

Workshop: kunnen we de levenscyclus van bestandsformaten voorspellen?

KIA Thematiendaagse 2024

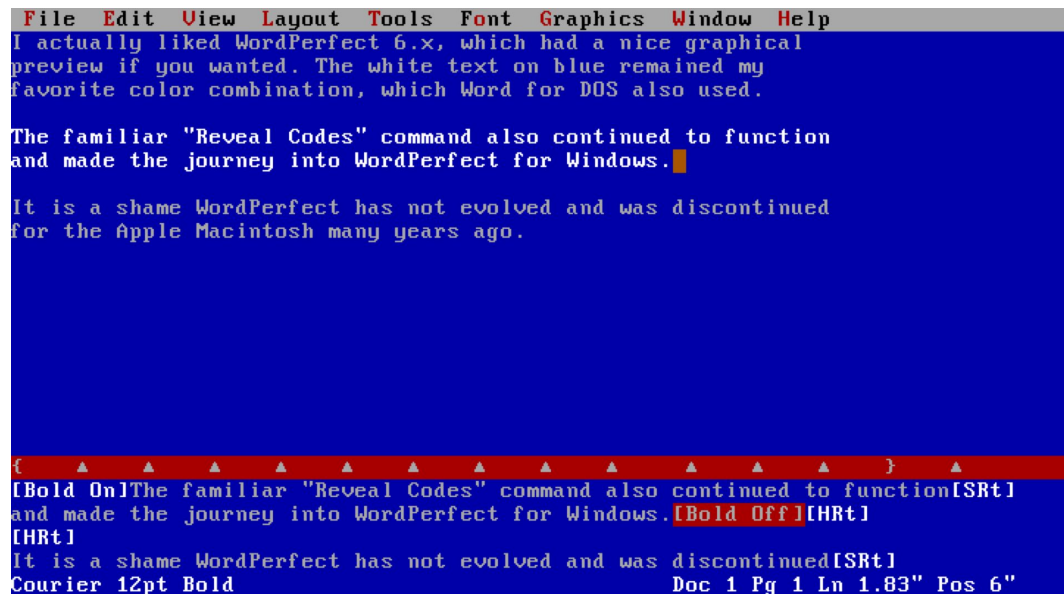
Sam Alloing
Valentijn Gilissen
Marin Rappard

Voorstelrondje

- Wie ben je?
- Wat is je functie?
- Waar werk je?
- Heb je data meegenomen?
- Wat hoop je vandaag te leren?

Analyse bestandsformaten, waarom?

Bestandsformaten komen en gaan (soms)



```
File Edit View Layout Tools Font Graphics Window Help
I actually liked WordPerfect 6.x, which had a nice graphical
preview if you wanted. The white text on blue remained my
favorite color combination, which Word for DOS also used.

The familiar "Reveal Codes" command also continued to function
and made the journey into WordPerfect for Windows.

It is a shame WordPerfect has not evolved and was discontinued
for the Apple Macintosh many years ago.

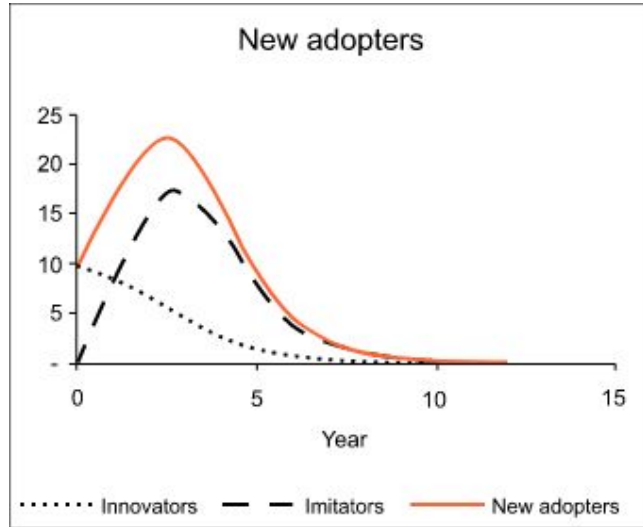
{
[Bold On]The familiar "Reveal Codes" command also continued to function[SRt]
and made the journey into WordPerfect for Windows.[Bold Off][HRt]
[HRt]
It is a shame WordPerfect has not evolved and was discontinued[SRt]
Courier 12pt Bold Doc 1 Pg 1 Ln 1.83" Pos 6"
```

Bron: <https://www.tameri.com/wordpress/blog/2010/06/09/i-sometimes-miss-worderfect-for-dos>

Context

- Bestandsformaten zijn niet eeuwig
 - Wanneer is een bestandsformaat “verouderd”?
- NDE onderzoek uit 2022
 - Onderdeel van de Preservation Watch groep van NDE
 - Praktische methode ontwikkelen om monitoring van bestandsformaten uit te voeren
 - Kan de monitoring geautomatiseerd worden?
 - Kan het gebruikt worden op instellingsniveau en op wereldwijd niveau?
 - Wat is de relatie tussen applicatie en bestandsformaat?

Methode



- Gebruikt in eerder onderzoek voor bestandsformaten
 - Kresimir Duretec en Christoph Becker, [Format Technology Lifecycle Analysis](#)
 - Webarchief data
- Bass Diffusion model (1969)
 - Frank Bass (1926 - 2006)
 - Economisch model over levenscyclus van producten
- Lineaire regressie als referentiemodel

Analyse van gebruiksfrequentie

De Bass-methode

Wat analyseren we hier?

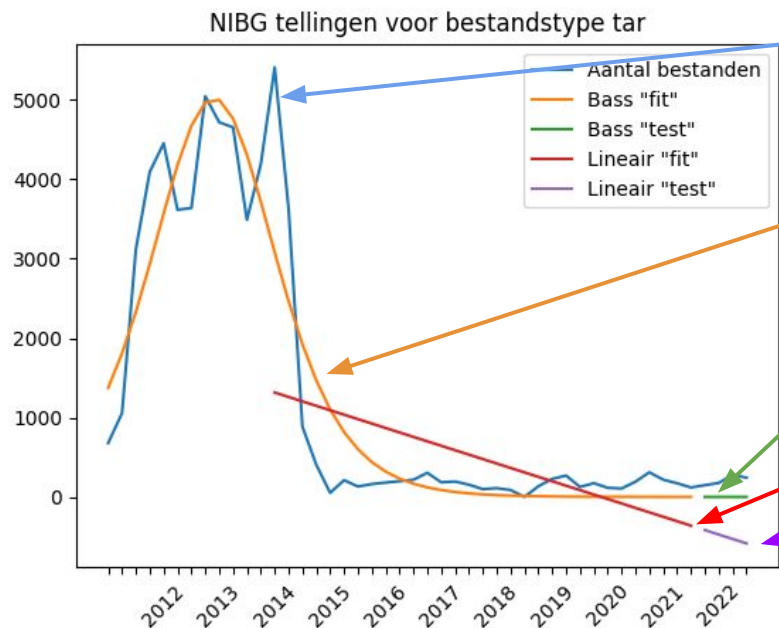
De aantallen bestanden voor “tar”,
gegroepeerd per kwartaal

De data waarop we de Bass curve hebben
toegepast

De voorspelling van de Bass curve

De data waarop we een eenvoudige rechte
trendlijn hebben toegepast

De voorspelling van de trendlijn



Bron:

<https://github.com/Anfield-Creations/NDE-monitoring-file-formats>

De praktijk!

