

## Rapport Werkgroep Duurzame Toegankelijkheid 1

Versie 1.0

Datum	8 mei 2018
Status	Definitief

## Colofon

Projectnaam	Project E-mailarchivering Rijk
Projectleider(s)	R.J.B. Hageman
Contactpersoon	R.J.B. Hageman T +31 6 55 26 78 86 F +31-70-331 5477 <a href="mailto:robert.jan.hageman@nationaalarchief.nl">robert.jan.hageman@nationaalarchief.nl</a> Postbus 90520   2509 LM Den Haag
Auteurs	V. Hoolt
Werkgroep	Vincent Hoolt, Pepijn Lucker, Buddy Joe Groeneveld, Roberto Wiredu, Mette van Essen, Remke Verdegem, Vincent Sleebe, Nico van Egmond, Vincent Kleijnen, Remco van Veenendaal, Tom Derks, Robbert Han Hageman, Joleyn Vrijmoeth
Meelezers	Bodien Abels, Jos Boerties, Erik Kraai, Jeroen Padmos
Versie	1.0
Bijlage(n)	2

## Project e-mailarchivering Rijk

Voor u ligt een rapport dat is opgesteld in het kader van het onderzoeksprogramma van het project e-mailarchivering Rijk, een samenwerkingsproject van de ministeries van BZK en OCW. Het project maakt onderdeel uit van het programma Rijk aan Informatie (RaI). Doel van het programma RaI is om nieuwe kaders op te stellen voor het informatiebeheer van het Rijk, die rekening houden met de ontwikkeling van digitaal werken en archiveren.

Wat betreft e-mail is geconstateerd dat dit type informatie bij het Rijk slechts sporadisch wordt gearchiveerd.<sup>1</sup> De bestaande werkwijze, waarbij iedere individuele medewerker wordt geacht om relevante e-mail handmatig op te slaan in een DMS of andere gemeenschappelijke archiveringsvoorziening, sluit niet goed aan op de praktijk, met als gevolg dat e-mail onvolledig en/of willekeurig wordt gearchiveerd. Dit levert voor de korte en middellange termijn (10 jaar) het risico dat door de organisaties van het Rijk niet adequaat kan worden gereageerd op informatieverzoeken en voor de lange termijn dat er bedenkelijke lacunes ontstaan in het nationaal geheugen, dat wil zeggen de toekomstige collectie van het Nationaal Archief.

Het project e-mailarchivering beoogt daarom een nieuwe werkwijze te ontwikkelen waardoor een tweetal doelen wordt gerealiseerd.

- Het veiligstellen van de inhoud van de e-mailboxen van het Rijk, zolang de desbetreffende informatie beschikbaar moet zijn voor het behandelen van WOB-verzoeken, parlementaire enquêtes, etc. Veiligstelling brengt met zich mee dat de e-mail door de betrokken organisaties in goede, geordende en toegankelijke staat wordt gearchiveerd.
- Het blijvend bewaren van de gearchiveerde e-mails die met het oog op verschillende doeleinden<sup>2</sup> van waarde worden geacht. Blijvende bewaring impliceert het op termijn overbrengen naar het Nationaal Archief, waar de betreffende e-mail door iedereen mag worden ingezien voor zover daar door de overbrengende instantie geen tijdelijke beperkingen aan de openbaarheid zijn gesteld.

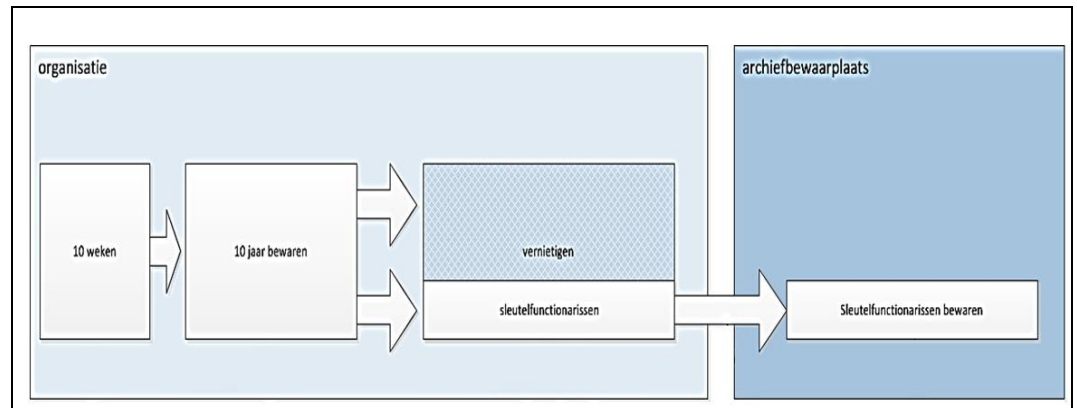
De voorgestelde nieuwe werkwijze bestaat uit de volgende elementen:

- E-mail verzonden of ontvangen door medewerkers van de Rijksoverheid wordt, tien weken na verzending of ontvangst, door de ICT-dienstverlener automatisch in een aparte omgeving opgeslagen.
- Medewerkers kunnen in de eerste tien weken na verzending of ontvangst privé e-mail, p-vertrouwelijke zaken of andere, niet-werkgerelateerde e-mail verwijderen.
- Alle e-mail wordt tien jaar opgeslagen, waarna deze wordt verwijderd. Van aan te wijzen sleutelfunctionarissen wordt de e-mail permanent bewaard.
- Na uiterlijk 20 jaar, of zoveel eerder als ministeries beslissen, worden de e-mails van sleutelfunctionarissen naar het Nationaal Archief overgebracht.

1 Zie de whitepaper 'Archiveren van E-mail in de Rijksoverheid' (november 2016) en het rapport van de Erfgoedinspectie 'Onvoltooid Digitaal' (oktober 2015).

2 Te weten het geheugen van de overheid, ten behoeve van de recht- en bewijszoekende burger, wetenschappelijk en historisch onderzoek, en cultureel erfgoed.

Deze wordt openbaar beschikbaar gesteld, voor zover door de zorgdrager geen openbaarheidsbeperkingen zijn aangebracht.



Figuur 1: schematische weergave van procesgang e-mail

De werkwijze beoogt het volgende toepassingsbereik: een werking voor de gehele rijksoverheid; invoering zonder terugwerkende kracht; formulering van generieke principes en aanpak.

Dat betekent dat de werkwijze:

- naast ministeries ook uitvoeringsorganisaties raakt;
- geen betrekking heeft op het veiligstellen of blijvend bewaren van de reeds gevormde inhoud van e-mailboxen;
- elk onderdeel van het Rijk verantwoordelijk is en blijft voor de manier waarop men nadere invulling geeft aan de gegeven werkwijze.

Ook moet worden bedacht dat de voorgestelde nieuwe werkwijze bedoeld is om een bestaand probleem met nu of binnenkort beschikbare middelen op te lossen. De bewaartermijn van 10 jaar voor het veiligstellen van alle e-mail is een indicatie voor de minimale tijdhorizon van de werkwijze.

Het onderzoeksprogramma van het project is bedoeld om de voorgestelde werkwijze op onderdelen te toetsen op bruikbaarheid en uitvoerbaarheid en om deze zo efficiënt en effectief mogelijk te kunnen inzetten. Het programma omvat vier onderzoekslijnen:

1. Een analyse van toepassing en gebruiksgemak van de nieuwe werkwijze
2. Waardering en selectie (sleutelfunctionarissen)
3. Duurzame toegankelijkheid
4. Openbaarheid en privacy.

Een overzicht van de hoofd- en subvragen per onderzoekslijn vindt u in Bijlage 1.

Het onderzoeksprogramma bestaat enerzijds uit bureauonderzoek en workshops, anderzijds uit het uitvoeren van enkele pilots. De algemene aanpak is als volgt: het bureauonderzoek levert input voor de pilots, namelijk kaders en uitgangspunten, scenario's en toetsbare aannames. In de pilots worden aannames getoetst en

scenario's uitgeprobeerd. Na analyse van de uitkomsten van de pilots volgt per onderzoekslijn een handreiking met aanbevelingen voor de nadere invulling en implementatie van de richtlijn.

