



Rijksinstituut voor Volksgezondheid
en Milieu
*Ministerie van Volksgezondheid,
Welzijn en Sport*

Visie op data storage RIVM

Hoe kan RIVM de storage dienstverlening
verder professionaliseren met waarborging
van

- Functionaliteit
- Prestaties
- Schaalbaarheid
- Flexibiliteit
- Transparantie
- Kosten-efficiëntie en
- Veiligheid en privacy?

Visie op storage | 22 april 2020



Inleiding

Dit Visiedocument geeft richting aan de doorontwikkeling van storage-voorziening van het RIVM. De storage-voorziening is het geheel van mensen, middelen en maatregelen, gericht op de opslag van data van het RIVM.

De kern van deze Visie wordt gevormd door de missie en ambitie van het RIVM zoals verwoord in RIVM2025 en de daarvan afgeleide iVisie 2025.

Onze missie

Het Rijksinstituut voor Volksgezondheid en Milieu (RIVM) zet zich in voor een gezonde bevolking en een duurzame, veilige en gezonde leefomgeving. We zijn daarvoor de betrouwbare adviseur voor de overheid, professionals en burgers.

RIVM staat midden in de samenleving en in die samenleving volgen maatschappelijke, technologische en wetenschappelijke ontwikkelingen elkaar snel op. Vraagstukken op het gebied van milieu, duurzaamheid en veiligheid. Maar ook ontwikkelingen als (de)centraliserende krachten in de maatschappij, waaronder de behoefte aan transparantie en draagvlak voor wetenschappelijk onderzoek. Naast deze ontwikkelingen bieden tal van technologische innovaties ons mogelijkheden. Hierdoor wordt ander onderzoek mogelijk en worden wij in staat gesteld om te reageren op de uitdagingen in de maatschappij.

De storage-voorziening van het RIVM dient tijdig aangepast te worden op deze ontwikkelingen.





Ontwikkelingen

De behoefte aan betere, kosten-efficiëntere, flexibele storage wordt gedreven door verschillende ontwikkelingen:

- De 'organische' groei van de data, die wij verwerken en gebruiken voor onze onderzoeken en wettelijke taken;
- 'Nieuwe' technologieën en ontwikkelingen, zoals bio informatica, artificial intelligence en machine learning, de Personal Health train en FAIR data services;
- Meer en (liefst) betere samenwerking, zowel internationaal als op nationaal niveau met onderzoekspartners en ketenpartners;
- Wettelijke kaders en rijks brede bewegingen, zoals de [WOG](#), [AVG](#) en de [Data Agenda Overheid](#);
- De beweging naar de cloud: steeds meer diensten en applicaties zijn – steeds vaker uitsluitend – als cloud-dienst af te nemen
- De wens om de kosten van ICT transparanter te maken en te kunnen doorbelasten.



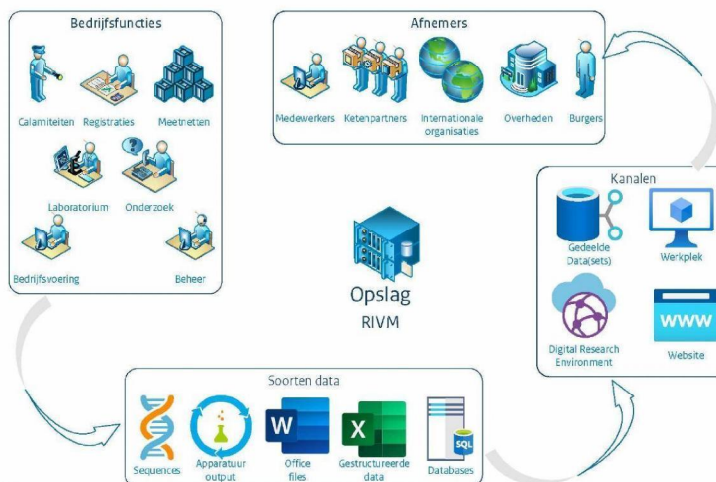
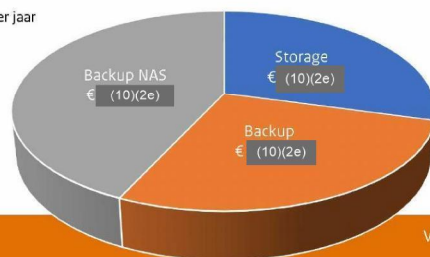
Data Driven Healthcare

Hoe kun je als zorginstelling met data de zorgkwaliteit en bedrijfsvoering verbeteren? Hoor van experts: Welke gegevens worden er gegenereerd? Wat is gemakkelijk te ontsluiten en wat niet? Hoe kun je die data het beste analyseren en met welk doel doe je dat? En hoe vertaal je dat naar toepassingen waarvan zowel de 'klant' (patient, client) als de medewerkers en organisatie profiteren? Zodat de zorg efficiënter, meer op maat, met minder fouten en daardoor goedkoper kan worden geleverd.