Overzicht werkwijze

1. Verzamelen van gegevens
 - a. CSV bestand met creatiedatum bestand en bestandsformaat
2. Omzetten van de data in bruikbaar formaat
 - a. Input CSV omzetten naar JSON
 - b. Data schonen
 - c. Formaten selecteren
3. Analyse uitvoeren
4. Analyseren van de uitkomsten

Stappen 1,2 en 4 zijn organisatie specifiek



Stap 1: Verzamelen van gegevens

- Informatie bevat bestandsnamen en aanmaakdatum
- Cijfers voor Beeld en Geluid
 - 4,7 miljoen records
 - ruim 800 MB
 - CSV-bestand
- opschonen van de data
 - 12% viel weg (of 600K records)
 - bijvoorbeeld geen aanmaakdatum

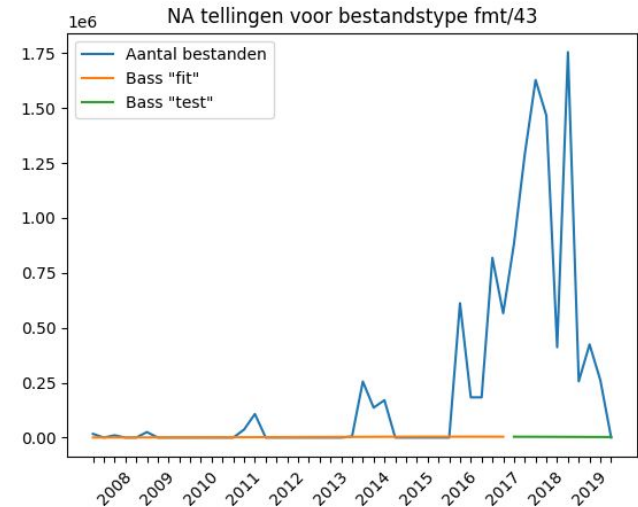
```
nibg:
  raw_csv_path: ~/data/NDE/digital_files.csv
  raw_csv_line_count: 4741017
  minimum_time_periods: 100
  json_output_dir: data/nibg/
  img_output_dir: images/nibg/
  num_test_measurements: 4
  linear_plots:
    - mxf
    - tif
    - tar
    - pdf
    - mpg
    - mp3
```



Datakwaliteit is een aandachtspunt



```
I,  
"fmt/43": {  
  "2017-11": 736903,  
  "\"\"\\n": 23114958,  
  "1601-01": 685,  
  "2016-11": 531885,
```



Stap 2: omzetten naar json

- Input data omzetten in json bestand
 - [nibg_aggregate_stats.json](#)
- Schrijven van stukje code om de data om te zetten
 - [nibg_aggregate.py](#)

Eerste opruimactie bestandsnamen

Tweede opruimactie datum

bestandsextensie overhouden

jaar en maand selecteren

```
with open(nibg_cfg['raw_csv_path'], 'rt') as f:
    for line_no, record in enumerate(tqdm(f, total=nibg_cfg['raw_csv_line_count'])):
        # Skip header with column names
        if line_no == 0:
            # header = record.split(',')
            continue

        parts = record.split(',')

        filename = str(parts[2])
        filetype = str(parts[3])
        # Skip if there is no dot in the filename to separate the file name extension
        if '.' not in filename:
            if filetype == '':
                skipped_records += 1
                continue
            else:
                filename = filetype

        create_date = parts[5]
        if create_date == '' or create_date == 'true' or create_date == 'false':
            skipped_records += 1
            continue

        extension = filename.split('.')[1].lower()
        file_temp_stats.setdefault(extension, {})

        year_month = '-'.join(create_date.split('-')[0:2])
        file_temp_stats[extension].setdefault(year_month, 0)
        file_temp_stats[extension][year_month] += 1
```

Formaten selecteren

```
# Prune stats for formats that have at least 10 entries
formats = list(file_temp_stats.keys())
dropped_formats: List[str] = []
min_measurements = nibg_cfg['minimum_time_periods']
```

Selecteren van formaten met voldoende voorkomen

```
for extension in formats:
    num_measurements = len(file_temp_stats[extension].keys())
    if num_measurements < min_measurements:
        logging.warning(f"Excluded {extension}: only {num_measurements} available, minimum is {min_measurements}")
        del file_temp_stats[extension]
        dropped_formats.append(extension)

output_path = os.path.join(nibg_cfg['json_output_dir'], 'nibg_aggregate_stats.json')
with open(output_path, 'wt') as f:
    f.write(json.dumps(file_temp_stats, indent=2))
```

Tellen voorkomen van een formaat

```
for format in file_temp_stats.keys():
    total = sum([count for count in file_temp_stats[format].values()])
    logging.info(f'{format=} {total=}')

```



Output bestand klaar

- json kan gebruikt worden voor analyse

```
"tif": {  
  "2011-02": 185702,  
  "2014-06": 46424,  
  "2014-05": 33926,  
  "2014-07": 3239,  
  "2014-02": 5714,  
  "2014-12": 267704,  
  "2014-03": 11055,  
  "2014-04": 20091,  
  "2014-10": 1357,  
  "2015-02": 9553,  
  "2014-09": 192,  
  "2017-10": 333,  
}
```



Stap 3: Analyse uitvoeren

- [nibg_analysis.py](#)

```
def main(config: Config) -> int:
```

```
    start = datetime.datetime.now()
```

```
    nibg_cfg = config['data']['nibg']
```

configuratie aanpassingen

```
    aggregate_stats_path = os.path.join(nibg_cfg['json_output_dir'], 'nibg_aggregate_stats.json')
```

```
    with open(aggregate_stats_path, 'rt') as f:
```

```
        aggregate_stats = json.loads(f.read())
```

```
    quarterly_counts = to_pruned_sorted_quarterly(aggregate_stats)
```

```
    plot_counts(quarterly_counts, nibg_cfg)
```

magie achter de schermen

```
    end = datetime.datetime.now()
```

```
    logging.info(f'Script took {end - start}')
```

```
    return 0
```



Stap 4: Analyseren van de uitkomsten

- Geen geautomatiseerd proces
- Context informatie van de instelling om gegevens te verklaren (zie voorbeelden andere presentaties)



Benodigdheden

- Python (getest van 3.10 -3.12)
- Pipenv
 - Creëert een aparte virtuele omgeving
 - Afhankelijkheden worden geïnstalleerd.



Installatie: Python en Pip

Zonder beheerdersrechten:

<https://apps.microsoft.com/store/detail/python-311/9NRWMJP3717K\>

Of anders:

<https://github.com/winpython/winpython/releases/download/6.4.20230625final/Winpython64-3.11.4.0dot.exe>

Download en voer uit. Je krijgt hiermee een “portable” installatie van Python.

Pipenv dependency+environment manager

```
pip install pipenv
```

Powershell:

```
Set-ItemProperty -path HKCU:\Environment\ -Name Path  
-Value "$((Get-ItemProperty -path HKCU:\Environment\ -Name  
Path).Path);$(python -c `"import site;  
print(site.USER_BASE + '\Python311\Scripts')`"))"
```

Log uit en weer in

Installatie van de analysetool

- Maak een nieuwe map aan genaamd, bijv “workshop”

- Pak uit: de inhoud van

<https://github.com/Antfield-Creations/NDE-monitoring-file-formats/archive/refs/heads/main.zip> in de map workshop

- Open de map workshop in je command shell.

