

# Webarchiv der KONINKLIJKE BIBLIOTHEEK



# Webarchiv der KONINKLIJKE BIBLIOTHEEK

Trienka Rohrbach

Funktionsmanager Webarchiv



# Webarchiv

- Begonnen in 2007 mit einer kleinen Sammlung
- Das Archiv ist eine begründete Auswahl
- Jede Webseite ist manuell ausgewählt worden, mit Fokus auf die Niederländische Sprache, Kultur und Geschichte.



# WEBARCHIV

## KEIN LEGALES E-DEPOT IN NL

- Jeder Eigentümer ist sein eigener Verlag
- Jeden individuellen Eigentümer müssen wir fragen, ob wir archivieren dürfen

# WEBARCHIV

WIE?

- Opt-in- oder Opt-out Verfahren
- BL in 2006 Opt-in 20% reagiert, davon Teil negativ
- Wir machen Opt-out
- Von insgesamt > 21.000 opt-out 220 negativ

# WEBARCHIV

## REAKTIONEN

- *Uw initiatief om mijn website gewoon te kapen en op uw eigen site te publiceren wordt niet getolereerd. Zo nodig neem ik juridische stappen.*
- *Ik wens niets met uw organisatie te maken te hebben. Ik geef geen toestemming*
- *Zowel de website [...] als de mensen die eraan meewerken worden stelselmatig gedemoniseerd en onrechtmatig behandeld door de onderdanen van de koning. Daardoor zijn wij genoodzaakt BEZWAAR te maken tegen het registreren van de website [...] door de webarchivering van de Koninklijke Bibliotheek.*
- *The website [...] has been inactive for the last 28 or 30 years. I feel that there is no reason to archive the data and would appreciate it if you would exclude the site from the process.*

# WEBARCHIV

## WAS BENUTZEN WIR

- WebCuratorTool



- Heritrix



- WARC



# WEBARCHIV

WANN

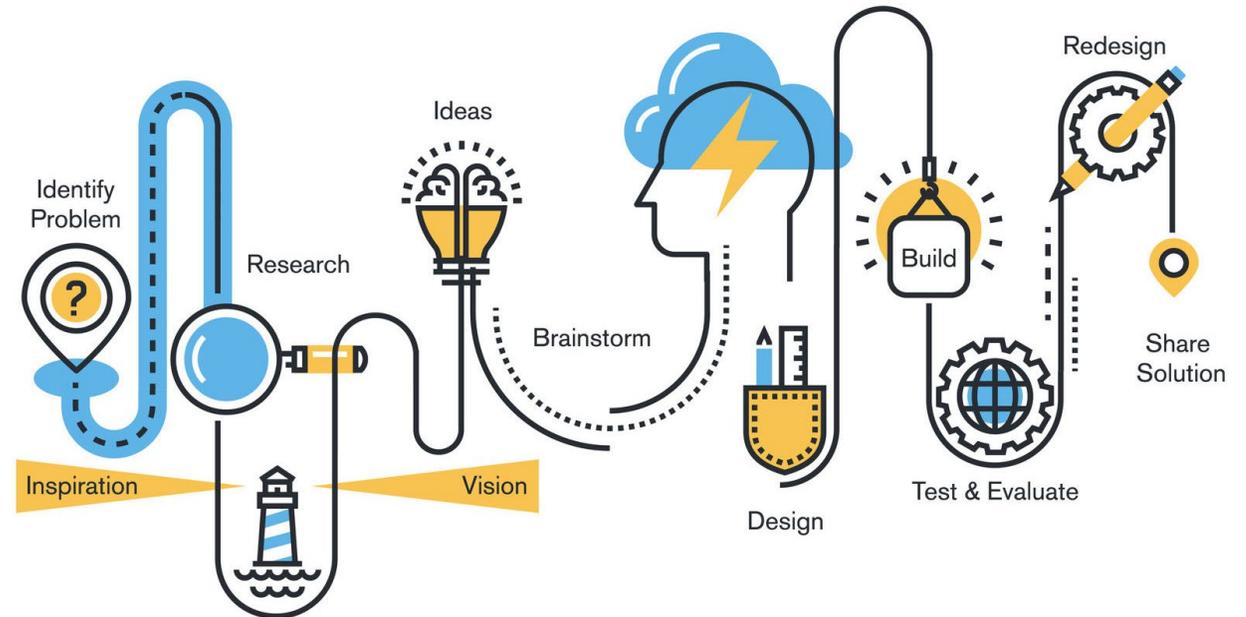
Datum	WebCuratorTool	Heritrix
Sept 2007 – August 2019	WCT 1.14.4	Heritrix 1
August 2019 – August 2020	WCT 2.0.2	Heritrix 3.3.0-SNAPSHOT
August 2020 – Mai 2021	WCT 3.0.0-SNAPSHOT	Heritrix 3.3.0-LBS2016-02
Mai 2021 – Heute	WCT 3.0.0	Heritrix 3.3.0-LBS-2016-02