## Inhoud

<b>Samenvatting</b>	<b>7</b>
<b>1 Inleiding</b>	<b>9</b>
<b>2 Deskresearch</b>	<b>10</b>
2.1 Volume en beheerslast	10
2.2 Duurzame toegankelijkheid	11
<b>3 Resultaten impactanalyse</b>	<b>13</b>
3.1 Normen voor opname e-mail	13
3.2 Impactanalyse goede, geordende en toegankelijke staat	13
3.3 Impactanalyse metadata	19
3.4 Impactanalyse preservering	21
3.5 Impactanalyse functionele en technische eisen	22
<b>4 Ontwikkelingen architectuur</b>	<b>26</b>
4.1 Transformatie als keuze in de keten	26
<b>5 Conclusie en aanbevelingen</b>	<b>29</b>
5.1 Aanbevelingen impactanalyse	29
<b>6 Bronnen en literatuur</b>	<b>31</b>
<b>7 Bijlage 1: Overzicht onderzoeksvragen</b>	<b>31</b>

## Samenvatting

Dit onderzoeksrapport maakt deel uit van het onderzoeksprogramma van het Nationaal Archief voor het project e-mailarchivering. Dit rapport is een tussenrapportage over duurzame toegankelijkheid na het uitvoeren van een impactanalyse en deskresearch voor de volgende vragen over duurzame toegankelijkheid:

- Kan het op deze wijze aangemaakte emailarchief voldoende worden ontsloten om aan de informatiebehoefte van toekomstige generaties te voldoen?
- Zijn organisaties [namelijk zorgdragers en NA] toegerust op het gebruik en beheer van dit archief?
- Welke vereisten, denk onder meer aan conservering, [zijn van toepassing uit oogpunt van duurzame toegankelijkheid] t.a.v. het realiseren van de oplossing?

Voor de impactanalyse zijn eerst normen vastgesteld voor de opname van e-mail in het e-depot. Deze normen hadden betrekking op:

- goede en geordende staat,
- metadata
- conservering
- functionaliteit
- techniek

Bij het toetsen van de normen was nog geen dataset beschikbaar om op deze basis de impact te bepalen. De impactanalyse bestond uit een tweetal workshops en enkele andere (kleinere) overleggen met vertegenwoordiging vanuit VWS, BZK, SSC IT en JenV en collega's vanuit het Nationaal Archief. De resultaten uit de ondertussen ontvangen dataset worden op een later moment gerapporteerd.

Bij de deskresearch is gebruik gemaakt van beschikbare informatie op internet, de informatie verstrekt op de briefing day van de Digital Preservation Coalition over e-mail archivering en het op handen zijnde rapport, en van een conceptversie van het rapport.

Uit de impactanalyse en de deskresearch komen vergelijkbare bevindingen. De volgende algemene bevindingen zijn het belangrijkste:

1. Samenwerking in de keten is essentieel om de toegankelijkheid van e-mail te borgen.
2. Er zijn verschillende use cases voor het beheer en de toegankelijkheid van e-mails: onderzoekers, ontwikkelaars en beheerders. Deze use cases moeten worden uitgewerkt. De voorziene duto-scan is hiervoor het goede vertrekpunt. Deze kan dan ook het beste z.s.m. gepland en uitgevoerd worden.
3. Er is onderzoek nodig naar de mogelijkheden van e-discovery en autoclassificatie software om e-mail toegankelijk te maken voor de verschillende gebruikersgroepen.
4. Probeer de eisen vanuit duurzame toegankelijkheid zo vroeg mogelijk in de keten te realiseren.
5. Plan voor de opname, beheer en beschikbaar stellen van grote volumes.

6. Leg voor sleutelfunctionarissen een register aan met contextuele informatie, ter vervanging van het verlies aan contextuele informatie bij de opname van de mailbox.
7. Herzie de standaard archivalistische ordeningsprincipes. Bekijk daarbij de gevolgen voor opname in het e-depot.

Na het uitvoeren van de DUTO scans en de analyse van de data die beschikbaar zijn, volgt een vervolgrapport.



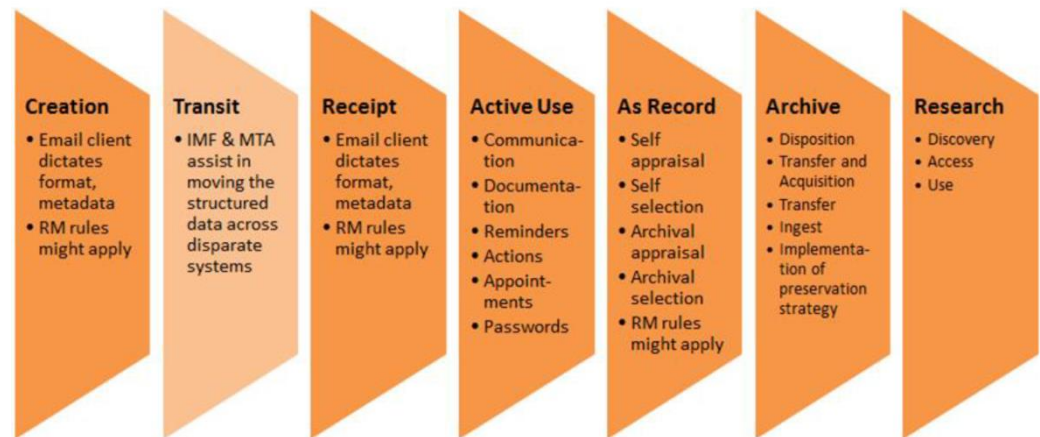
# 1 Inleiding

Dit onderzoeksrapport is opgesteld als onderdeel van onderzoekslijn 3: duurzame toegankelijkheid

Onder duurzame toegankelijkheid verstaat het Nationaal Archief het volgende:

*Informatie is vanaf het moment van ontstaan beschikbaar en bruikbaar voor iedereen die daar recht op heeft en voor zolang noodzakelijk is. Burgers, bedrijven, ambtenaren, journalisten en bijvoorbeeld onderzoekers kunnen de informatie die ze nodig hebben vinden, raadplegen en interpreteren. Niet alleen nu, maar ook over pakweg 100 jaar.<sup>3</sup>*

Het gebruik van e-mail kan hierbij worden weergegeven als een levenscyclus:



4

Duurzame toegankelijkheid kijkt terug vanuit de onderzoeksbehoefte naar het begin van de levenscyclus om te bepalen wat er nodig is in elke fase om de duurzame toegankelijkheid te borgen.

De onderzoeksvragen zijn:

- Kan het op deze wijze aangemaakte emailarchief voldoende worden ontsloten om aan de informatiebehoefte van toekomstige generaties te voldoen?
- Zijn organisaties (namelijk zorgdragers en NA) toegerust op het gebruik en beheer van dit archief?
- Welke vereisten, denk onder meer aan preservering, zijn van toepassing uit oogpunt van duurzame toegankelijkheid t.a.v. het realiseren van de oplossing?

<sup>3</sup> NORA Katern Duurzame Toegankelijkheid, [https://www.noraonline.nl/wiki/Duurzame\\_Toegankelijkheid](https://www.noraonline.nl/wiki/Duurzame_Toegankelijkheid)

<sup>4</sup> Taskforce on technical approaches to email archives

## 2 Deskresearch

Bij de deskresearch voor dit vraagstuk is gekeken naar de ontwikkelingen en gedachten in het werkveld over de archivering, preserving en beschikbaar stellen via archieven van e-mail. Een belangrijke rol hierbij speelt de "International Email Preservation Task Force".<sup>5</sup> Van hun voorlopige rapport en de informatie verstrekt tijdens een specifiek seminar is dan ook dankbaar gebruik gemaakt.

### 2.1 Volume en beheerslast

Het archiveren en toegankelijk maken en houden van e-mail is geen nieuw probleem. Het probleem is niet zozeer het formaat van e-mail, maar het heel persoonlijke gebruik van e-mail en het volume.

Volgens een onderzoek van de Radicati Group<sup>6</sup> ontvangt de gemiddelde kantoormedewerker 76 mails per dag (exc. Spam), en verstuurt de medewerker 32 mails per dag.<sup>7</sup>

Business Email	2015	2016	2017	2018	2019
<b>Average Number of Emails Sent/Received per</b>	<b>122</b>	<b>123</b>	<b>124</b>	<b>125</b>	<b>126</b>
<b>Average Number of Emails Received</b>	<b>88</b>	<b>90</b>	<b>92</b>	<b>94</b>	<b>96</b>
<i>Average Number of Legitimate Emails</i>	<i>76</i>	<i>76</i>	<i>76</i>	<i>76</i>	<i>77</i>
<i>Average Number of Spam Emails</i>	<i>12</i>	<i>14</i>	<i>16</i>	<i>18</i>	<i>19</i>
<b>Average Number of Emails Sent</b>	<b>34</b>	<b>33</b>	<b>32</b>	<b>31</b>	<b>30</b>

Table 3: Business Emails Sent/Received Per User/Day, 2015 - 2019

Voor dit project is een getal van van 1, 1 miljard e-mails op jaarbasis als uitgangspunt genomen voor de rijksoverheid. Dit komt neer op een aantal van 50 mail per werknemer per dag.<sup>8</sup> Beide aantallen zijn vele malen hoger dan het aantal nota's, rapporten e.d. dat een medewerker produceert. Realistisch gesproken betekent dit dat elke sleutelfunctionaris tussen de 10.000 en 24.000 e-mails per jaar ontvangt en verzendt. Het Nationaal Archief moet dan ook rekening houden met grote volumes. Het volume is een orde van grootte die digitaal nog niet eerder zo op archieven af is gekomen. Dit stelt andere eisen, niet alleen aan de opslagcapaciteit, maar ook aan het beheer van de informatie en het uiteindelijk beschikbaar stellen. N.B. Het NARA verwacht vanaf 2019 miljarden e-mail te moeten opnemen.<sup>9</sup>

Het beheer van grote volumes betekent dat de opslagkosten hoger worden voor het e-depot. Echter ook de capaciteit van het collectiebeheersysteem moet worden verruimd en de indexeer- en zoekcapaciteit zoals deze via de website beschikbaar wordt gesteld door de toename van de hoeveelheid informatie. Dit vertaalt zich ook naar een grotere beheerslast.