Open de map

```
pipenv install --python=3.11
```

```
[requires]
```

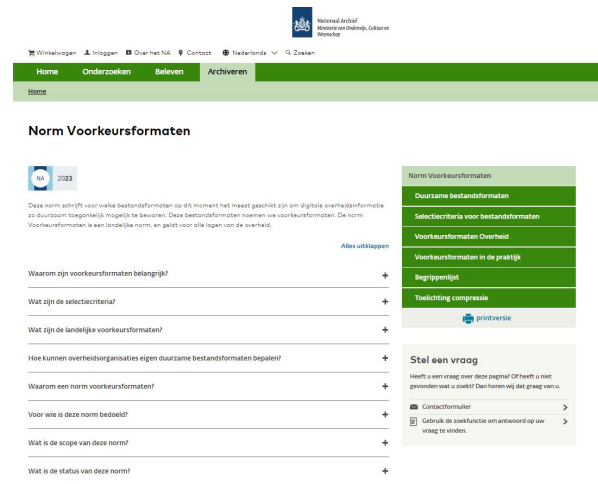
```
python_version = "3.10"
```

Je moet nog de PipFile aanpassen aangezien die de versie bevat

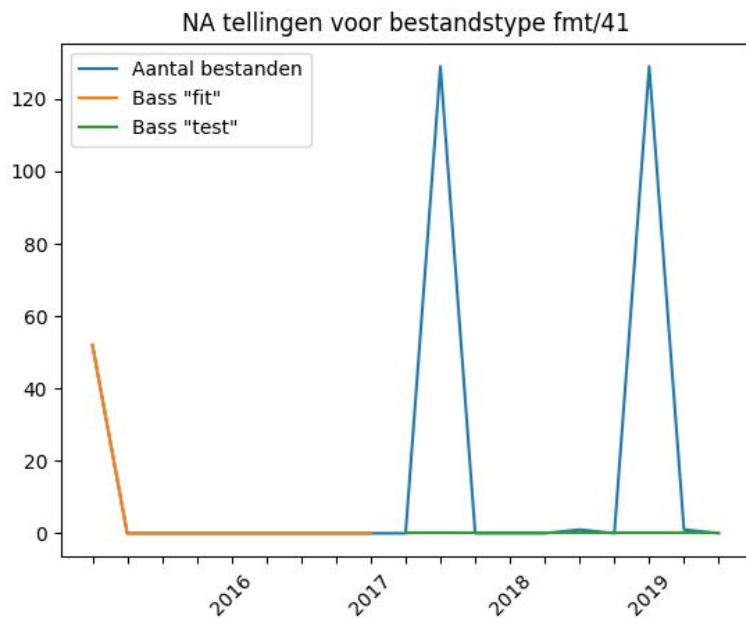
Uitvoeren van analyses

- `pipenv run python -m analysis.nibg_aggregate`
- `pipenv run python -m analysis.nibg_analysis`

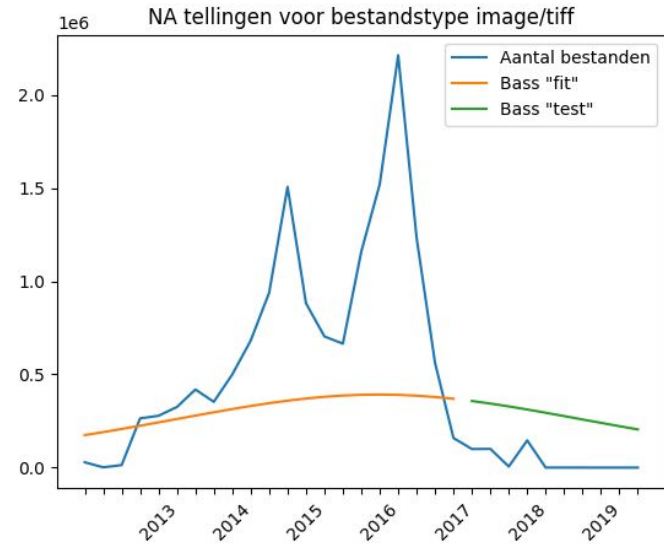
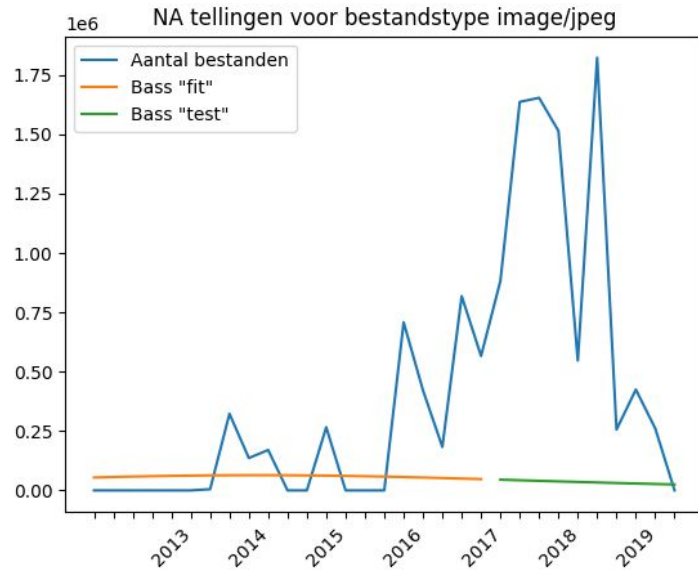
- Krijgen voor nu alles na een langere periode (± 20 jaar)
- Voorkeursformatenbeleid
- PRONOM Persistent Unique Identifier (**PUID**)
- Geen creatiedatum na 2019 (wel ingest)
- Nu vooral:
 - Komt haperend binnen
 - Vooral veel gedigitaliseerd materiaal



Soms past de schoen niet...

