

00.631/20 DR



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

nota IV board

A. van Leeuwenhoeklaan 9
3721 MA Bilthoven
Postbus 1
3720 BA Bilthoven
www.rivm.nl

Nota nummer

Datum

22 oktober 2020

Betreft	Behoefte reken- en opslagcapaciteit RIVM		
Auteur	(10)(2e)	(10)(2e)	(10)(2e)
Alvast met	(10)(2e)	(10)(2e)	(10)(2e)
Besluit	Ja		
Document	n.v.t.		
Externe opdrachtgevers	n.v.t.		
Medezeggenschap	n.v.t.		
Gevolgen voor ROS	Nee		
Bijlage(n)			

Aanleiding en doel van de nota

Aanleiding voor deze nota zijn de geluiden vanuit de centra, dat er een explosieve groei van de hoeveelheid data te verwachten is. Voornamelijk als gevolg van de vele extra COVID-19 onderzoeken en de sterke toename op het gebied van Next Generation Sequencing (NGS). NGS is reken- en data-intensief en er is dus ook een substantiële toename van de behoefte aan reken- en opslagcapaciteit te verwachten. De extra inzet van mensen en middelen voor Corona is begroot op twee jaar. Data zal blijven groeien, maar rekenen moet (deels) flexibel zijn ander moet de oplossing in 2 jaar worden afgeschreven.

Doel van deze nota is om maatregelen te treffen, c.q. activiteiten te initiëren om op deze groei te anticiperen en te zorgen voor passende IV voorzieningen.

Kernpunten

- Naar verwachting ontstaat op korte termijn een substantieel groeiende behoefte aan opslag- en rekencapaciteit, mede door COVID-19.
- Het RIVM en de IV organisatie moeten maatregelen treffen om voorbereid te zijn op de toekomst met betrekking tot de veranderende behoefte aan reken- en opslagcapaciteit.

- Deze maatregelen zijn deels van technische en deels van organisatorische aard.
- Organisatorisch:
 - Lever (maandelijkse) prognoses vanuit de centra met betrekking tot de verwachte toe- of afname van NGS in het kader van diagnostiek en surveillance.
 - Maak het gebruik van Data Management Plannen verplicht. Uit een Data Management Plan (DMP) kan worden afgeleid wat de behoefte (van dat) onderzoek zal zijn. Een DMP wordt vóór ieder onderzoek in het kader van de onderzoek planning opgesteld en gedurende het onderzoek onderhouden. Op basis van het aantal monsters, dat zal worden geanalyseerd kan berekend worden hoeveel reken- en opslagcapaciteit nodig zal zijn.
- Technisch:
 - Implementeer monitoring van het daadwerkelijke gebruik van reken- en opslagcapaciteit, voor zover dit niet reeds het geval is. Trends in de behoefte aan reken- en opslagcapaciteit kunnen worden geëxtrapoleerd om de toekomstige behoefte te voorspellen en kunnen dan worden opgenomen in de offertes.
 - Werk een (of meer) oplossingsscenario('s) uit om ook op langere termijn in de reken- en opslagbehoefte te kunnen blijven voorzien. Doe dit onder architectuur, zodat de oplossing(en) passen in het totale landschap en houdt hierbij rekening met de verschillende use cases en niet-functionele wensen en eisen.

Toelichting

Situatie

Een door het bio informatica programma uitgevoerde analyse laat zien, dat thans¹ gemiddeld ca. 25.000 uur/maand gebruik gemaakt wordt van de rekencapaciteit van de High Performance Computing omgeving (HPC) en 9 TB aan opslagcapaciteit. Voor de korte termijn (2021) zijn, c.q. worden (op basis van gebruik 1e half jr 2020 en inschatting op lineaire groei daarvan) de volgende voorzieningen getroffen:

- Uitbreiding van de CPU-capaciteit van het HPC door uitbreiding van het aantal cores van 200 naar 450
- Aansluiting op het [SURFSara Data Archive](#). Ruwe data en tussenresultaten, die voor het verdere onderzoek niet meer nodig zijn kunnen – geautomatiseerd – naar deze voorziening worden verplaatst, waarmee ruimte wordt vrijgemaakt.