<sup>5</sup> Taskforce on Technical Approaches to Email Archives, 'The future of Email Archives'.

<sup>6</sup> <http://www.radicati.com/>

<sup>7</sup> The Radicati Group, 'Email statistics report 2015-2019'.

<sup>8</sup> Whitepaper 'Archiveren van E-mail in de Rijksoverheid'.

<sup>9</sup> Jason Baron, 'The future of Email Archiving: Four Propositions'.

Daarnaast is meer capaciteit nodig voor de opname en verwerking van deze stroom aan informatie.

De huidige vorm van beheer en beschikbaar stellen zijn daarbij mogelijk onvoldoende. De gebruiker van een archief wil meer en anders met deze informatie omgaan dan de standaard werkwijze die nu wordt aangeboden, de ontsluiting van de informatie enkel op basis van metadata.<sup>10</sup> Gezien de hoeveelheid en de inhoudelijke diversiteit ontstaat de vraag om meer en betere zoekmethoden om enkel de benodigde informatie boven water te krijgen, met zekerheid dat er geen informatie is gemist.

Er is hiervoor additionele software nodig, met de eigen beheer last, hardware benodigdheden en licentiekosten. Vooral de licentiekosten kunnen een uitdaging zijn omdat het NA, standaard voor het publiek toegankelijk is, en via het web portaal is het bereik groot. Het NA zal keuzes moeten maken over welke functionaliteit al dan niet via de website kan worden aangeboden.

## 2.2 Duurzame toegankelijkheid

Vanuit het werkveld wordt gesproken over andere werkwijzen voor de zorgvuldige opname en beschikbaar stellen van de informatie uit de mailboxen, juist door de hoeveelheid aan informatie. Zoals eerder vermeld verwacht het NARA miljarden mail om te ontsluiten. Wil het publiek hier kennis van nemen dan vraagt dat andere middelen om de informatie aan te bieden.

Dit betekent waar mogelijk het gebruik van geavanceerde technologieën, zoals "predictive coding" en "Technology Assisted Review (TAR)". Deze bieden mogelijkheden om op deze basis grote hoeveelheden informatie geautomatiseerd te doorzoeken en classificeren, niet enkel op steekwoorden, maar op concepten, op relaties en op steeds verder verfijnde voorbeeld sets, waarbij de algoritmes getraind worden en uiteindelijk kan op deze manier een grote hoeveelheid informatie met een grote mate van betrouwbaarheid geselecteerd en gepresenteerd worden<sup>11</sup>.

Zoeken is het uitsluiten van informatie die niet voldoen aan criteria die je van tevoren opgeeft (bijv. keywords) Zoektechnologie wordt complexer en slimmer, denk aan semantisch zoeken, zoeken op concepten, clustering en het inzetten van zelflerende systemen om de zoekvraag/query te verbeteren. Zoekmachines en zoektechnologieën zijn niet geschikt voor het vinden van alle relevante informatie.

Gegevensanalyse- en visualisatie software helpen met het organiseren en analyseren van grote hoeveelheden data op nieuwe en andere manieren. Je verkrijgt inzicht in de informatie. Analysetechnologieën geven de informatie weer die daadwerkelijk in de data aanwezig is (dat gebeurt via patroonherkenning, clustering van gelijksoortige informatie, visualisaties van relaties, e.d.) De informatie beschrijft zichzelf. Software ontdekt thema's, terugkerende woorden, zinnen of concepten in tekst automatisch, zonder de input van een zoekterm of een mens.

Predictive coding is een proces waarbij supervised machine learning-technologie wordt ingezet. Het systeem voorspelt hoe documenten moeten worden geclassificeerd op basis van beslissingen die aan het systeem zijn gevoed. Machine learning-technologie is niet nieuw en wordt ook toegepast binnen de hierboven beschreven technologieën. Het verschil is dat het systeem niet getraind wordt aan

<sup>10</sup> James Baker, 'Using Email archives in research'.

<sup>11</sup> Larry Chapin en Simon Atfield, 'The reconstruction of narrative in e-discovery investigation'.

de hand van woorden, patronen, analyse technieken en/of monitoring van gedrag, maar dat het systeem een menselijke beslissing probeert na te bootsen.

Deze technieken zijn bekend vanuit de e-discovery wereld en daar al in gebruik, maar zijn nog niet ingesteld op de eisen binnen de archiefwereld, bijv. voor het bepalen van de openbaarheid van e-mail. Er is onderzoek en tijd nodig om deze methodieken ook hiervoor geschikt te maken.<sup>12</sup>

Binnen het Nationaal Archief loopt op een onderzoek naar automatische classificatie van e-mails. De resultaten daarvan en de verdere ontwikkelingen van dit onderzoek zijn voor het toekomstig beheer en de toegankelijkheid van e-mail van groot belang.

<sup>12</sup> Tim Gollins, 'Technically Assisted Review: Fact, Fiction or Jam tomorrow.'

## 3 Resultaten impactanalyse

Met een impactanalyse kijkt het Nationaal Archief in welke mate uit te plaatsen informatieobjecten voldoen aan onze normen. Deze normen hebben betrekking op het uitplaatsen of overbrengen van digitale informatie naar het e-depot. Ook kijken we welke maatregelen genomen moeten worden om de normen te bereiken. Voor de (fictieve) opname van de mailboxen van sleutelfunctionarissen zijn de bestaande normen voor de opname van digitaal archief tegen het licht gehouden in een workshop binnen het NA en hier en daar aangepast voor de opname van e-mail.

In een tweetal workshops is samen met de pilot departementen gekeken in hoeverre de mailboxen en e-mail nu voldeed aan deze normen. Omdat er nog geen sprake is van daadwerkelijke overbrenging van e-mail betekenen afwijkingen van de norm niet dat direct maatregelen moeten worden genomen, wel dat er aandacht moet komen voor maatregelen.

Bij het vaststellen van de normering is uitgegaan van opname in het e-depot vanuit een exchange (outlook) omgeving.

Bij de resultaten van de impact analyse is er vaak een keuze voor de oplossing door het NA of de ministeries. De Archiefregeling legt hierbij de verantwoordelijkheid voor het in goede en geordende staat bewaren en daarmee ook aanbieden bij de zorgdrager. Het is vooral belangrijk hierin samen op te trekken.

### 3.1 Normen voor opname e-mail

De normen die zijn gebruikt bij de impactanalyse zijn geordend per deel gebied:

- Goede en geordende staat
- Metadata
- Preservation
- Functionele voorwaarden
- Technische voorwaarden

### 3.2 Impactanalyse goede, geordende en toegankelijke staat

In de analyse voor de goede, geordende en toegankelijke staat zijn onderstaande bevindingen en de mogelijke impact daarvan beschreven. Waar mogelijk zijn scenario's voor het oplossen van de bevinding vermeld; in sommige gevallen zijn er zelfs meerdere scenario's denkbaar. Een advies voor welke zaken opgepakt kunnen worden is te vinden in het hoofdstuk conclusies en aanbevelingen.