# WEBARCHIV

# WEBCURATOR TOOL

WebCuratorTool wird weiter entwickelt von

KB } nationale  
bibliothek



# WEBARCHIV

## WEBCURATORTOOL

Ambassade van de Bondsrepubliek Duitsland [incl. vertegenwoordigingen zonder eigen website]

General Seeds Profile Schedule Annotations Description Groups Access

**Id:** 25690116

**Name:**

**Description:**

**Seed**

<http://www.niederlande.diplo.de/>

**Schedule**

35 14 15 4 ? \*

**Notes**

Ambassades en consulaten

# WEBARCHIV

## WEBCURATORTOOL

dec 2013	jan 2014 - dec 2015	jan 2016 - dec 2017	jan 2018 - dec 2019	jan 2020 - dec 2021
s	2 pages	2 pages	2 pages	6 pages
	<a href="#">apr 15, 2014</a> *	<a href="#">apr 15, 2016</a> *	<a href="#">apr 15, 2018</a> *	<a href="#">apr 15, 2020</a>
	<a href="#">apr 15, 2015</a> *	<a href="#">apr 15, 2017</a> *	<a href="#">apr 15, 2019</a>	<a href="#">apr 15, 2020</a>
				<a href="#">apr 15, 2020</a> *
				<a href="#">apr 15, 2021</a> *
				<a href="#">apr 15, 2021</a>
				<a href="#">apr 15, 2021</a> *

## Deutsche Vertretungen in den Niederlanden

Herzlich willkommen auf der gemeinsamen Webseite der Deutschen Botschaft Den Haag und des Generalkonsulats Amsterdam!

Bitte beachten Sie, dass die Botschaft in Den Haag und das Generalkonsulat in Amsterdam am 27.04.2021 geschlossen sind.

Den Haag

10 °C  
Den Haag  
© dwd.de

Beschränkungen aufgrund Coronavirus in den Niederlanden

Ein- bzw. Rückreise und Transit aus den Niederlanden nach Deutschland

Ambassade van de Bondsrepubliek Duitsland [incl. vertegenwoordigingen zonder eigen website]	15/04/2014 14:38:39	Harvested			
Ambassade van de Bondsrepubliek Duitsland [incl. vertegenwoordigingen zonder eigen website]	15/04/2015 16:08:27	Harvested			
Ambassade van de Bondsrepubliek Duitsland [incl. vertegenwoordigingen zonder eigen website]	15/04/2016 17:13:14	Harvested			
Ambassade van de Bondsrepubliek Duitsland [incl. vertegenwoordigingen zonder eigen website]	15/04/2017 17:48:59	Harvested	00:00:33:03	193.71 MB	2823
Ambassade van de Bondsrepubliek Duitsland [incl. vertegenwoordigingen zonder eigen website]	15/04/2018 14:58:28	Harvested	00:00:00:20	1.36 KB	2
Ambassade van de Bondsrepubliek Duitsland [incl. vertegenwoordigingen zonder eigen website]	15/04/2019 15:23:00	Harvested	00:00:00:38	1.66 KB	4
Ambassade van de Bondsrepubliek Duitsland [incl. vertegenwoordigingen zonder eigen website]	15/04/2020 14:52:05	Harvested	00:00:22:56	113.97 MB	1359
Ambassade van de Bondsrepubliek Duitsland [incl. vertegenwoordigingen zonder eigen website]	15/04/2021 15:12:18	Harvested	00:00:37:39	187.79 MB	2126