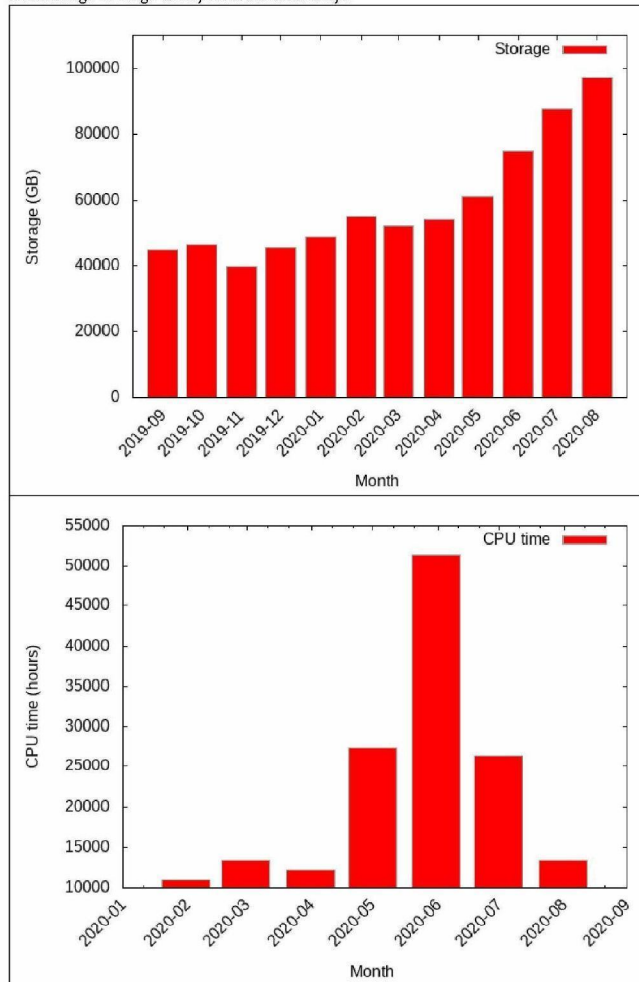
Naar verwachting zijn deze maatregelen voldoende om de groei op korte termijn op te vangen.

De toepassing van NGS zal echter in de komende tijd substantieel, zo niet exponentieel toenemen. Vanuit IDS klinken geluiden, dat men 10x meer zal gaan sequencen dan nu het geval is². Daarbij komen nog de ad-hoc uitbraakonderzoeken, herhalingen, fylogenetische analyses en analyses van de SARS-COV-2 Illumina data bovenop, en ook nog de huidige reguliere non-SARS-COV-2 Illumina data. ZGO gaat de SARS-COV-2 sewage surveillance opschalen, dus daar verwacht ik ook een toegenomen vraag. IDS-bacteriologie gaat ook meer in-house sequencen en analyseren. We zien soms dat het hele grid vol staat met statistische inferenties en, langer geleden, ook met NOX berekeningen. De huidige

¹ Peildatum: augustus 2020

² Voor SARS-COV-2 surveillance: 120 monsters per week, hetgeen neerkomt op 750 gb – 1 tb.

voorzieningen in combinatie met de hierboven genoemde uitbreidingen zullen dus op de middellange en lange termijn niet toereikend zijn.



Complicaties

- De te verwachte toename is onbekend; dit geldt voor diagnostiek en surveillance, maar ook voor onderzoek.
- NGS en de voorzieningen daarvoor staan niet op zichzelf en zijn slechts een deel van de functionaliteit, waarin het geheel van IV voorzieningen (het landschap) moet voorzien om onze taken naar behoren te kunnen uitvoeren. Denk hierbij met name aan:
 - Laboratorium informatiemanagement systemen (LIMS);
 - Referentie databases
 - Surveillance databases
 - Samenwerkingsfunctionaliteit
 - Omgevingen en tools voor (verdere) analyse

Dit betekent dat een eventuele oplossing altijd in samenhang met bestaande voorzieningen moet worden bekeken, i.e. moet passen binnen de architectuur van het bestaande landschap.