<b>Gestelde normen</b>	<b>Bevindingen t.o.v. de norm</b>	<b>Scenario's</b>	<b>Wat is de impact</b>
Rol en functie sleutelpersoon.	De beschrijving van de rol en functie van de sleutelfunctionaris, voor de periode waarvan de mail wordt overgedragen, leveren belangrijke contextinformatie voor het duiden van de mail voor de onderzoeker. Dit vereist het vastleggen van deze informatie bij de ministeries. Op dit moment wordt dit waarschijnlijk niet vastgelegd. Het kan zijn dat deze informatie wel herleidbaar is, dit moet nog worden onderzocht.	Het bijhouden van een register van sleutelfunctionarissen.	Additionele werkzaamheden voor het opzetten en bijhouden van registers bij de ministeries.
Werkprocessen (waar mogelijk).	Indien mogelijk moet ook een link met werkprocessen en selectielijst als context op de mailbox gelegd worden. Op dit moment kan dit niet.	<ul style="list-style-type: none"> <li>• Beschrijving met vermelding welke dossiers in een DMS onder de verantwoordelijkheid van de sleutelfunctionaris vielen.</li> <li>• Gebruik van e-discovery software, vergelijkbaar met de huidige zoek &amp; vind die in staat is om de informatie te koppelen.</li> </ul>	Onderzoek noodzakelijk naar de haalbaarheid van deze scenario's.

Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
<p>Mail heeft een titel/subject.</p>	<ul style="list-style-type: none"> <li>• E-mail heeft niet altijd een titel/subject; Voor het big data platform maakt de onderwerpsregel voor het zoeken niet uit en daarmee wordt dit door de ministeries niet als een probleem ervaren. Voor de goede, geordende en toegankelijke staat ziet het NA dit wel als een probleem. Denk bijvoorbeeld aan het groeperen van de e-mail. Een duidelijke onderwerpregel maakt e-mail bovendien herleidbaar tot de context, wat het een potentieel zoekveld maakt.</li> <li>• Het vullen van de onderwerpregel is niet verplicht voor e-mail, wel volgt er een waarschuwing als deze niet is ingevuld.</li> <li>• De kwaliteit van de onderwerpsregel is verschillend. Het ontbreken of gevuld met beperkte kwaliteit van de onderwerpregel heeft impact op de toegankelijkheid van de e-mail.</li> </ul>	<ul style="list-style-type: none"> <li>• Medewerkers meer bewust moeten worden gemaakt wat het belang is van het vullen van deze onderwerpregel.</li> <li>• Opname in de richtlijn e-mail archivering.</li> </ul>	<ul style="list-style-type: none"> <li>• Ontwikkelen van een bewustwording-campagne rondom e-mail archivering. Hierbij extra nadruk op sleutelfunctionarissen.</li> <li>• Opname van tekst in de richtlijn e-mail archivering kan tot discussie leiden.</li> </ul>

Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
Structuur van de mailbox wordt behouden.	<ul style="list-style-type: none"> <li>• Belang van de structuur van de mailbox wordt onderstreept door alle gesprekspartners als contextinformatie.</li> <li>• Het gaat om weergave en informatiebehoud. Het hoeft niet als de standaard structuur in het e-depot terug te komen.</li> <li>• "Platte" opslag met een index is voor het proces van automatische classificatie de beste oplossing.</li> </ul>	Vastleggen bij de e-mail van het "folder pad" als metadata.	<ul style="list-style-type: none"> <li>• Er dient verder onderzoek gedaan te worden naar de mogelijkheden om e-mail op te slaan.</li> <li>• De vorm waarin de structuur wordt vastgelegd om deze te kunnen reproduceren moet onderzocht worden.</li> </ul>
E-mailbox sleutelpersoon geldt als een serie	De e-mailbox van een sleutelfunctionaris, gekoppeld aan de persoon geldt als aggregatie niveau; serie. De periode kan als aggregatieniveau "dossier" of subserie gezien worden.	Subserie is duidelijker dan dossier omdat het enkel een specifieke periode betreft van correspondentie en geen relatie kent met een specifiek onderwerp. Dit komt dan eerder voort uit folders in de mailbox	Aanpassing noodzakelijk in het e-depot van de huidige aggregatieniveaus.
Verzend-/ ontvangstdatum van e-mail beschikbaar	Er is altijd een verzend- of ontvangstdatum. Dit is metadata en kan als zodanig ook meegenomen worden.		



Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
<p>E-mails bevatten geen links naar documenten of DMS-objecten maar naar de daadwerkelijke bestanden.</p>	<p>Uitgangspunt is dat een e-mailbox een aparte archiefeenheid is. Er is hierdoor een discrepantie tussen het moment van overbrenging van de archiefbescheiden uit een DMS en dat van de e-mail uit de e-mailboxen. Om betrouwbare informatie te kunnen garanderen is wel het originele bestand nodig. De mails dienen daarom geen links naar documenten of DMS-objecten te bevatten, maar de daadwerkelijke bestanden, want:</p> <ul style="list-style-type: none"> <li>▪ Een link verwijst altijd naar de recentste versie, terwijl juist de versie van verwijzing op dat moment belangrijk is.</li> <li>▪ Een link is na overbrenging ook niet meer valide, omdat de DMS bijvoorbeeld niet meer bestaat, en de e-mail is losgeknipt.</li> <li>▪ De bijlagen kunnen bestaan uit documenten die vernietigd zijn volgens de geldende selectielijsten.</li> </ul>	<ul style="list-style-type: none"> <li>• Bij overbrenging moet er een (kopie van het) bestand mee. Mogelijke oplossing is om de link naar de DMS reconstrueerbaar te maken.</li> <li>• Via e-discovery ontsluiten van beide bronnen en samenstellen van 1 dossier. Hierbij valt mogelijk het versieverschil niet te voorkomen.</li> </ul>	<ul style="list-style-type: none"> <li>• Het in kaart brengen van de risico's als de links verloren gaan.</li> <li>• Nader onderzoek naar opties om e-mails en bijlagen bij elkaar te brengen noodzakelijk.</li> </ul>
<p>De mail moet authentiek zijn; is het mailbestand (en evt de bijlage) een correcte weergave van de communicatie op dat moment? (Bijv. geen wijzigingen etc. na versturen of ontvangen.)</p>	<p>De e-mail en bijlage moeten een correcte weergave zijn van de communicatie op dat moment. Er moeten garanties zijn dat wijzigingen na de tien weken niet meer toegepast kunnen worden.</p>	<p>Een optie die nader onderzocht moet worden is het toepassen van een <i>litigation hold</i> op een mailbox. Deze zorgt ervoor dat het niet mogelijk is om e-mails te wijzigen of verwijderen.</p>	<p>De <i>litigation hold</i> maakt geen onderscheid in tijd, maar enkel in de mailbox. Als deze erop staat wordt dus ook de eerste 10 weken niets verwijderd. Of de medewerkers hier moeite mee hebben moet onderzocht worden.</p>

Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
Relatie tussen e-mail in een string zichtbaar.	Threadindex als metadata verstrekt deze informatie.	Opname bij metadataveld referentie.	Metadata moeten worden opgenomen ( zie ook metadata).
Alle e-mail wordt individueel overgedragen.	Standaard export van e-mail is als .pst, dus een samenvoeging van alle e-mail. Deze kan wel worden uitgepakt	Zie functionele en technische eisen voor de scenario's.	
Relatie van e-mail met bijlage(n) blijft behouden	Als de mail als mail wordt overgedragen is dit geen probleem.		
Periode persoon rol sleutelfunctionaris bekend	<ul style="list-style-type: none"> <li>• De sleutel-functionaris-markering moet afhankelijk zijn van rol, niet van persoon. Dit betekent ook dat bijgehouden moet worden wie wanneer welke sleutelfunctionaris rol vervult. De selectiecriteria voor de sleutel-functionaris worden in de richtlijn besproken. Er kan een sleutelrol voorkomen op lagere schalen dan schaal 16, maar dat zal nog getoetst worden.</li> <li>• Het is handig als een persoonlijk account bij export losgeknipt wordt van de sleutel-functie. Bij wijziging van functie moet er binnen de organisatie zelf een signaal komen.</li> </ul>	Ontwikkelen van een signaal functie zodra medewerkers als sleutelfunctionaris worden aangewezen, als hun functie wijzigt en als ze geen sleutel-functionaris meer zijn.	Uitwerking nodig van de signaleringsfunctie binnen de ministeries.
Periode overgedragen e-mail bestanden (van.. tot..) beschikbaar	De overdracht gaat met jaarlijkse batches.		Planning van overdracht noodzakelijk zodat niet alles gelijk wordt aangeboden bij het NA.

Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
Selectie periode van de sleutelfunctionaris is mogelijk	De start- en einddatum van de verbinding van een persoon aan een sleutelfunctie moet bekend zijn.	Ook dit vraagt een registratie van de sleutelfunctionarissen door de ministeries met daaraan gekoppeld de naam van de betreffende functionaris en de periode.	Uitwerken van de registratie eisen en processen bij de ministeries.
	Afhankelijk van de gebruikte oplossing. Op dit moment kan binnen het e-depot wel al full-tekst, inclusief bijlagen worden gezocht. Dit kan ook via de zoek & vind oplossing.	Inrichten van een nieuwe oplossing voor het zoeken van informatie door gebruikers / onderzoekers bij het Nationaal Archief	Onderzoek naar mogelijkheden voor gebruikers voor zoekoplossingen is noodzakelijk.
De openbaarheid moet na overbrenging niet of zo kort mogelijk beperkt zijn.	E-mail bevat mogelijk privacygevoelige gegevens of informatie die anderszins niet gelijk openbaar kan worden. Deze e-mail is niet apart herkenbaar.	<ul style="list-style-type: none"> <li>• Met behulp van e-discovery het classificeren van de mail die beperkt openbaar is.</li> <li>• Mailbox als geheel beperkt openbaar.</li> </ul>	<ul style="list-style-type: none"> <li>• Onderzoek gewenst naar mogelijkheden e-discovery voor vaststellen openbaarheidsbeperkingen.</li> <li>• E-mail voor langere tijd beperkt openbaar, onafhankelijk van de inhoud.</li> </ul>

### 3.3

#### Impactanalyse metadata

Archieven worden beschikbaar gesteld aan het publiek. In verband met beheer en dienstverlening naar het publiek dient er een set metadata beschikbaar te zijn. Deze metadata wordt niet door NA-medewerkers toegevoegd aan alle documenten, maar is afkomstig van de informatie zelf. Het NA kent voor het e-depot een xml-standaard, gebaseerd op de metadata standaarden van het Rijk en van de lokale overheden. Bij de invoering van DMS systemen kon rekening gehouden met de eerder genoemde metadata standaarden en aan de xml-standaard voor de opname in het e-depot worden voldaan.

E-mail kent ook metadata, maar is, als standaard, internationaal systeem niet op dezelfde wijze ingericht. Voor de impactanalyse is de normering, nl het bezitten van de vereiste metadata zoals verwacht in het e-depot (TOPx)<sup>13</sup> tegen het licht gehouden en nog enigszins versimpeld tot de absoluut noodzakelijke velden. Op basis van deze set zijn de mailbox, de folders en de e-mailmessage bekeken om te zien of en hoe deze konden voldoen aan -in ieder geval deze- vereiste metadata. Als er additionele metadata beschikbaar waren is ook gepoogd deze een plaats te geven, waarbij het niet aanwezig zijn van de metadata geen belemmering is.

De resultaten van deze mapping zijn te vinden in het bijgevoegde mappingsheet. Hierin is alle informatie bijeengebracht over de aanwezige en niet aanwezige, maar

<sup>13</sup> TOPx is gebaseerd op het Toepassingsprofiel Rijk en het Toepassingsprofiel Lagere Overheden. Dit is een 1-entiteitsmodel.

wel verplichte metadata op alle aggregatieniveaus. Deze details zijn dan ook niet hier opgenomen.

Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
Alle verplichte metadata voor de serie zijnaanwezig.	<ul style="list-style-type: none"> <li>• De mailbox van een persoon geldt als serie.</li> <li>• De functie/rol van de persoon achter een e-mailadres zit niet in exchange, maar moet uit een medewerkersregister of iets dergelijks vandaan gehaald worden.</li> <li>• Elke mailbox kent een unieke identificatie.</li> </ul>		Zonder de koppeling van de functie/ rol met de medewerker is de metadata niet compleet en kan niet worden opgenomen.
Alle metadata voor het dossier zijn aanwezig.	<ul style="list-style-type: none"> <li>• Er zijn geen typerende metadata voor een dossier, zoals begin- en einddatum beschikbaar in de mailbox.</li> <li>• Er is op en in de mailbox veel additionele metadata aanwezig die nuttig kan zijn, maar niet past in het huidige schema.</li> </ul>	Een representatie van een dossier kan alleen de overdrachtsperiode zijn met van te voren bepaalde invulling van de metadata.	<ul style="list-style-type: none"> <li>• Dossiers zijn niet gerelateerd aan activiteiten en zaken. Als dit gewenst is moet dit door andere mogelijkheden worden uitgevoerd.</li> <li>• Er moet gekeken worden naar een volledig aangepast en specifiek schema voor metadata, gebaseerd op beschikbare informatie</li> </ul>

Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
Alle verplichte metadata voor het record zijn aanwezig.	<p>Aanname is dat één e-mail ook één archiefstuk is.</p> <ul style="list-style-type: none"> <li>Er is geen informatie over openbaarheid van een e-mail: die metadata ontbreken.</li> <li>Informatie over de vertrouwelijkheid is niet aanwezig.</li> <li>Elke mail kent een uniek identificatienummer.</li> <li>De verzend-/ ontvangstdatum van e-mail is beschikbaar.</li> <li>Verzender &amp; ontvanger(s) gelden als actoren. Deze kunnen vaak herhaald worden.</li> <li>De thread-index is een belangrijk veld om mee te nemen. Dit refereert alle mails in een conversatie.</li> <li>Datum dat mappen zijn aangemaakt is beschikbaar.</li> <li>Verwachting is dat een onderscheid gemaakt kan worden tussen verzender of ontvanger. Dit past niet in de huidige opzet.</li> <li>Er is veel eventinformatie beschikbaar.</li> </ul>	De id moet in de header mee met de export.	<ul style="list-style-type: none"> <li>Er moet gekeken worden naar de opname van verzender en ontvanger als actoren om het onderscheid wel te behouden.</li> <li>Er moet in nader overleg bepaald worden welke standaard metadata nog van waarde is. Dit betreft met name event informatie, zoals viruschecks bij ontvangst e.d.</li> </ul>
Alle verplichte metadata voor het bestand zijn aanwezig.	Er is een verschil tussen e-mailbestand en bijlage. Zolang de bijlage deel uitmaakt van de e-mail als geheel is dit geen probleem.		

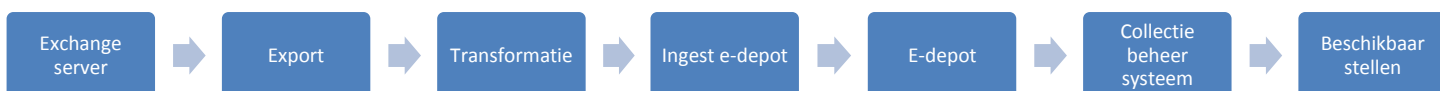
### 3.4 Impactanalyse preservering

Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
Betreft het informatiesoorten met extra impact op preservatie (zoals databases, vector-images)?	Nee, allemaal e-mails.		Geen.
Wijken bestandsformaten af van de NA-voorkeursformaten?	Enkel .pst op dot moment mogelijk als export. (al brengt pst wel de nodige problemen met zich mee – bijvoorbeeld maximale omvang - en wordt dit niet meer door Microsoft ondersteund).		Onderzoeken of voorkeursformaat EML kan worden aangeleverd.

Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
Bevat de set bestanden met veel of ongebruikelijke (interactieve) gedragselementen?	E-mails an sich zijn statisch (zijn door een aantal 'wasstraten' gehaald, bijvoorbeeld tegen SPAM en virussen). Bijlagen kan van alles zijn.		Nader onderzoek nodig. Risico is aanwezig.
Bevat de set versleutelde bestanden (encryptie)? NB: Encryptie: niet, tenzij noodzakelijk en dan krijgt het NA de decryptiesleutel en/of het wachtwoord.	E-mails van binnen naar buiten en vice versa worden gecontroleerd op versleutelde bestanden (deze worden geblok). Maar bijlagen bij e-mail kunnen ge-encrypt zijn. En sleutels zijn lang niet altijd beschikbaar. Hoogstwaarschijnlijk zijn er wel pdf's waar beperkingen aan zitten (printen, kopiëren). Niet met zekerheid te zeggen.  Er worden sporadisch versleutelde e-mails verstuurd op topniveau		Nader onderzoek is nodig hoe vaak versleuteling of wachtwoorden voorkomen. Het Nationaal Archief kan deze e-mail niet opnemen in een e-depot.
Bevat de set digitaal ondertekende bestanden waarbij rechtsgeldigheid van de handtekening relevant is? NB: Digitale handtekening: niet, tenzij (bijvoorbeeld juridisch) noodzakelijk.	Niet met zekerheid bekend. E-mails sowieso niet.		
Vindt migratie / conversie (Bijvoorbeeld van MS Office documenten naar PDF/A) in bronomgeving plaats, zo ja, (hoe) wordt kwaliteit gemeten?	Bij overgang van IND naar J&V zijn mails geëxporteerd (uit Lotus Notes systeem o.i.d.) naar Exchange. Gehanteerde kwaliteitscriteria zijn niet bekend.		Een ministerie dient in de toekomst de migratie activiteiten vast te leggen, incl. alle uitkomsten. De uitkomst moet bij overdracht aan het NA worden meegeleverd.

