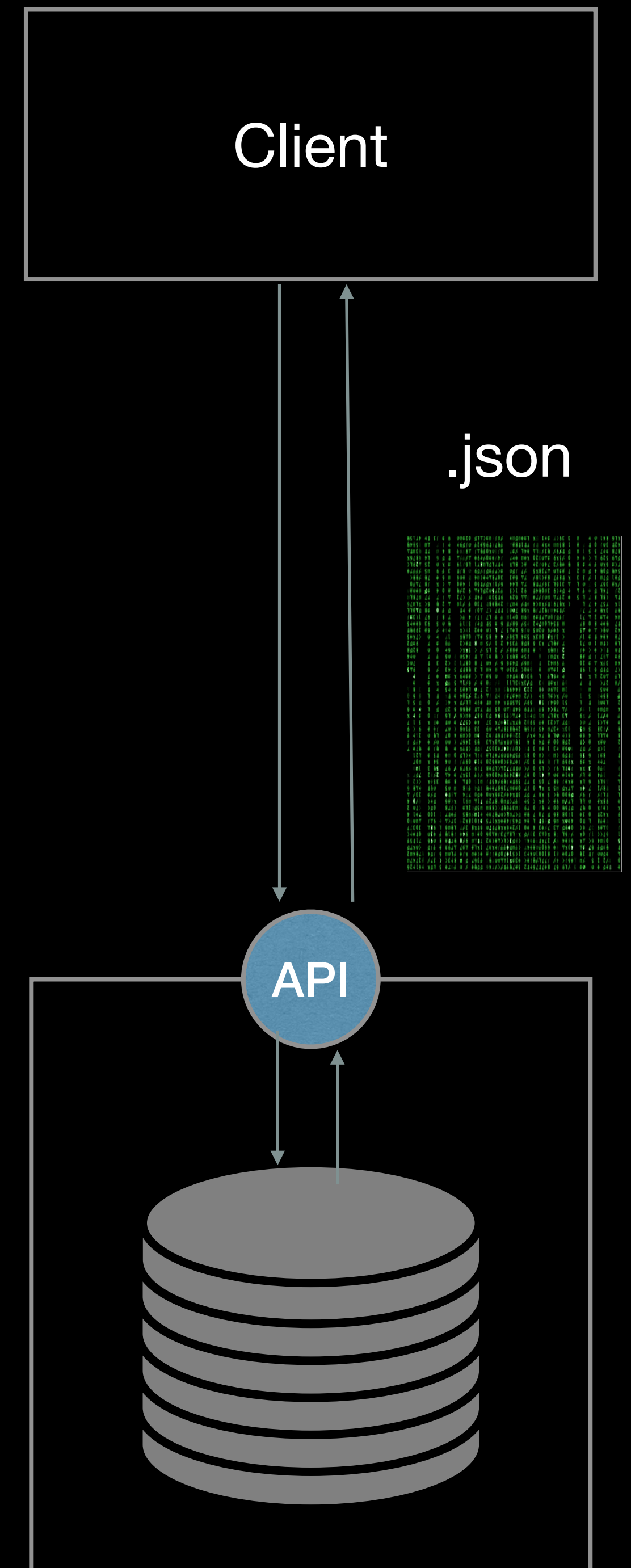
Benoem altijd de **bron** van je data

Beargumenteer, in hoeverre je denkt dat data **betrouwbaar** is

Gebruik een **API** voor data die up-to-date is

Schrijf leesbare code

Check je verhaal



Check je verhaal

ICT'ers beheren de API

CMD'ers ontwerpen en maken de **User Interface**

Data Scientists kunnen je helpen om het verhaal achter de data te checken



CMD'er

Data scientist

ICT'er



So what?