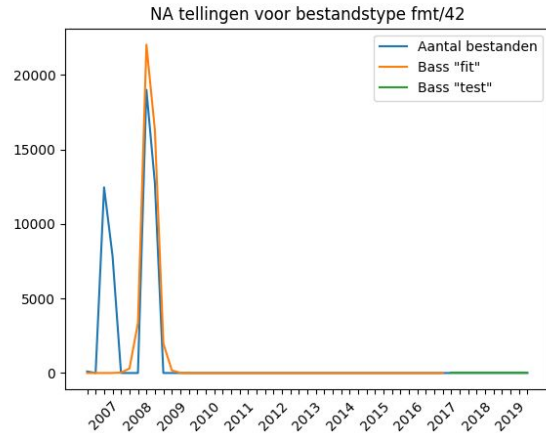


JPEG en TIFF

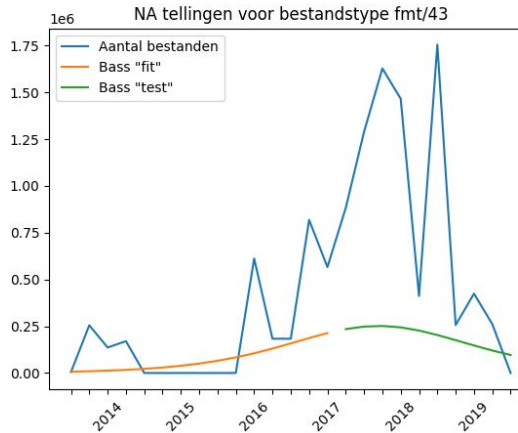


JPEG versies

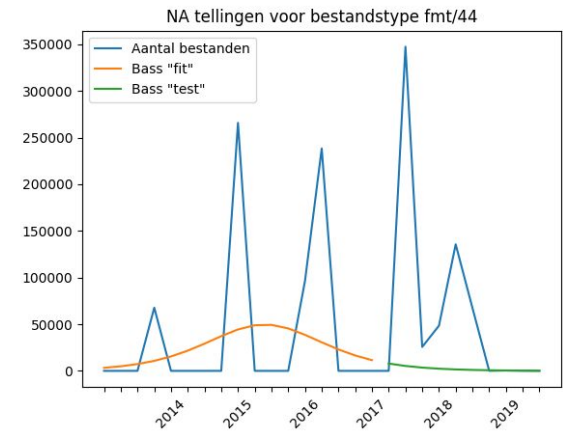
JPEG 1.00



JPEG 1.01

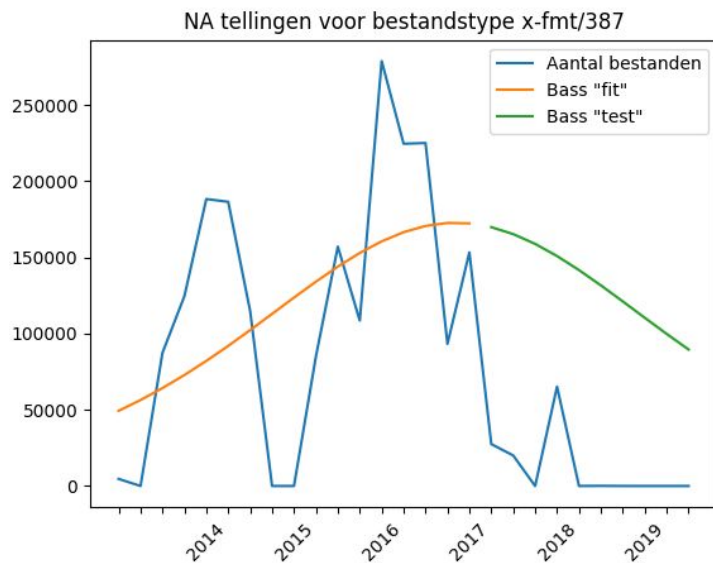


JPEG 1.02

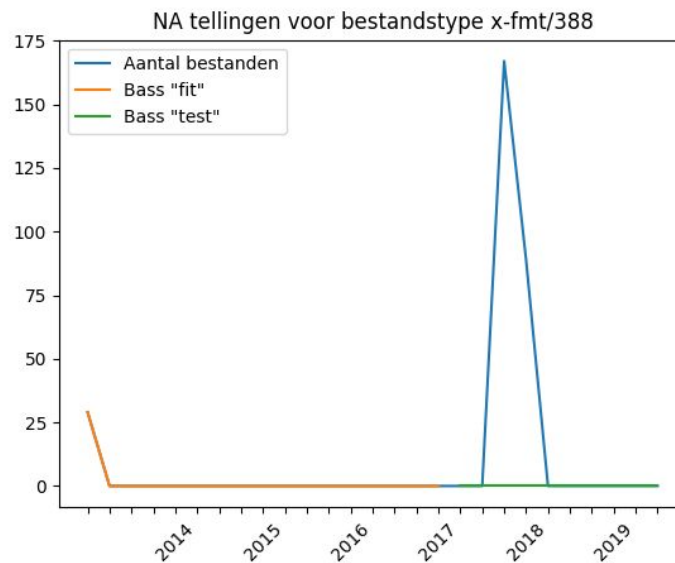


TIFF versies

TIFF versie 2.2

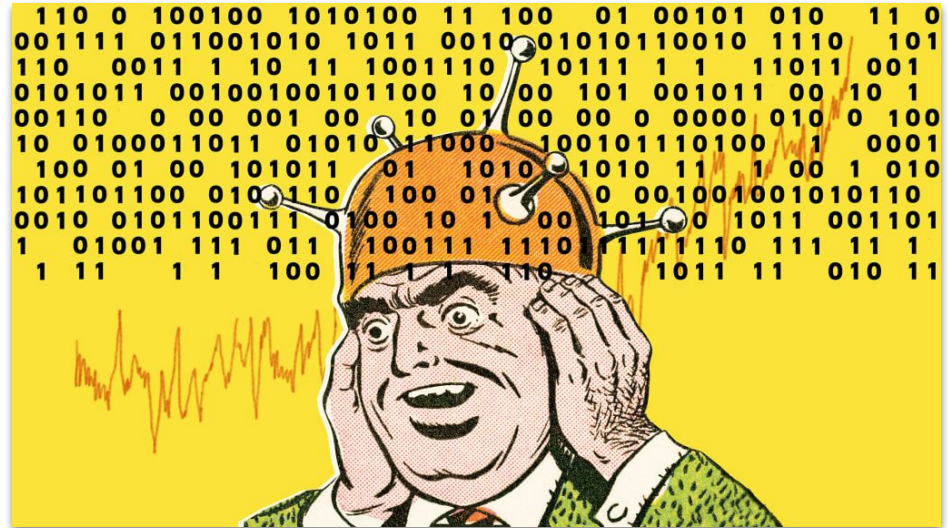


TIFF versie 2.1



Bevindingen

- Data heeft context nodig
- Levenscyclus detecteerbaar
- Overzicht



DANS

Werkwijze, zoals eerder gesteld:

1. Verzamelen van gegevens
 - a. CSV bestand met **creatiedatum bestand** en bestandsformaat
2. Omzetten van de data in bruikbaar formaat
 - a. Input CSV omzetten naar JSON
 - b. Data schonen**
 - c. Formaten selecteren**
3. Analyse uitvoeren
4. **Analysen** van de uitkomsten

Stappen 1,2 en 4 zijn organisatie specifiek

NB: Bij DANS is een JSON bestand gebruikt in plaats van een CSV

EASY (2005-2022)

Dataset_database_V (va. 15-01) X +

https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:155018

[edit]

EASY offers sustainable archiving of research data and access to thousands of datasets.

Search... SEARCH Search help

Advanced search Browse

OOIJEN-WANSSUM, GEBIEDSONTWIKKELING VINDPLAATS W4

Back to list

Status: Maintenance 2022-07-01 20:00 Delete Republish...

Assigned to: V.L. Glissen Change

Depositor: R.J.W.M. Gruben Change...

Overview Description Data files (208) Administration Activity log

Cite as:

Weterings, drs. P. (BAAC) (2019): Ooijen-Wanssum, gebiedsontwikkeling vindplaats W4. DANS. <https://doi.org/10.17026/dans-293-vbqg>

2019-10-10 | Weterings, drs. P. (BAAC) | [10.17026/dans-293-vbqg](https://doi.org/10.17026/dans-293-vbqg)

In opdracht van de Combinatie Moeder Maas heeft BAAC een Inventariserend Veldonderzoek door middel van Proefsluizen (IVO-F) uitgevoerd in onderzoekgebied W4 te Wanssum, gemeente Veenray. Aanleiding voor dit onderzoek was de voorgenomen gebiedsontwikkeling langs de Maas tussen Ooijen en Wanssum zal plaatsvinden. Op basis van een archeologisch vooronderzoek in de vorm van een kartierend booronderzoek-plus werd binnen onderzoekgebied W4 bewoning verwacht uit de ijertijd en de Romeinse tijd.

Het onderzoekgebied bevindt zich binnen een relatief laag gelegen deel van een terras uit het late Dryas. Tijdens deze periode maakte het gebied deel uit van de dalvlakte van een vlechtend riviersysteem. Binnen het terras zijn nog tal van restanten behorend tot het vlechtende systeem herkenbaar, die in het vroege Holoceen zullen zijn opgevoeld met kleine afzettingen van de Maas die in het Holoceen haar huidige loop ging aanmerken.

Tijdens het onderzoek werden geen sporen aangetroffen, met uitzondering van twee kleine recente verstoringen, een spoor van natuurlijke aard en een ondiepe kuit waarvan vermoed wordt dat deze ook een natuurlijke oorsprong heeft. Vondsten zijn daarentegen talrijker en werden in totaal 108 vondsten verzameld, variërend van aardewerk tot bouwmaterialen en natuursteen (waaronder vuursteen). Het vondstmateriaal kent een brede datering vanaf het neolithicum/bronstijd, via de late ijertijd tot en met de nieuwe tijd en is zonder uitzondering in de 21e eeuw aangetroffen; het post-Romeinse overblijfsel dat dit gebied tijdens overstromingen heeft afgedekt. Vanwege de grote variatie in datering, afname van de ligging in een jong overstromingspakket én het gegeven dat er op één mogelijke kuit na, geen sporen werden aangetroffen, is geconcludeerd dat zich binnen onderzoekgebied W4 geen vindplaats bevindt, wat leidt tot de conclusie dat het terrein niet behoudenswaardig is en vrijgegeven kan worden van nader archeologisch onderzoek.

Add jumpoff page

Metadata bevat:
Date Created
Date Deposited
Date Published

Dataset_database_V (va. 15-01) X +

https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:155018/tab/2

[edit]

EASY offers sustainable archiving of research data and access to thousands of datasets.

Search... SEARCH Search help

Advanced search Browse

OOIJEN-WANSSUM, GEBIEDSONTWIKKELING VINDPLAATS W4

Back to list

Status: Maintenance 2022-07-01 20:00 Delete Republish...