# WEBARCHIV

## KWALITÄTSKONTROLLE

n	extract_ja	extract_ja	block_url	block_url	from_date	last_proc	type	schedules	annotation	dublin_co	dublin_co	aantal_ur	aantal_ur	harvest_j	data_amc	elapsed_t	start_time	
t	f	f	{}	f				{}	{"XS4ALL	XS4ALL	19	48	0	Finished	1.97	00:00:45	2021-10-20 12:01:28.889	
t	f	f	{}	f	2008-08-1	2021-07-1	-6	{"00 36 15	{}			8	10669	0	Finished	262.46	00:11:19	2021-10-11 22:32:28.751

- Zeitliche grenzen (> 24:00:00)
- Daten grenzen (> 10 GB)
- Maximale tiefe in webseiten (> 2 oder 3 oder 4 oder ..)
- Dokuments grenzen (> 50.000)
- Angegeben das etwas ausgeschlossen soll werden aber keine Werte
  
- Alle Probleme werden behandelt, alles wird festgelegt in einer Anmerkung in der WCT

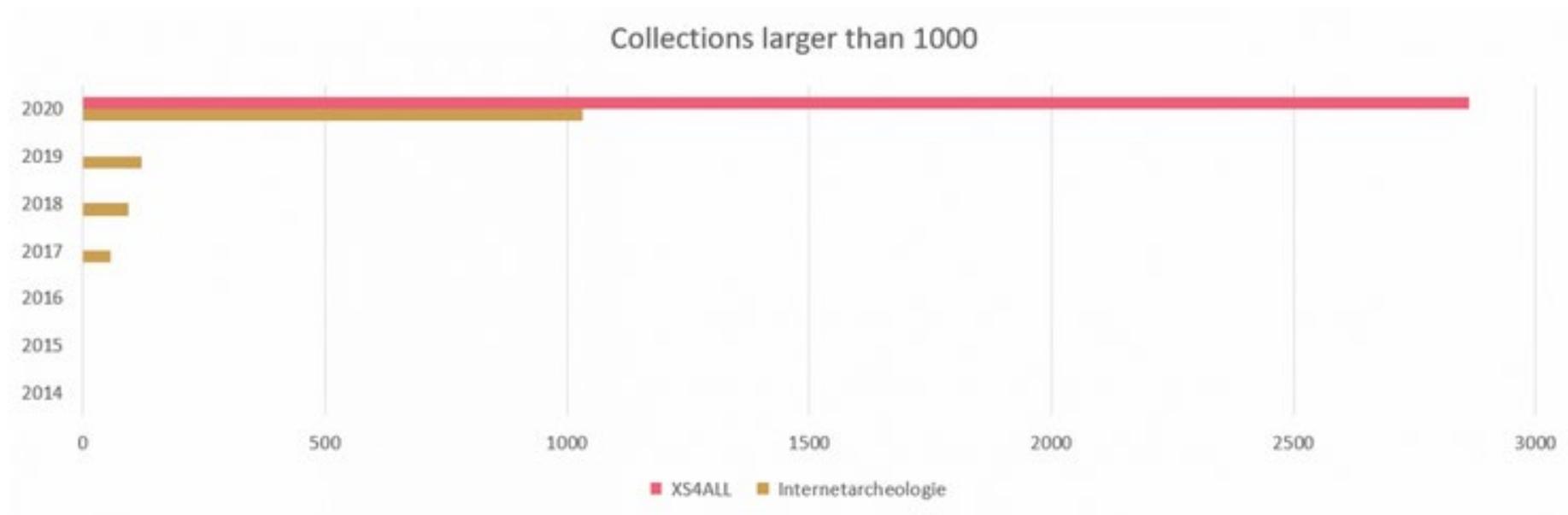
# WEBARCHIV

## SPEZIELLE SAMMLUNGEN

- 100<sup>e</sup> Vierdaagse von Nimwegen (4 Tage wandern rundum und durch Nimwegen in 2016)
- 200 Jahre Königreich der Niederlande (in 2013)
- Botschaften und Konsulate
- (Natürlich) Covid 19
- Kloster
- Heilige Nicolaus
- Wattenmeer
- XS4all