### 3.5 Impactanalyse functionele en technische eisen

Voor de functionele en technische impact is gekeken naar de mogelijkheden om van de exchange server waar de mail bewaard wordt te kunnen komen. De basis vormt het volgende versimpelde procesoverzicht:



### 3.5.1 Exchange en export

Vanuit dit overzicht is eerst ingezoomd op de exportmogelijkheden vanuit exchange:

Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
Er moet vooraf duidelijk zijn wat de inhoud is van een overdracht.	<ul style="list-style-type: none"> <li>Overdracht van alle mail, behalve uit de folder privé. Dit is te realiseren.</li> <li>De mailbox bevat alle e-mail, dit is inclusief de agenda items. Er is geen onderscheid in formaat of anderszins tussen agenda-items en overige e-mail.</li> <li>Als de <i>litigation hold</i> aanstaat komen alle persoonlijke e-mails ook mee.</li> </ul>	<ul style="list-style-type: none"> <li>Selecteren van de over te dragen mail via rapportages in exchange</li> <li>Overdracht via een exporttool waarin selectie mogelijk is.</li> </ul>	<ul style="list-style-type: none"> <li>Als de agenda items meekomen in de export kan de agenda gereconstrueerd worden.</li> <li>Onderzoek noodzakelijk naar gebruik van <i>litigation hold</i> met alle voor- en nadelen.</li> <li>Privémail kan in het e-depot terecht komen. De gevolgen voor openbaarheid en vanuit de AVG moeten onderzocht worden.</li> </ul>
De tijd van de overdracht moet bepaald worden. 20 jaar na welk moment?	Overdracht per jaar.		Jaarlijkse overdrachten zijn mogelijk maar breken wel in op een e-mail thread.
De export bestaat uit de authentieke e-mail.	De authentieke mail staat op de exchange server. Export van de mail zoals deze op de exchange server staat vereist het aanzetten van de <i>litigation hold</i>	Aanzetten <i>litigation hold</i> in exchange. Vernietiging van e-mails pas na export en voor overbrenging	Onderzoek nodig naar gebruik van <i>litigation hold</i> met alle voor en nadelen.
Overbrenging voldoet aan exportvoorwaarden omtrent bestandsformaten van het NA.	Het NA wil graag de mail in .eml-formaat ontvangen. Standaard is een .pst het exportformaat, waarbij wel werd aangegeven dat deze niet meer door MS wordt ondersteund.		
Bijlagen zijn bestanden en geen links.	Grote variëteit, waarbij steeds vaker links gestuurd worden in plaats van bijlagen.	Capture van de site of het bestand noodzakelijk voor een overdracht.	Of dit mogelijk is moet worden onderzocht.

Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
Metadata zijn aanwezig.	<ul style="list-style-type: none"> <li>• Metadata zitten in de header van een e-mail.</li> <li>• Alle folders kennen dezelfde metadata. Er is geen verschil tussen inbox en onderliggende folders.</li> </ul>		Metadata moeten uit de header worden ingelezen.
Alle verplichte metadata velden zijn gevuld met een waarde.	Er zijn geen verplichte velden en dus zullen ook velden als subject af en toe leeg zijn of met nietszeggende titels.	Door een exporttool laten invullen van een fictieve waarde bij lege velden.	Zonder vulling van verplichte velden worden e-mailbestanden niet opgenomen in het e-depot. Er is een risico op uitval.
Gebruik van "verboden" tekens zoals leestekens.	In de subjectheading van mail kunnen leestekens voorkomen.		Er is onderzoek nodig of deze beperking in het e-depot kan worden opgeheven.

### Transformatie

Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
Gebruik van digikoppeling voor transport <ul style="list-style-type: none"> <li>○ Standaard locatie mail klaarzetten.</li> <li>○ Automatisering van opname.</li> </ul>	Digikoppeling is bekend en kan ingezet worden.		Aansluiten met digikoppeling.
Formaat van aanlevering is een .TAR.	Standaard wordt vanuit exchange enkel een .pst geleverd. Onduidelijk is of een .pst bestand ook gelimiteerd kan worden voor het aantal bestanden, mogelijk wel in de grootte van het bestand.	<ul style="list-style-type: none"> <li>• Aanpassen e-depot zodat .pst bestanden opgenomen en uitgepakt kunnen worden.</li> <li>• Voor ingest in het e-depot een nieuw (standaard) exporttool inzetten.</li> </ul>	<ul style="list-style-type: none"> <li>• Bouw nieuwe workflow bij het Nationaal Archief, of</li> <li>• Bouw standaard exporttool door SSC - ICT (namens de ministeries) die wel een .TAR - bestand kan leveren.</li> </ul>



Gestelde normen	Bevindingen t.o.v. de norm	Scenario's	Wat is de impact
Er mogen max. 15.000 bestanden in een .TAR geleverd worden.	Een mailbox kan uit meer dan 15.000 items (e-mails bestaan, De export uit exchange in .pst neemt standaard alle e-mail mee, zonder beperking. Deze kan niet zo worden aangeboden aan het e-depot.	<ul style="list-style-type: none"> <li>• Inzetten exporttool.</li> <li>• Selecteren van aantallen voor een export van een .pst bestand.</li> </ul>	<ul style="list-style-type: none"> <li>• Bouw standaard exporttool.</li> <li>• Onderzoek selecteer mogelijkheden in exchange.</li> </ul>

#### *Ingest e-depot*

Het transporteren van het te leveren bestand dient plaats te vinden via, idealiter, de digikoppeling.

Dit lijkt geen probleem te zijn om te realiseren aangezien deze al eerder ook is ontwikkeld.

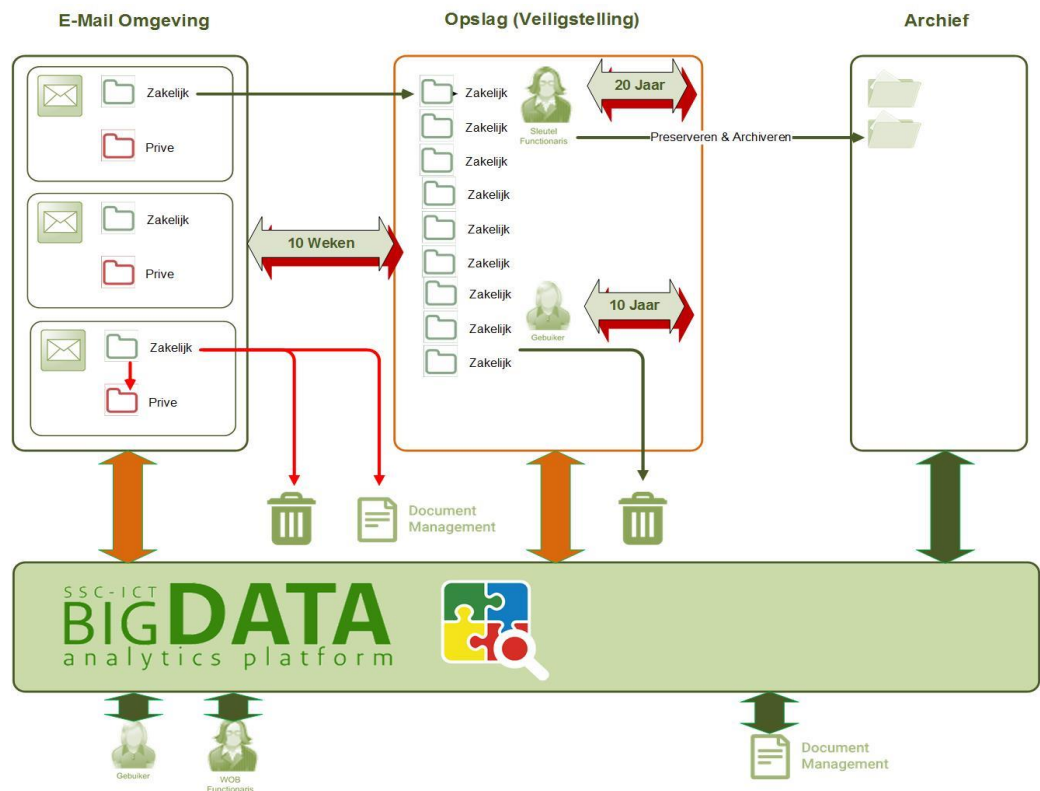
#### *E-depot*

De vereiste ordening voor opname in het e-depot van archief – serie- dossier – record - bestand kent nog wel een paar vraagtekens.