Assigned to: V.L. Glissen Change

Depositor: R.J.W.M. Gruben Change...

Overview Description Data files (208) Administration Activity log

Download View details Upload... Import... Delete -Visibility- -Accessibility- Change Filters...

Dataset Contents

☐	Name	Size	Creator	Visible to	Accessible to
☐	Documentatie		Archivist	Anonymous	Anonymous
☐	Fotos		Archivist	Anonymous	Anonymous
☐	GIS		Archivist	Anonymous	Anonymous
☐	Tabellen		Archivist	Anonymous	Anonymous
☐	original		Depositor	None	None

CONTACT

+31 88 003 46 66

info@dans.knaw.nl

More »

Newsletter

Twitter

YouTube

LinkedIn

Disclaimer

Legal information

Privacy statement

How to cite data

DANS is an institute of KNAW and NWO

KNAW NWO

Driven by data

CORE SEAL

nestor 2016

Data Files: oorspronkelijke
aanlevering in folder genaamd
original.
Gecureerde data staat buiten
deze folder.

Data Station

**EASY Date Created =>
Production Date**

The screenshot shows the DANS Data Station Archaeology homepage. The header includes the DANS logo and navigation links: About, User Guide, Support, Log In. Below the header, there's a search bar and a 'Search this dataverse...' field. The main content area displays search results for 'Archeologisch booronderzoek (IVO-B) Den Iip 53a te Den Iip, gemeente Landsmeer'. The results list includes the title, date (jun 16, 2023), and a brief description of the research project. The left sidebar shows filters for 'Dataverses (0)', 'Datasets (130,655)', and 'Files (1,954,741)'. The bottom section shows 'Methods of Recovery IABR'.

The screenshot shows the DANS Data Station Archaeology dataset page for 'Doorn, Driebergsestraatweg 63 Doorn, Driebergsestraatweg 63'. The page header includes the DANS logo and navigation links: About, User Guide, Support, V.L. Glissen (DANS) 1065. Below the header, there's a search bar and a 'Search this dataset...' field. The main content area displays the dataset title, version (Version 1.0), and a brief description of the research project. The left sidebar shows filters for 'Files', 'Metadata', 'Terms', and 'Versions'. The bottom section shows 'Change View' options (Table, Tree) and a list of files with download links.

Data Station

original => Version 1.0

The screenshot shows the main page for the dataset 'Plangebied Villa Nuova terrein te Groningen' on the DANS Data Station Archaeology platform. The page includes a header with the DANS logo and navigation links. The main content area features a description of the dataset, a 'Cite Dataset' button, and a 'Description' section with a detailed text about the archaeological findings. A 'Subject' section is also present. At the bottom, there is a table with columns for 'Dataset Version', 'Summary', 'Contributors', and 'Published on'.

Dataset Version	Summary	Contributors	Published on
2.0	Citation Metadata: DV PID Version (Changed); Bag ID (Changed); Files (Added: 2; Removed: 2); View Details	Abel Vries, de (DANS)	2023-06-15
1.0	This is the first published version.	E.J.M. Zwet, van der, Abel Vries, de (DANS)	2023-06-15

The screenshot shows a 'Version Differences Details' dialog box overlaid on the main page. The dialog box provides a comparison between Version 1.0 and Version 2.0. It includes a table with columns for 'Data Vault Metadata', 'Files', and 'Subject'. The 'Files' section lists three files with their IDs, SHA-1 hashes, and names. The 'Subject' section lists the same three files with their IDs, SHA-1 hashes, and names. The dialog box also includes a 'Done' button.

Data Vault Metadata	Files	Subject
DV PID Version: 1.0 Bag ID: urn:uuid:9a767264-455e-4705-91f0-fa5327b5c700	File ID 2665484 SHA-1: 56ab838113128f9aae7dfda62e5cb12e10b8dc0 Name: Gdri2_tabellijst.xlsx Type: MS Excel Spreadsheet Access: Public	File ID 2665503 SHA-1: 49a23cf076d55a81c84f0be0d1f52c6a9a5b0b0 Name: Gdri2_rapport_meta.xlsx Type: OpenOffice Spreadsheet Description: Access: Public
DV PID Version: 2.0 Bag ID: urn:uuid:6bcd563a-7475-4d02-a4ad-e2aea4e4fd4b	File ID 2665486 SHA-1: 7a6a066a7d7f6e6a4cd5807af941835af780c Name: Gdri2_rapport_meta.xlsx Type: MS Excel Spreadsheet Access: Public	File ID 2665504 SHA-1: bb686e573d2f5dd1c4d87de148c7426686de4580 Name: Gdri2_rapport_meta.xls Type: OpenOffice Spreadsheet Description: Access: Public

Voor de doelen van deze analyse kijken we alleen naar datasets waar we beschikken over een versie 1.0 met origineel aangeleverde bestandsnamen.

Analyse van data uit het Data Station

“We filteren op voorhand de datasets uit die zijn gemigreerd uit EASY, en datasets die ons de originele bestandsnamen kunnen vertellen. De complete versie-metadata van iedere dataset beschikbaar in de Archaeology Datastation wordt weggeschreven naar een bestand, met op iedere regel de JSON metadata voor een dataset. Om op deze manier door de meer dan 120.000 datasets heen te lopen kost het script iets meer dan een dag, resulterend in een bestand van 2,4 Gigabyte.”

“We extraheren de bestandstypes en de productiedatum voor de versie 1.0 van iedere dataset, die we samen aggregeren naar een nieuw JSON-bestand, gegroepeerd per bestandstype en per maand.”

“De filters hebben geresulteerd in een selectie van ruim 117 duizend datasets, voor een totaal van ruim 1,3 miljoen bestanden in 86 verschillende bestandstypes, gemiddeld ruwweg elf bestanden per dataset. De data van vóór 1997 laten we buiten beschouwing.”

Bron: Rein van ‘t Veer, 2022-12-06: *Monitoring van bestandsformaten 4: formaten in gebruik bij Data Archiving and Networked Services (DANS).*

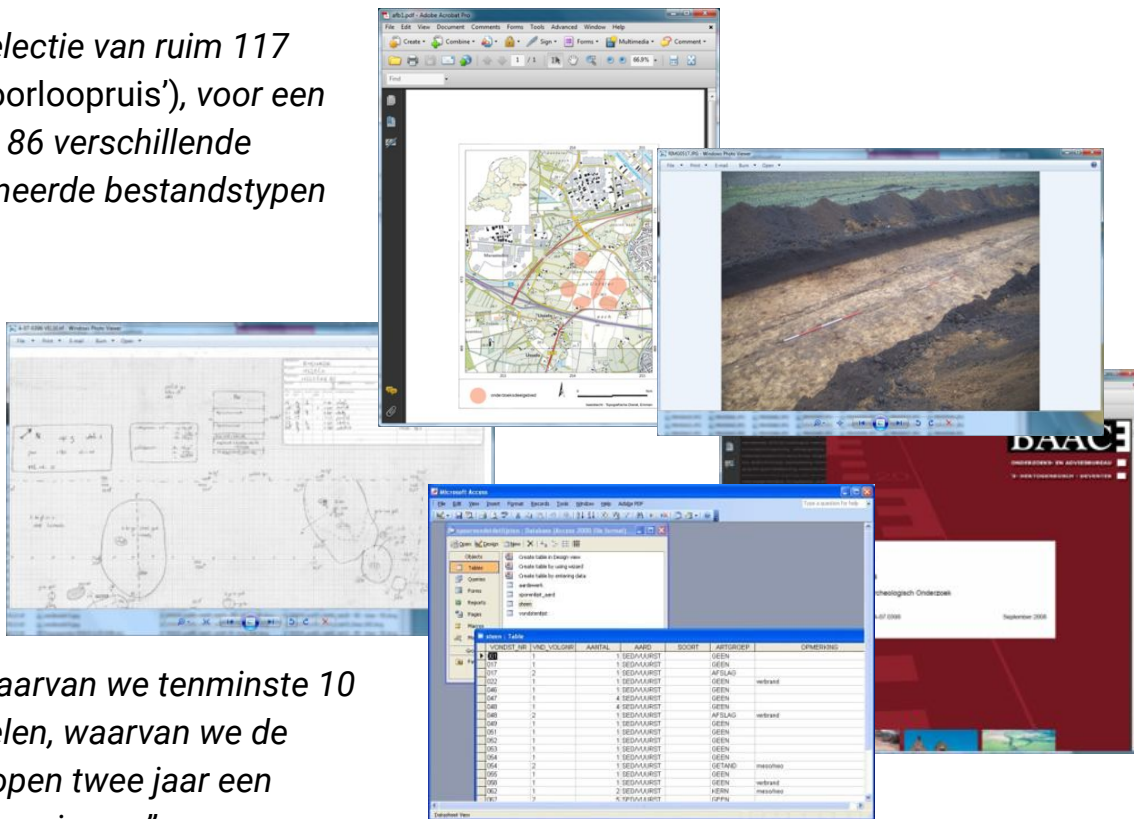
<https://kia.pleio.nl/groups/view/4fc4e83a-f55b-4000-b1cb-3fe9a16d3f93/kennisplatform-preservation/blog/view/b92f7a1e-fd6a-4c88-875c-5bccc470554c/monitoring-van-bestandsformaten-formaten-in-gebruik-bij-data-archiving-and-networked-services-dans>