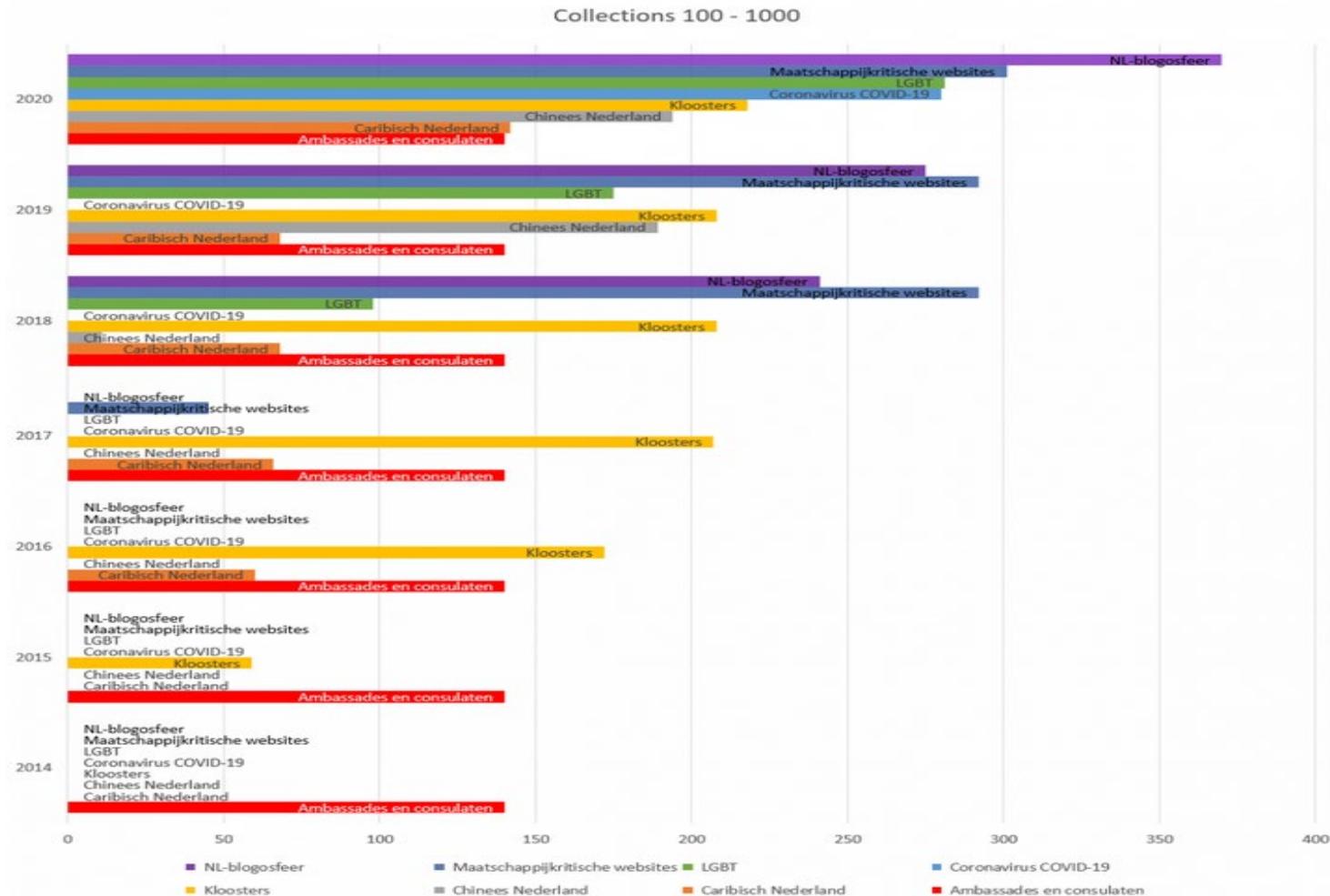
# WEBARCHIV

## SPEZIELLE SAMMLUNGEN