- Er moet gekeken worden naar de bijlagen. Deze zijn als ze enkel op bestandsniveau worden meegestuurd niet direct opvraagbaar en zichtbaar voor klanten van het NA. Als ze ook als record worden meegestuurd dan moet de relatie tot de e-mail goed worden beschreven.
- De mailbox is voorzien als serie niveau. Het gebruik van de verschillende aggregatie niveaus moet worden aangepast. Hierbij moet rekening worden gehouden met een opname van de structuur van e-mailboxen zodat deze bij elke overbrenging herhaalbaar is. Als de mappenstructuur direct vanuit gekopieerd wordt met de huidige structuur moet bij elke aanlevering opnieuw gemapt worden om het juiste dossierniveau te vinden.
- Er is onderzoek nodig hoe namen van personen kunnen worden geanonimiseerd (lakken).
- Er dient bepaald te worden hoe wordt omgegaan met een verandering van de persoon die de rol van sleutelfunctionaris vervuld. Is dit een nieuwe serie of wordt doorgeslagen met de bestaande?

## 4 Ontwikkelingen architectuur

Voor de oplossing om de e-mail langdurig te bewaren voor gebruik bij WOB-verzoeken en andere onderzoeken binnen de overheid en de overdracht aan het NA van de mail van sleutelfunctionarissen dient een voorziening te worden opgezet. Op hoofdlijnen was in het project de volgende architectuur voorzien:



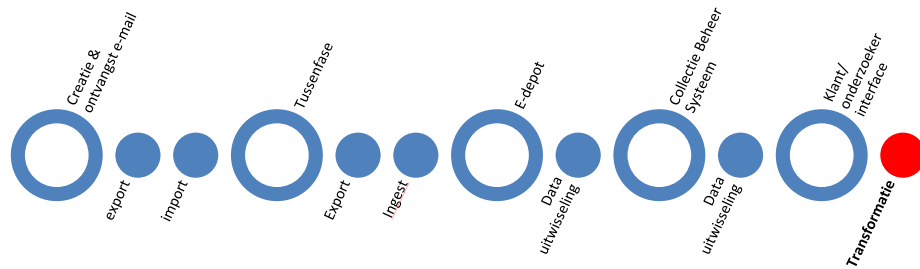
Tijdens het project is verder gekeken naar de invulling. Als NA is de keten nog niet volledig bij ontvangst van de archiefstukken, maar is deze pas volledig bij het daadwerkelijk aanbieden van de stukken. In deze keten transformeert de metadata van de archiefstukken om te voldoen aan de eisen voor toegankelijkheid, zoals deze zijn bepaald in het toepassingsprofiel rijk en de internationale standaarden op archiefgebied (m.n. EAD). In de huidige keten voor digitale informatie is de belangrijkste transformatie van de metadata het moment van overbrenging vanuit een DMS naar het e-depot. Er is echter niet gezegd dat de keten voor de overbrenging van e-mail precies hetzelfde moet verlopen. (N.B. voor het uitvoeren van de impact analyse is dat wel als uitgangspunt genomen)

### 4.1 Transformatie als keuze in de keten

In onderstaande opties voor dit ketenproces is het moment van transformatie verschoven. Tijdens de impactanalyse zijn deze opties besproken om een eerste voorkeur te bepalen. Er is verder onderzoek nodig naar de haalbaarheid van deze opties.

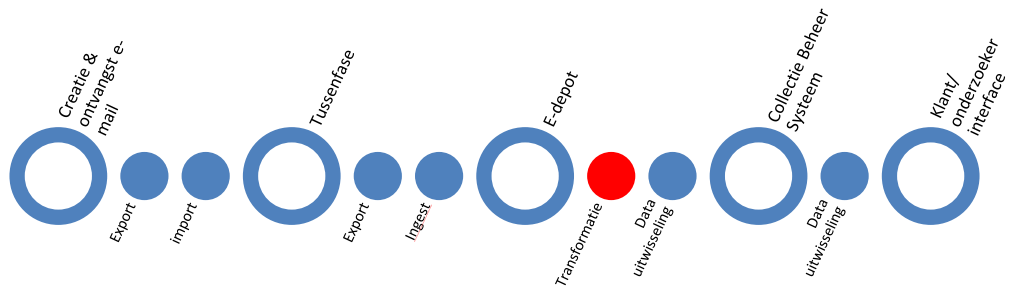
*Transformatie door de gebruiker*

Als eerste optie vindt de transformatie pas plaats aan het einde van de keten. De onderzoeker wordt hierbij de instrumenten (software) gegeven om zelf uit de brij van informatie een logisch geheel te maken. In de hele keten wordt, om dat doel te bereiken, geen enkele informatie aangepast. Het grote voordeel van de optie is dat de onderzoeker vrijheid krijgt en er geen aanpassingen nodig zijn. Het NA moet echter hiervoor wel investeren in de goede software en in de mogelijkheid om onderzoekers op te leiden in het gebruik hiervan. Om in de keten geen transformatie te eisen zijn ook aanpassingen nodig. Op dit moment kan deze optie nog niet worden aangeboden.



*Transformatie door het Nationaal Archief*

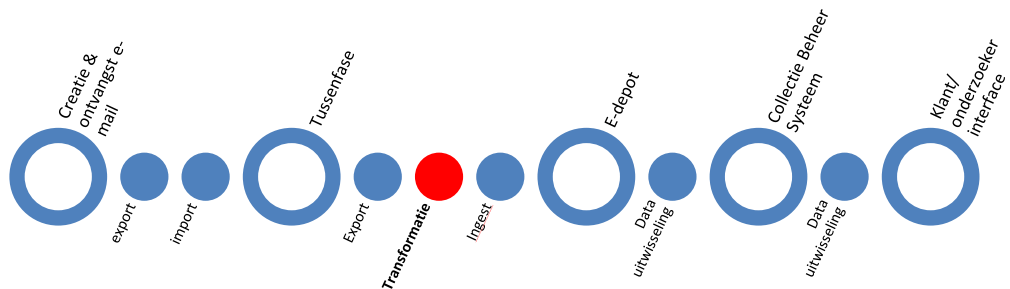
De transformatie kan ook eerder in het proces liggen, zoals na opname in het e-depot:



Hierbij moet het NA de handelingen verrichten om de mail toegankelijk te maken voor de onderzoeker. De zorgdrager dient enkel aan te leveren.

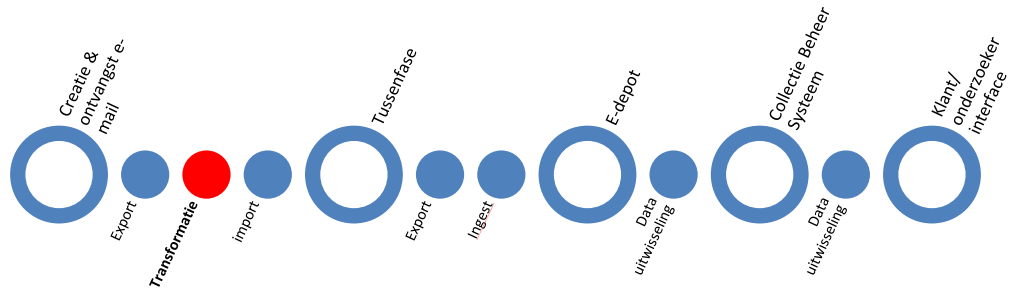
*Transformatie door de zorgdrager*

Als NA verplichten we ook nu al de zorgdrager om bij overbrenging deze vorm te geven volgens onze exportvoorwaarden. Dit kan uiteraard ook bij e-mail. Hiervoor moet dan een export tool worden gebouwd. Deze is uiteraard wel herbruikbaar.



*Transformatie aan de bron*

Vanuit de filosofie van archivering bij de bron is het ook mogelijk om nog eerder in het proces al rekening te houden met de eisen rondom duurzame toegankelijkheid en de daarbij behorende opname in een e-depot. En dat is in deze keten op het moment dat de e-mail na 10 weken naar een tussenoplossing wordt gemigreerd. Vanuit het zicht van het NA is dit mogelijk de beste oplossing, waarbij de inspanning volledig bij de zorgdrager zal plaatsvinden.