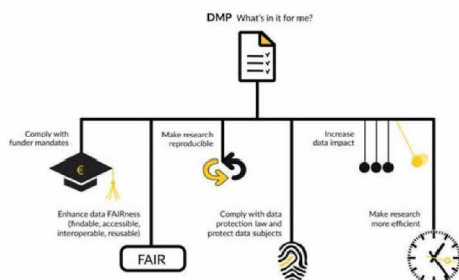
Kernvraag

Welke maatregelen moeten het RIVM en de IV organisatie treffen om voorbereid te zijn op de toekomst met betrekking tot de veranderende behoefte aan reken- en opslagcapaciteit?

Oplossing

1. Verkrijg inzicht in de te verwachten behoefte aan opslag- en rekencapaciteit:
 - a. Voor 'productiewerk' (diagnostiek en surveillance) kan dit door het daadwerkelijke gebruik te monitoren. Trends in de behoefte aan reken- en opslagcapaciteit kunnen worden geëxtrapoleerd om de toekomstige behoefte te voorspellen en worden opgenomen in de offertes.
 - b. Leg hiernaast frequente (maandelijkse) prognoses vanuit de centra met betrekking tot de verwachte toe- of afname van NGS in het kader van diagnostiek en surveillance.
 - c. Voor wetenschappelijk onderzoek kan uit een Data Management Plan (DMP) worden afgeleid wat de behoefte (van dat) onderzoek zal zijn. Een DMP wordt vóór ieder onderzoek in het kader van de onderzoek planning opgesteld en gedurende het onderzoek onderhouden. Op basis van het aantal monsters, dat zal worden geanalyseerd kan berekend worden hoeveel reken- en opslagcapaciteit nodig zal zijn.

Een datamanagementplan is een formeel document dat aan het begin van een onderzoeksproject ontwikkeld wordt en waarin alle aspecten van het beheer van gegevens worden beschreven, zowel tijdens als na het project. Bij databeheer draait alles om het efficiënt maken van het werk en het creëren van meerwaarde voor de data voor het project en anderen, tijdens en na uw onderzoek. Bijvoorbeeld door toe te werken naar FAIR (vindbaar, toegankelijk, interoperabel, herbruikbaar) of Open data. Door aan het begin van het project na te denken over datamanagement, is men goed voorbereid. De volgende voordelen liggen binnen handbereik met een goede datamanagementplanning:



2. Werk een (of meer) oplossingsscenario('s) uit om ook op langere termijn in de reken- en opslagbehoefte te kunnen blijven voorzien. Doe dit onder architectuur, zodat de oplossing(en) passen in het totale landschap en houd hierbij rekening met de verschillende use cases:
 - a. Diagnostiek
 - b. Surveillance
 - c. Onderzoek, zowel:
 - i. Uitsluitend binnen het RIVM als
 - ii. Samen met (inter)nationale onderzoekspartners, waarbij in sommige gevallen alleen data hoeft te worden gedeeld en in andere gevallen samengewerkt moet kunnen worden in een digitale samenwerkingsruimte of een digitale onderzoeksomgeving (DRE). Dit houdt in, dat er de behoefte is om opslagcapaciteit, rekencapaciteit of een combinatie daarvan te kunnen delen.