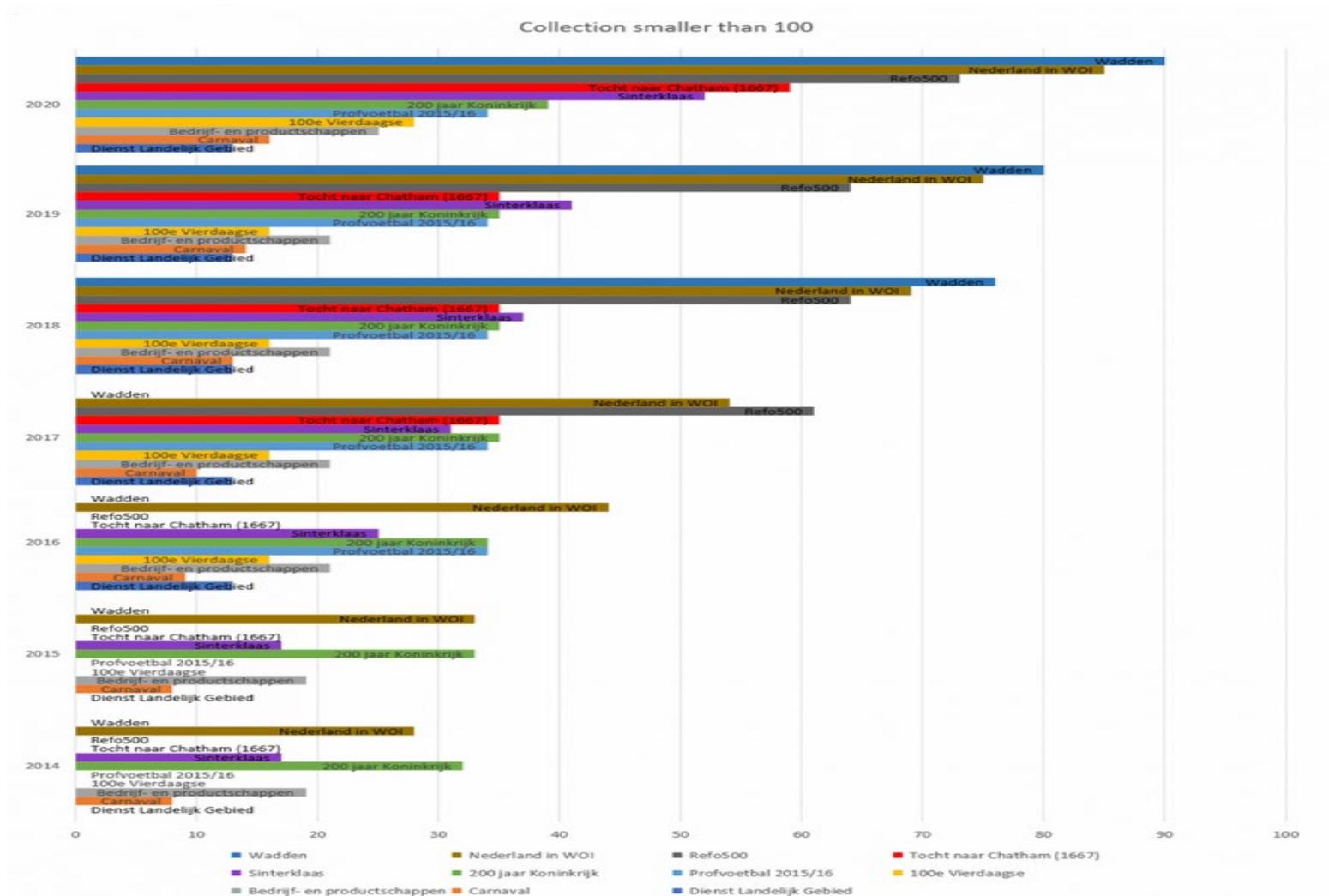
# WEBARCHIV

## SPEZIELLE SAMMLUNGEN



# WEBARCHIV

## SPEZIELLE