Huidige situatie

- Block storage (SAN)
Deze storage faciliteert de diensten Managed Serverhosting en Generieke Werkplek.
- File storage (NAS)
Dit zijn de R\ en de S\schijf, waarop losse bestanden kunnen worden opgeslagen.
- Backup en recovery
Een faciliteit voor het opslaan en kunnen terughalen van reservekopieën van zowel block als file storage.
- Beheer
Het beheer van de storage oplossingen wordt technisch uitgevoerd door de afdeling Basisinfrastructuur. Het overgrote deel van de beheerwerkzaamheden wordt daarbij gedaan door het team "Platform". Voor configuratie en exploitatie van de file storage wordt veelvuldig gebruik gemaakt van het Windows serverbeheer-team en het het Linux serverbeheer-team.
- Kosten
ca. € (10)(2e) per jaar



Verskillende soorten data, met

- verschillende requirements ten aanzien van beschikbaarheid, betrouwbaarheid, etc.
 - verschillende levenscycli
 - ten behoeve van verschillende doeleinden en
 - afkomstig uit verschillende bedrijfsprocessen en -functies
- worden opgeslagen op één en dezelfde file storage van het RIVM.



Knelpunten

In de huidige storage omgeving heeft de file storage problematiek de aandacht en zijn er acute problemen op het vlak van de file-storage (NAS). Een van de technische issues is dat de storagecapaciteit haar maximale grenzen (>90%) heeft bereikt en RIVM hiermee het risico loopt van een grote verstoring, waarbij de NAS diensten (webapplicaties, websites, homeshares, databases) dan niet meer beschikbaar zullen zijn voor een groot aantal klanten. Er is onvoldoende storage aanwezig in de technische infrastructuur om in de behoefte van de business te voorzien. Om te kunnen blijven voldoen aan de korte termijnsvraag, dient de NAS storage zo snel mogelijk uitgebreid te worden.

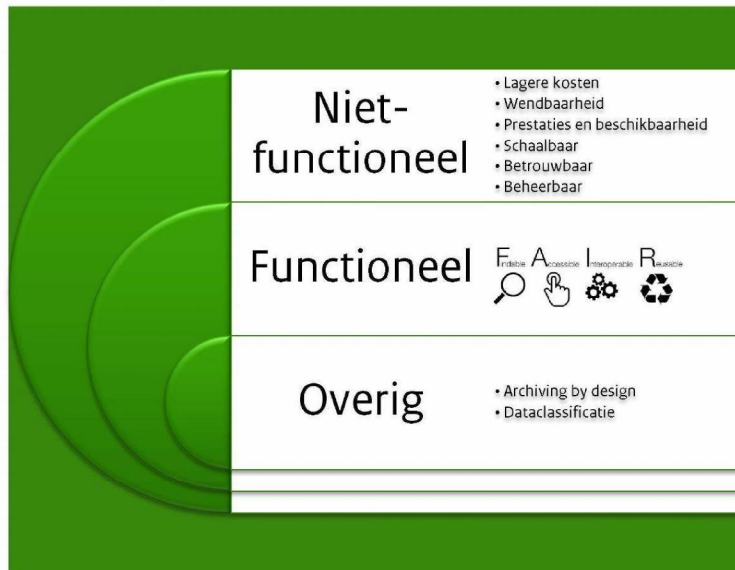
It's time to Get Rid of
the Bottlenecks



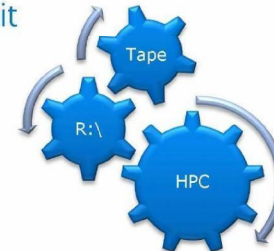
- Archiveringsfunctie (archiefwet) is niet geïmplementeerd op de NAS, met als gevolg dat archiefdata op de primaire storage achterblijft na verloop van de bewaartermijnen en niet verplaatst wordt naar het archiveringsplatform.
- Idem voor de backup- en restore functionaliteit.
- De huidige opslag biedt geen goede mogelijkheid tot het loggen van toegang; wie heeft wanneer in welke folder/data gekeken?
- De netwerkschijven worden gebruikt voor vrijwel alle opslag, er is geen goede optie voor het archiveren van data
- Verschillende mensen hebben problemen met de traagheid van de netwerkschijven
- DVD beheer kan niet optimaal toegang tot folders verlenen en verwijderen, vooral het verwijderen van toegang voor een individu is erg omslachtig wanneer het om meerdere locaties gaat
- Transfer van data tussen opslaglocaties, apparaten, etc. wordt niet altijd goed gefaciliteerd.
- De huidige workflow/dataflow (ruwe data > analyses/bewerkingen > dataset > analyses/resultaten) staat verdeeld over de netwerkschijven R en S, dit wordt vaak ad hoc uitgevoerd.