Houdt hierbij rekening met de volgende eisen:

1. Maak gebruik van de status van data (eg. brondata, tijdelijke data, resultaat, archief) met betrekking tot de technologie waar deze wordt opgeslagen en verwerkt.
2. Geef inzicht hoe de classificatie van data zich verhoudt tot de technologie waar deze wordt opgeslagen en verwerkt.
3. Onderzoek en adviseer hoe de huidige technische mogelijkheden om (tijdelijk) opslag en verwerkingsdiensten van buiten het RIVM benut kunnen worden. Onderverdeeld in twee typen: permanent en tijdelijk gebruik.
4. Geef inzicht in de financiën van de opslag en verwerkingsscenario's in relatie met omvang en prestaties.
5. Geef inzicht in de mate van flexibiliteit in capaciteit (kunnen groeien en krimpen) van opslag en rekenkracht van de scenario's.
6. Geef inzicht in de impact op de dienstverlening naar de afnemers (eg. prestaties, keuzes, kosten, support) alsmede op de beheerorganisatie (kennis & competenties) van de oplossingsscenario's.

De volgende eisen worden gesteld aan de opslag- en rekenomgeving:

- a. De oplossingsscenario's moeten toekomstvast zijn (3-5 jaar), waardoor ze flexibel mee moeten kunnen groeien met de verwachte capaciteitsgroei.
- b. De oplossingsscenario's moeten voldoen aan wet- en regelgeving. Met name op het gebied van informatiebeveiliging en privacy.
- c. Het eigenaarschap van de data blijft ten alle tijde bij het RIVM, ook bij het gebruik van 'externe' dienstverleners.

- d. De initiële kosten en de groei door capaciteitstoename dienen voorspelbaar en betaalbaar te zijn. Bij gelijkblijvende prestaties gaan we uit van meer capaciteit voor relatief minder geld.
- e. Elk scenario levert een passende prestatie voor de afnemer.
- f. Elk scenario garandeert beschikbaarheid en continuïteit door (inzicht in) beheerorganisatie(s).

EXTERNE OPDRACHTGEVERS

Geen impact.

MEDEZEGGENSCHAP

Geen impact.

RISICO'S (t.a.v. financiële, juridische of personele aspecten)

Geen.

FINANCIELE CONSEQUENTIES

Geen.

COMMUNICATIE

VERDERE AFHANDELING

Beslispunt

- Het aanwijzen van een opdrachtgever (uit het primair proces) en de instelling van een multidisciplinaire werkgroep om (een) oplossingsscenario('s) op te stellen om aan de behoefte aan reken- en opslagcapaciteit van het RIVM op langere termijn te kunnen blijven voldoen. In deze werkgroep moeten minimaal vertegenwoordigd zijn:
 - Een enterprise architect van het CIO office RIVM;
 - Een (of meer) domeinarchitect(en) vanuit (voormalig) SSC Campus;
 - Een (of meer) bio informaticus
 - Een (of meer) datamanager(s)
 - Een data steward
 - Een (of meer) informatiemanager(s)

De leden van deze werkgroep moeten minstens 2 dagen per week van hun tijd worden vrijgespeeld voor deelname aan deze werkgroep. De werkgroep wordt in staat gesteld om agile (waarbij regelmatig in één ruimte moet kunnen werken) in ca. 1 maand tot een eindresultaat te komen.

Deze scenario's moeten passen in de I&D strategie om te komen tot een landschap creëren waarin de verschillende cases kunnen landen.

Daarnaast moet er hier ook nagedacht worden over de permanente ondersteuning via (bijvoorbeeld) een Competence Center. De verschillende oplossingen moeten goed beheerd en door-ontwikkeld kunnen worden. Daarnaast is van belang dat er goede intakes gaan plaatsvinden om een case goed te laten landen in een oplossing.

- Inrichting van monitoring op de huidige opslag voorzieningen en het HPC, teneinde trends in het gebruik daarvan te kunnen zien.
- Maandelijkse prognose vanuit de centra met betrekking tot de verwachte behoefte aan reken- en opslagcapaciteit.
- Het gebruik van Data Management Plannen (DMP) voor onderzoeken verplicht te stellen.

Besluit IV board

[Formuleer kort en helder het besluit van het de iv board.]