Analyse van data uit het Data Station

“De filters hebben geresulteerd in een selectie van ruim 117 duizend datasets (van ná 1997 i.v.m. ‘voorloopruis’), voor een totaal van ruim 1,3 miljoen bestanden in 86 verschillende bestandstypes. De veruit meest gedeponeerde bestandstypen zijn, afnemend gesorteerd:

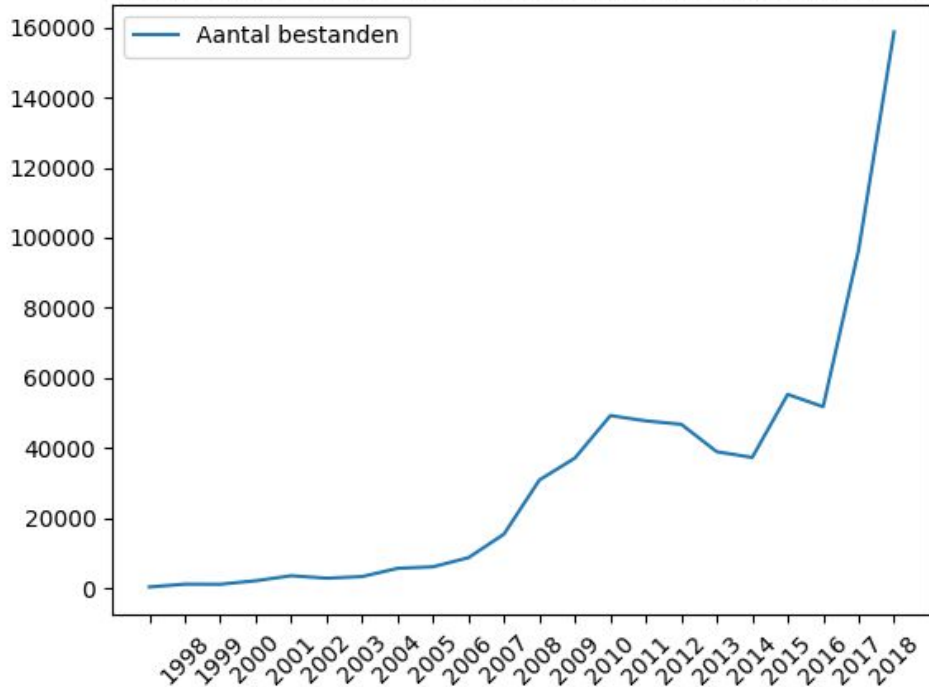
- JPEG (680 duizend),
- XML (316 duizend),
- PDF (78 duizend),
- CSV (33 duizend),
- DBase DBF (18 duizend),
- TIFF (16 duizend),
- MapInfo TAB (14 duizend)”

“We analyseren de bestandsformaten waarvan we tenminste 10 kwartalen aan tellingen kunnen verzamelen, waarvan we de bestandstypes overhouden die de afgelopen twee jaar een vermindering zagen in de aantallen deponeringen.”



DANS

Tellingen per periode voor alle bestandstypen gecombineerd



2005–2007: eDNA-project
2007: KNA-verplichting

2012–2013 economische crisis

2017: opkomst ArcheoDepot,
afstemming met formaatbeleid
DANS

Aanlevertermijn voor 2020+ is nog
niet verlopen: 2020–2022 is buiten
beschouwing gelaten

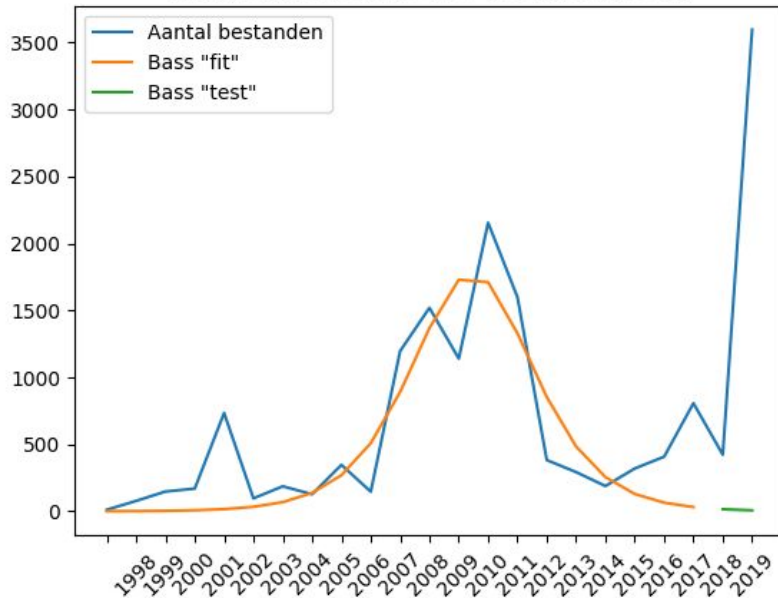
Lineaire-regressiemethode kijkt
vanaf piekmoment.
=> niet bruikbaar, enkel naar
Bass-methode gekeken