Behoefte RIVM



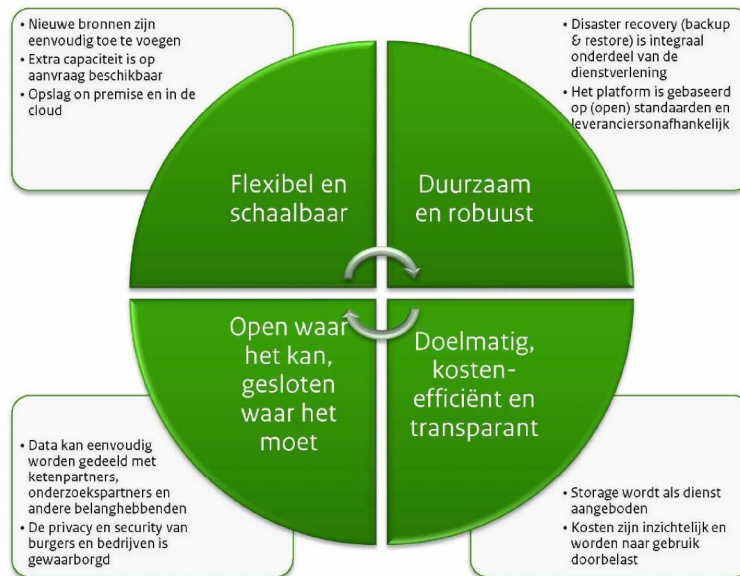
Voorbeeld van de gewenste flexibiliteit



Data uit NGS-sequencers wordt gedownload naar de R:\ schijf (gemiddelde toegangstijden). Deze data moeten vervolgens in een High Performance Cluster beschikbaar zijn met zeer snelle toegangstijden. Nadat alle berekeningen hebben plaatsgevonden, kan de bron-data worden gearchiveerd op langzame (en goedkopere) storage



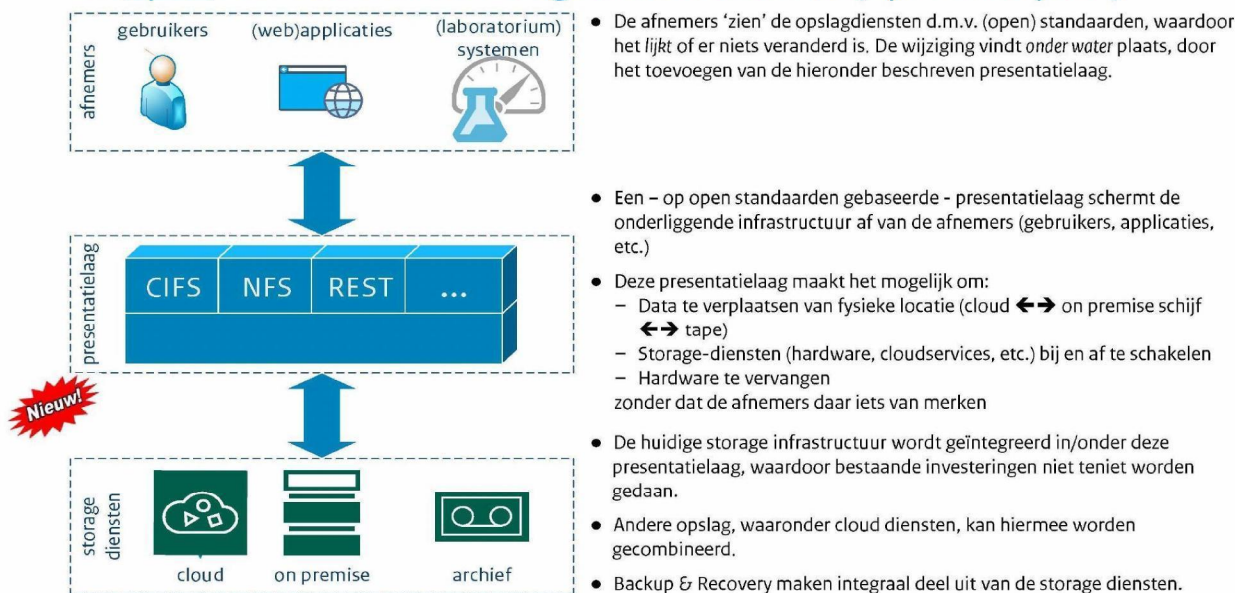
Gewenste situatie



De wetenschappelijke wereld is volop in beweging. Onderzoekers werken vaker en intensiever met elkaar samen, en doen dat steeds meer over de grenzen van instellingen heen. Een datacentrische oriëntatie is cruciaal - niet alleen om efficiënter en betrouwbaarder te werken, maar ook om nieuwe manieren van onderzoek te faciliteren waarbij data uit verschillende kennisdomeinen kan worden gecombineerd om antwoorden te formuleren op de vaak complexe en multidisciplinaire vraagstukken van deze tijd. Deze beweging vindt ook internationaal weerklank in onder andere de FAIR principes en de impuls voor Open Data en Open Science en open overheid.



De open architectuur van de gewenste situatie (op hoofdlijnen)





Voorgestelde marsroute