Wees kritisch

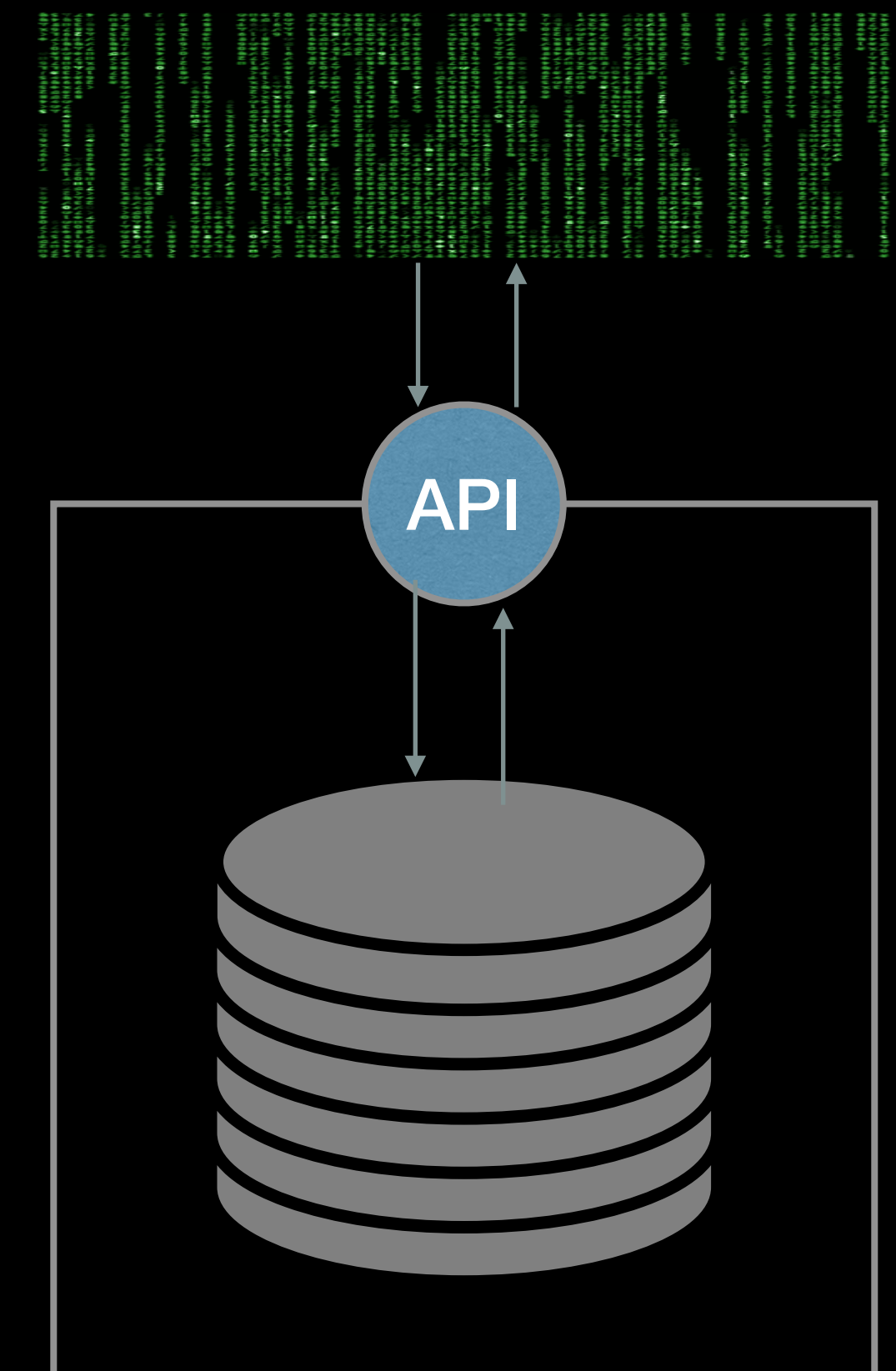
Streef naar een kloppend verhaal (kloppend met wat je weet over de werkelijkheid)

Check, dubbelcheck: **Trianguleer**

Check, dubbelcheck: raadpleeg evt. een specialist

Of een **Data Scientist**

Trek daar tijd voor uit!



Data

1. Ooh data... kennen we al
2. **Data in het wild - formats for data transfer**

Data in het wild: formats for data transfer

CSV: Comma Separated Values

XML: eXtensible Markup Language

JSON: Java Script Object Notation

CSV

CSV: Comma Separated Values

Oudste format voor data overdracht (= data transfer)

“Studentnummer”, “Achternaam”, “Tussenvoegsel”, “Roepnaam”, “Telefoonnummer”, “Woonplaats”, “Postcode”, “Huisnummer”, “Diploma”, “Opleiding”

“50087953”, “Benvenuti”, “”, “Laura”, “06-21157151”, “Amsterdam”, “1091CM”, “2”, “Mediacollege Zaandam”, “CMD”

+ Lichtgewicht, snel

- alleen geschikt voor tabellen (niet voor complexe data), kan niet worden verwerkt met JS

XML

Familie van HTML, maar eXtensible - developer mag tags toevoegen

```
<publicatie>
```

```
  <artikel>Benvenuti, Laura, Erik Barendsen, Gerrit C. van der Veer, and Johan Versendaal.
```

```
  "Understanding Computing in a Hybrid World: On the Undergraduate Curriculum Front-End Development." In  
  Proceedings of the 49th ACM Technical Symposium on Computer Science Education, pp. 580-585. ACM,  
  2018.</artikel>
```

```
    <link>../docs/UnderstandingComputing.pdf</link>
```

```
</publicatie>
```

+ Geschikt voor complexe data; kan data valideren (betrouwbaarheid!)

```
pattern value="[0-9]{4}[][a-zA-Z]{2}"
```

- Kan traag zijn, kan niet worden verwerkt met JavaScript - converteren!

JSON

Java Script Object Notation

```
{ artikel: "Benvenuti L.,G.C.van der Veer: 'Multimedia design kan je leren, kunnen we het ook doceren?', In: F.J. Verbeek, D. Lenior and M. Steen (Eds.) Change! - Proceedings CHI-NL 2009. CHI-NL, Leiden, Netherlands, 39-41",
```

```
  link: "../docs/MultimediaDesignDoceren.pdf",
```

```
  hoofdstuk: "chapter 3"
```

```
}
```

- + “snel”, lichtgewicht, kan complexe data aan, verwerking met JavaScript
- kan data niet valideren

So what?

API's kunnen data teruggeven in verschillende formaten

Wij gebruiken JSON

Data, aangeleverd in andere formaten, moeten wij converteren

Er zijn converters online!



Vragen?