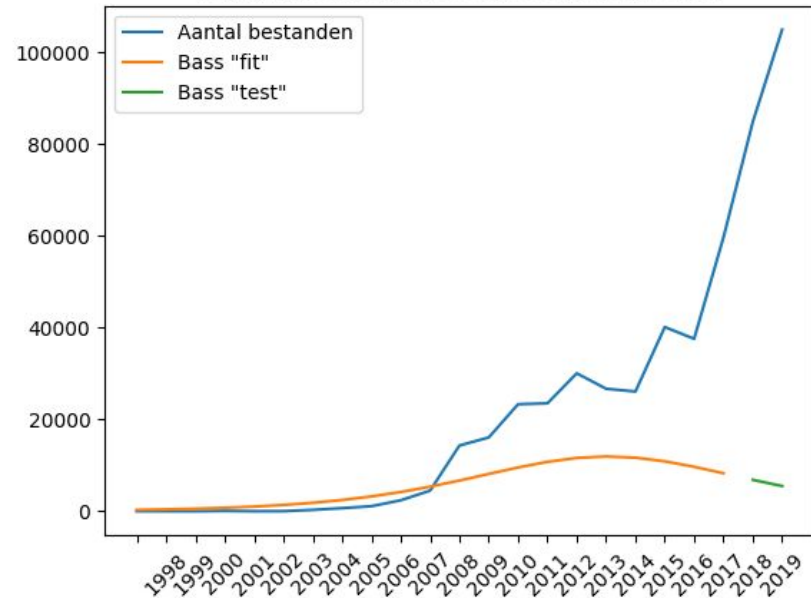
DANS

Veldfoto's = .jpg; Scans van tekeningen = .tif

Tellingen per periode voor bestandstype .tif



Tellingen per periode voor bestandstype .jpg

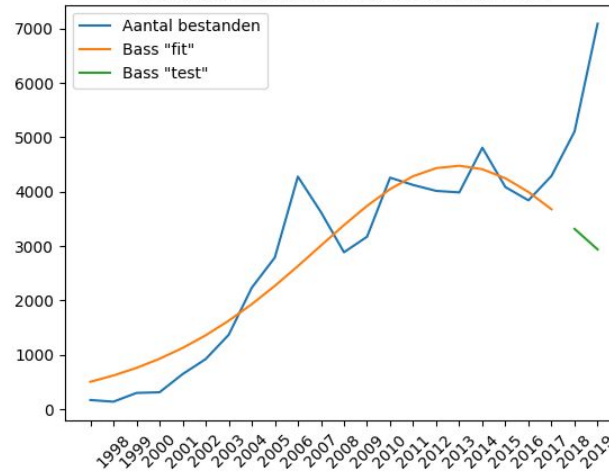


Bass-model weet niet goed te passen



DANS

Tellingen per periode voor bestandstype .pdf

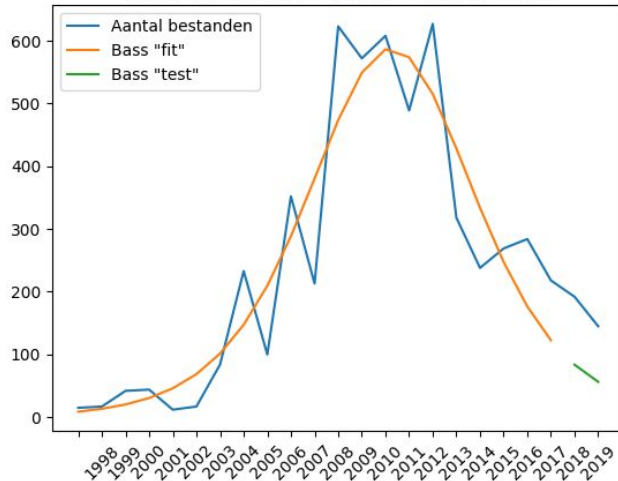


PDF = publicatieformat
voor rapportages.

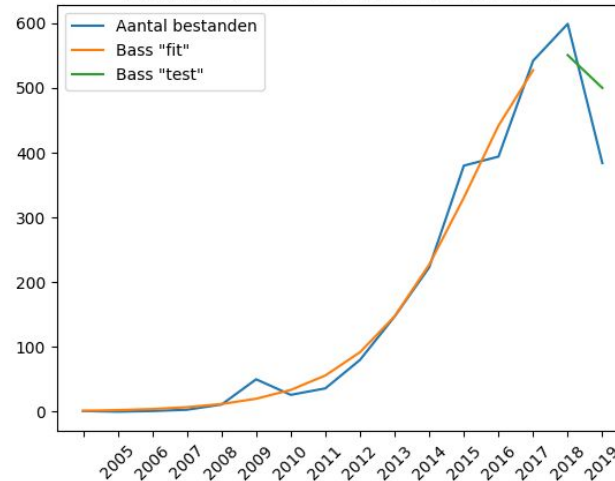
MS-Word =
dagrapporten en
andere rapportages

Duidelijke afname in
gebruik van .doc en
opkomst .docx

Tellingen per periode voor bestandstype .doc



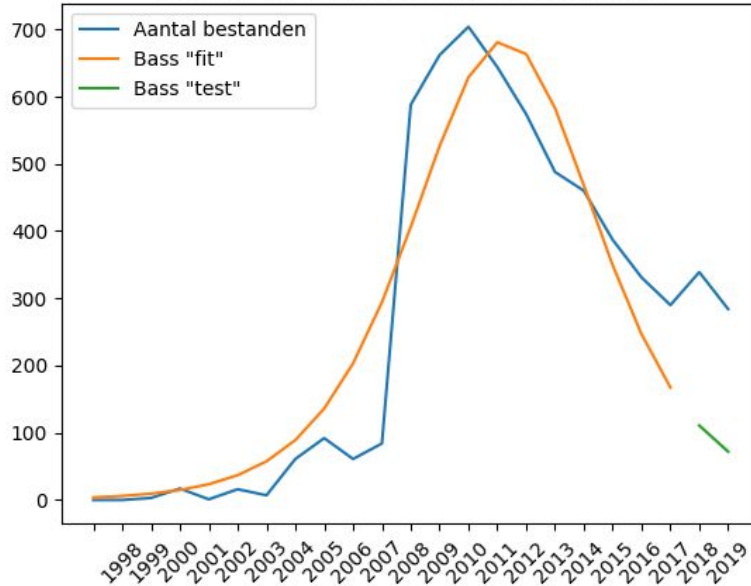
Tellingen per periode voor bestandstype .docx



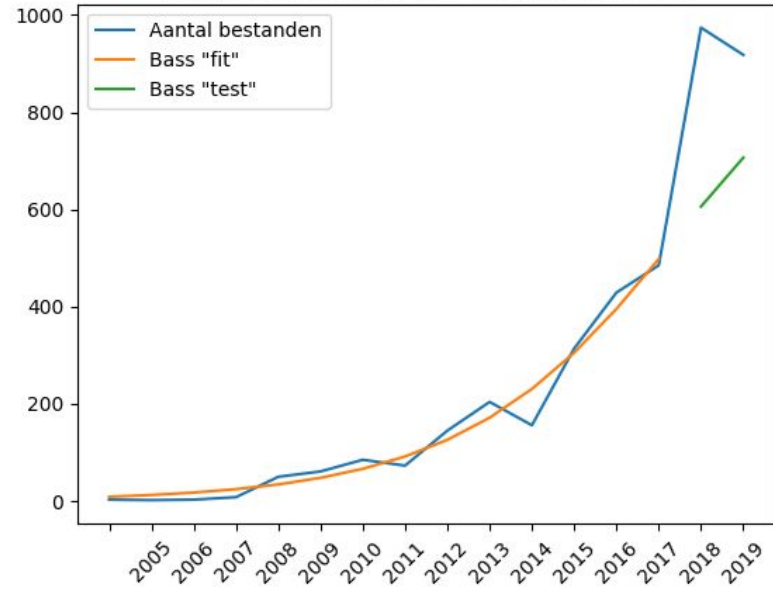
DANS

Duidelijke afname in
gebruik van .xls en
opkomst .xlsx

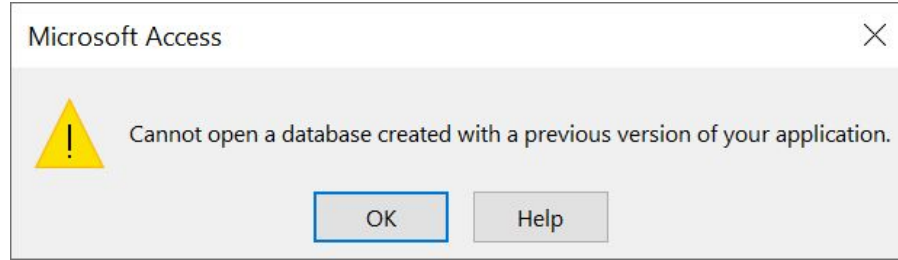
Tellingen per periode voor bestandstype .xls