## 5 Conclusie en aanbevelingen

De volgende aanbevelingen komen uit de deskresearch en de impactanalyse:

1. Werk samen in de keten om de toegankelijkheid van e-mail te borgen.
2. Er zijn verschillende use cases voor het beheer en de toegankelijkheid van e-mails: onderzoekers, ontwikkelaars en beheerders. Deze use cases moeten worden uitgewerkt. De voorziene duto-scan is hiervoor het goede vertrekpunt. Deze kan dan ook het beste z.s.m. gepland en uitgevoerd worden.
3. Er is onderzoek nodig naar de mogelijkheden van e-discovery en autoclassificatie software om e-mail toegankelijk te maken voor de verschillende gebruikersgroepen.
4. Probeer de eisen vanuit duurzame toegankelijkheid zo vroeg mogelijk in de keten te realiseren.
5. Plan voor de opname, beheer en beschikbaar stellen van grote volumes.
6. Leg voor sleutelfunctionarissen een register aan met contextuele informatie, ter vervanging van het verlies aan contextuele informatie bij de opname van de mailbox.
7. Herzie de standaard archivalistische ordeningsprincipes. Bekijk daarbij de gevolgen voor opname in het e-depot.

### 5.1 Aanbevelingen impactanalyse

In de impactanalyse zijn een aantal afwijkingen van de normen geconstateerd. Voor opname in het e-depot moeten deze afwijkingen worden opgelost. De mogelijkheden om deze op te lossen liggen bij de archiefvormer of het NA. Er zullen in een aantal gevallen keuzes moeten worden gemaakt over wie de oplossing gaat realiseren.

De volgende aanbevelingen zijn uit de impactanalyse voortgekomen:

- Onderzoek de mogelijkheden om mail van sleutelfunctionarissen op een andere wijze in het e-depot op te nemen.  
Outlook stelt medewerkers in staat om een eigen folderstructuur te maken. Deze structuur moet ook beschikbaar zijn voor het publiek. Dit past niet in de huidige inrichting van het e-depot. Voor het e-depot geldt op dit moment dat er een bepaalde structuur is waarbij het laagste aggregatie niveau van de mailbox zal moeten dienen als dossier. Een dossier is echter niet herkenbaar in een mailbox. Als voor e-mail bewust deze indeling en inrichting wordt losgelaten dan wordt recht gedaan aan de bijzondere vorm van de mailbox.
- Laat bij e-mail de dossier-gedachte los. Zie e-mail als een persoonlijke verzameling van correspondentie, met een persoonlijke ordening binnen de standaard gegeven door de mailbox structuur.
- Onderzoek hoe de rol- en functiebeschrijving van sleutelfunctionarissen kan worden vastgehouden door de ministeries. Dit is essentiële contextinformatie voor het publiek om de waarde van de e-mail informatie te kunnen inschatten. De rol moet worden gekoppeld aan de persoon die deze functie bekleedt. Het ministerie dient ook de periode dat iemand een functie bekleedt vast te houden. Bij overdracht levert het ministerie deze informatie aan het Nationaal Archief.
- Onderzoek de oplossingsrichtingen voor de geconstateerde afwijking van de norm op het gebied van aanlevering van een .tar-bestand:

- De zorgdrager ontwikkelt een export tool om te voldoen aan de vereiste levering. Dit is in lijn met de wijze waarop vanuit een DMS wordt aangeleverd.
  - Het NA ontwikkelt het e-depot om .pst-bestanden te kunnen opnemen en uitpakken en voor de verdere verwerking in het collectiebeheersysteem.
- Ga uit van een jaarlijkse overdracht van de e-mails uit een mailbox.
- De compleetheid en kwaliteit van de metadata, m.n. subject, wisselt. Het verdient aanbeveling om medewerkers bewust te maken van het belang om het subject goed te gebruiken om altijd een duidelijk onderwerp mee te geven.
- Doe verder onderzoek naar mogelijkheden om de informatie van de mailbox aan te bieden aan de gebruiker.
- Ontwikkel een nieuw metadata schema (analoog aan TopX) voor e-mail.
- Doe onderzoek naar een mogelijke koppeling met het Actorenregister voor het bijhouden en de beschrijving van de sleutelfunctionarissen.
- Onderzoek de mogelijkheden om met autotaxonomie en e-discovery mogelijkheden metadata, dossierkoppelingen en eventuele beperkingen op de openbaarheid te realiseren.
- Onderzoek hoe e-mails als .eml-bestanden kunnen worden geleverd.

## 6 Bronnen en literatuur

James Baker. "Using E-mail Archives in Research".

Jason Baron. "The Future of Email Archiving: Four Propositions". Januari 2018.

Larry Chapin en Simon Atfield, "The Reconstruction of Narrative in E-Discovery Investigations". 2018.

Erfgoedinspectie. "Onvoltooid Digitaal". Den Haag, oktober 2015.

Tim Gollins. "Technically Assisted Review: Fact, Fiction of Jam Tomorrow?". Juni 2017.

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. "Archiveren van E-mail in de Rijksoverheid". Den Haag, november 2016.

Taskforce on technical approaches to email archives, "The future of Email Archives". 2018.

The Radicati Group. "E-mail Statistics Report 2015-2019".

*[https://www.noraonline.nl/wiki/Duurzame\\_Toegankelijkheid](https://www.noraonline.nl/wiki/Duurzame_Toegankelijkheid)*

## 7 Bijlage 1: Overzicht onderzoeksvragen

### Onderzoekslijn 1. toepassing en gebruiksgemak

Hoofdvragen	Subvragen
Is deze werkmethode bruikbaar?	Is de methode begrijpelijk en helder voor medewerkers?
	Is het duidelijk wat weg mag en wat er moet blijven in de eerste tien weken?
	Hoe wordt het gebruiksgemak ervaren?
	(eventueel aanvullende vragen in de evaluatie afstemmen met ministeries die een pilot uitvoeren)

### Onderzoekslijn 2. (selectie van) sleutelfunctionarissen

Hoofdvragen	Subvragen
Wie zijn sleutelfunctionarissen?	Welke aanpassingen zijn nodig aan het bestaande waardering en selectie instrumentarium om de werkwijze integraal te incorporeren?
	Hoe kunnen organisaties sleutelfunctionarissen het beste aanmerken?
	Is de informatie die in de e-mail van een sleutelfunctionaris zit, voldoende voor reconstructie van dossiers op lange termijn?

### Onderzoekslijn 3. Duurzame toegankelijkheid

Hoofdvragen	Subvragen
Kunnen wij bij WOB-verzoeken, Kamervragen, parlementaire enquêtes etc. (informatieverzoeken als bedoeld in artikel 68 in de GW) iets terugvinden?	Hoe snel kan een informatieverzoek worden verwerkt in een opslag met veel e-mailberichten?
	Hoe nauwkeurig is het antwoord op een zoekvraag? (hoeveel overbodige bijvangst is er, of hoeveel gemiste informatie)
	Hoe groot is de kans dat informatie 'wordt gemist'?
	Hoeveel hits levert een gemiddelde zoekvraag op?
Kan het op deze wijze aangemaakte emailarchief voldoende worden ontsloten om aan de informatiebehoefte van	is de informatie volledig en accuraat te achterhalen, of is er een koppeling nodig met een database die e-mailberichten rangschikt naar dossiers, afdelingen/ministeries e.d.?



toekomstige generaties te voldoen?	Kan software helpen bij het ordenen en waarden van de informatie in het bedoelde e-mailarchief?
Zijn organisaties toegerust op het gebruik en beheer van dit archief?	Hoeveel email wordt op deze wijze langjarig (10 jaar) opgeslagen?
	Hoeveel email wordt op deze wijze op termijn overgebracht naar het Nationaal Archief?
	Wat zijn de (langjarige) opslag kosten?
	Wat zijn de implementatiekosten van de nieuwe werkwijze?
Weke vereisten, denk onder meer aan preservering, komen voort uit DUTO ten aanzien van het realiseren van de oplossing?	Welke maatregelen moeten dienstverleners nemen bij de inrichting van de e-mailapplicaties en e-mailbeheeromgevingen?

#### Onderzoekslijn 4. Openbaarheid en Privacy

Hoofdvragen	Subvragen
Op welke manier kan het beste uitvoering worden gegeven aan de voorschriften ten aanzien van openbaarheid, en hoe moet worden om gegaan met eventuele knelpunten daarbij?	Onder welke voorwaarden kan de inbox van een sleutelfunctionaris openbaar worden gemaakt, mede gezien voorschriften uit WBP, Archiefwet e.d.?
	Op welke wijze kan optimaal worden voldaan aan toepassing beperkingen openbaarheid bij overbrenging?
	Kan software helpen bij het toepassen van openbaarheidsbeperkingen?
Is voldoende invulling gegeven aan de voorschriften vanuit de wetgeving en andere bepalingen ten aanzien van privacy?	Is het privacyreglement (op te stellen op basis van de uitgevoerde PIA) afdoende, is er meer nodig op basis van de WBP of de komende AVG?
	Hoe dient de privacy van externen die met de overheid informatie uitwisselen via email, te worden geborgd, gezien de openbaarheid die bij sleutelfunctionarissen wordt beoogd?