SAMMLUNGEN



# Facten & Zahlen

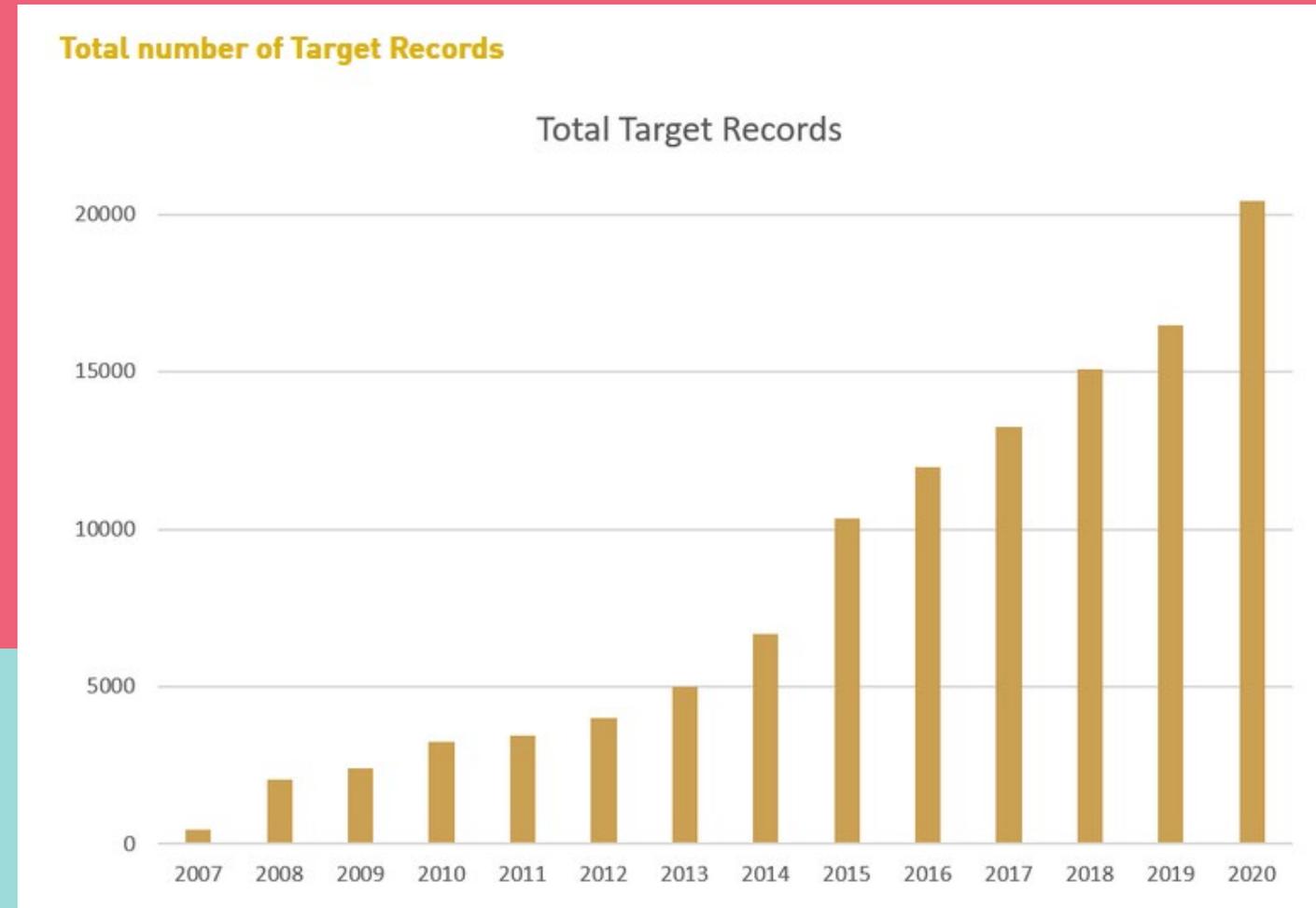
# WEBARCHIV

In 2021:

- > 21.000 Seiten
- > 430.000.000 Dateien
- > 56 Terabyte

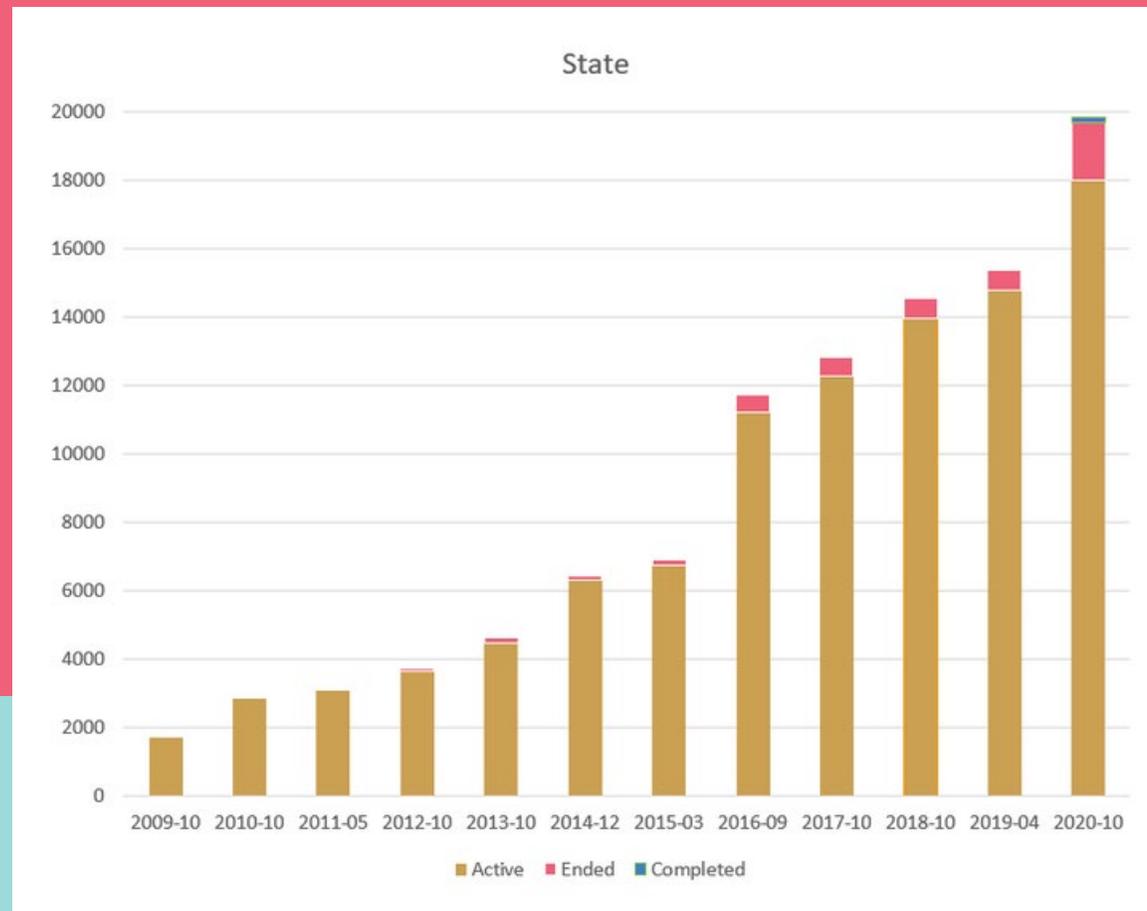
# Facten & Zahlen

## TOTAL TARGET RECORDS



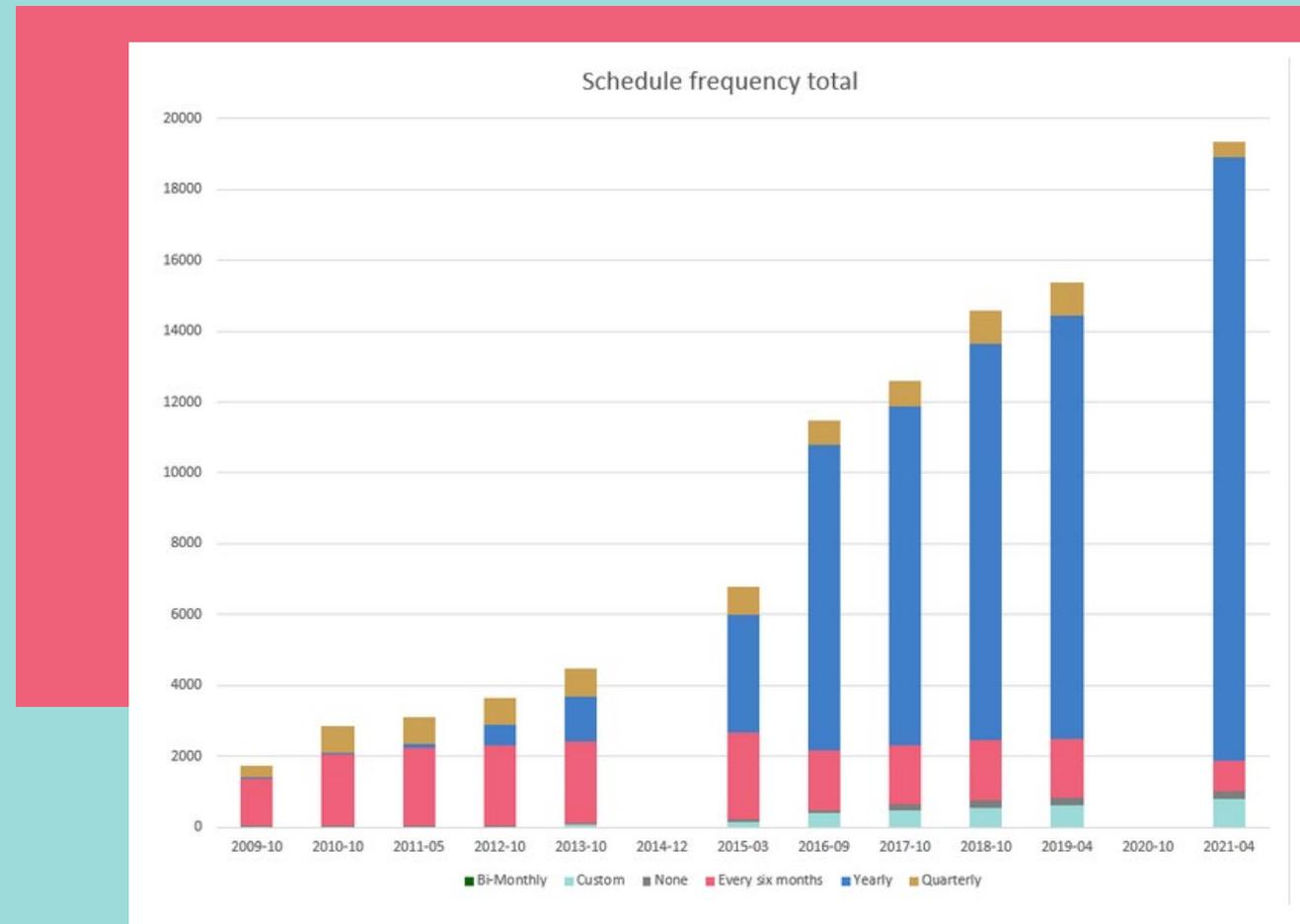
# Facten & Zahlen in 2021

## STATUS DER TARGET RECORDS



# Facten & Zahlen

## FREQUENZ



# Fragen?

Mehr information: [www.kb.nl](http://www.kb.nl)  
(Niederländisch und Englisch)

# Danke schön für die Aufmerksamkeit

Die daten sind hergestellt worden von Iris Geldermans

Mehr information: [www.kb.nl](http://www.kb.nl)  
(Niederländisch und Englisch)



KB } nationale  
bibliotheek