Tellingen per periode voor bestandstype .xlsx

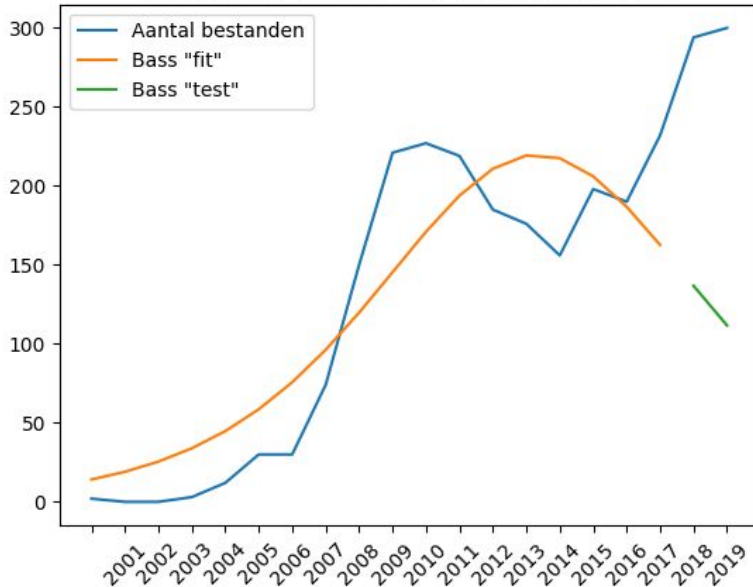


DANS

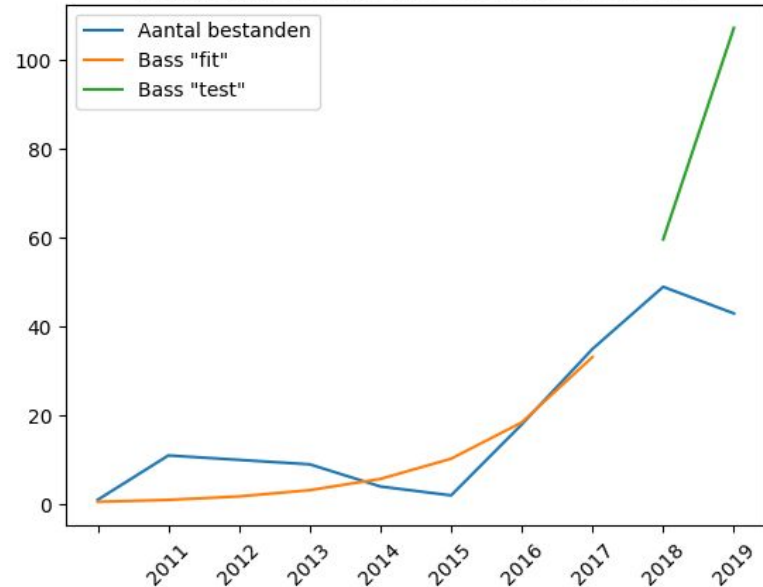


Nog weinig .accdb te zien. Mogelijke verklaring vanuit context: standaard opbouw van database bij organisaties.

Tellingen per periode voor bestandstype .mdb

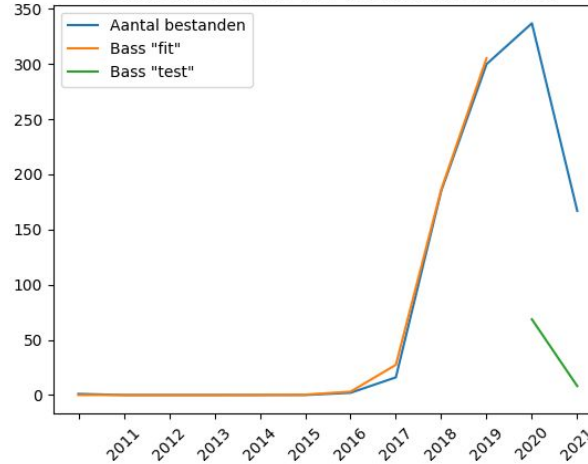


Tellingen per periode voor bestandstype .accdb



DANS

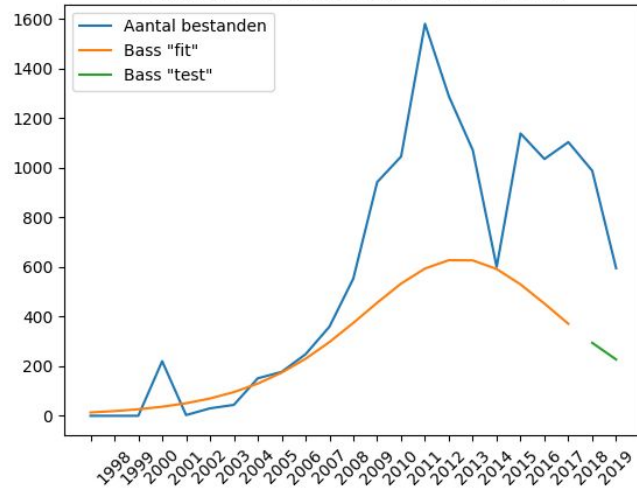
Tellingen per periode voor bestandstype .geojson



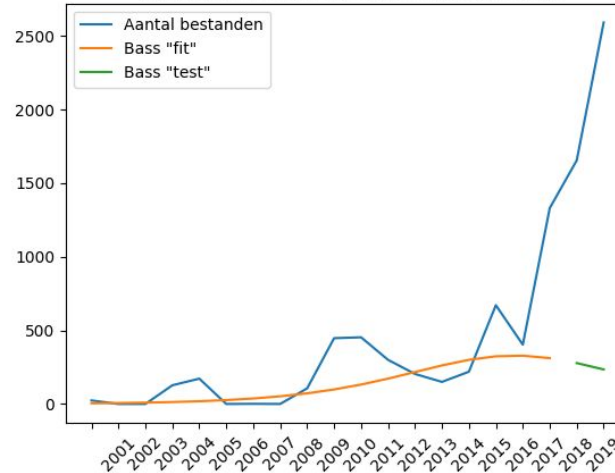
MapInfo .tab neemt
zichtbaar af.

Opkomst GeoPackage
is nog niet in beeld te
brengen

Tellingen per periode voor bestandstype .tab



Tellingen per periode voor bestandstype .shp



DANS

- Duidelijke transitie van oudere naar nieuwere formaten, met name bij bestandsformaten van Microsoft Office
- Uitzondering: Microsoft Access, met mogelijke verklaring vanuit de context
- Duidelijke opkomst van bepaalde formaten: GeoJSON
- Duidelijke afname van andere formaten: MapInfo TAB
- Bass-model algoritme kan moeilijk omgaan met trends van grote pieken en dalen



Eindconclusie

- Levert belangrijke inzichten, maar:
 - Datakwaliteit is belangrijk
 - Context is belangrijk!
 - Geen geautomatiseerd proces
 - Gedigitaliseerde bestanden leveren geen goede resultaten op
 - Voor het wereldwijde niveau ontbreekt de data nog
 - (Webarchief is niet hetzelfde als een instelling)
 - Relatie met applicaties is nog onduidelijk
 - Goede bron ontbreekt nog
 - Voorspellend karakter Bass model is te beperkt
 - Grillige grafieken geven slechte resultaten

Praktisch aan de slag!

Vragen?

Sam Alloing:

- Digital Preservation Officer KB
- sam.alloing@kb.nl

Valentijn Gilissen:

- Data Processing Team Leader, Preservation Officer, Data Steward
- valentijn.gilissen@dans.knaw.nl

Marin Rappard:

- Digital Preservation Adviseur Nationaal Archief
- marin.rappard@nationaalarchief.nl

Links

- Code: <https://github.com/Antfield-Creations/NDE-monitoring-file-formats>
- Blogs van het onderzoek
 - blog 1: [Introductie](#)
 - blog 2: [Het internet als archief, het Bass-model in de praktijk](#)
 - blog 3: [Formaten in gebruik bij Beeld en Geluid](#)
 - blog 4: [Formaten in gebruik bij DANS](#)
 - blog 5: [Applicaties voor verdwijnende bestandsformaten](#)