

Overzicht archief data bij SURF onder grant NWO-2020.022 18 jan 2023

Introductie

Hieronder een overzicht van onze data op tape in het archief onder de grant NWO-2020.022.
Er is ook een Excell formulier met meer informatie over elke run die we op IMAU gedaan hebben
(omschrijving, aantal jaren, project etc).

Deze staat hier: https://webspace.science.uu.nl/~kliph103/Projects/Archive_SURF/
Dit Excell overzicht bevat niet alleen palaeo runs maar ook runs die ik voor projecten van Henk heb gedaan.

Er staat in ons archief in totaal \approx 1200 TB in directory /archive/heydtacc:

We mogen echter maar 970 TB onder deze grant bewaren.

Ik heb de data waarvan ik denk dat deze wel weg kan in de overzichten hieronder in **rood** gemarkeerd.
Voor een later overleg vlak voor de kerstvakantie heb ik nog tekst in blauw toegevoegd, dit gaan we nu ook met Michiel bespreken.

Als we alles weghalen zoals vorige keer met Anna besproken dan houden we nog zo'n 280 TB van de 1200 TB over!

Overzicht /archive/heydtacc

1200 TB

pop 175 TB

highres paleo

EO_38Ma_paleomag_2pic_tx0.1	(volgens mij een testrun)	2.6 TB
p21a.EO38Ma.tx0.1.2pic_control_bulk_ncep		55 TB
p21a.EO38Ma.tx0.1.4pic_control_bulk_ncep		50 TB

De eerste run kan weg, dan hebben we nog de paleo runs van Peter Nooteboom:

2pic run	43 jaar
4pic run	27 jaar

3D variabelen: UVEL, VVEL, WVVEL, KE, TEMP, SALT, UET, VNT, UES, VN, PD -> totaal 11
2D variabelen: SSH, H2, SHF, SFWF, TAUX, TAUy, HMXL, XMXL, TMXL -> totaal 9

Laatste 7 jaar van beide runs hebben we dagelijks geoutput, in totaal 45 TB (2pic) en 40 TB (4pic)

Als we van deze dagelijkse velden alleen TEMP,SALT,UVEL en VVEL opslaan en voor de rest van de runs

Met opmerkingen [BM(1)]: Kan volgens mij weg

Met opmerkingen [BM(2)]: Simulaties van Peter, eventueel T,S,U,V velden bijhouden van laatste x jaar?

Met opmerkingen [HAvd(3)]: Grootste data volume zijn de laatste 7 jaar met dagelijks output. Is het mogelijk het aantal variabelen (vooral 3D) te beperken? Anders is dit typisch iets voor een data archief (10 jaar bewaren) buiten dit project.

Alleen de laatste 20 jaar dan blijft er ongeveer 17 TB per run over, voor beiden dus 34 TB

→ Totale besparing voor deze runs is dan: 74 TB

lowres paleo

EO_38Ma_4pic_g16_NESSC_fHuber	18 TB
EO_38Ma_hotspot_4pic_g16_NESSC_fHuber	12 TB
EO_38Ma_paleomag_4pic_g16_NESSC_fHuber	15 TB
EO_38Ma_f19g16_forcing_shicheng	1.6 TB

Met opmerkingen [BM(4)]: POP simulaties, enkel laatste 200 jaar bijhouden? Misschien jaarlijkse oceaan/MSF velden van de hele run.

Met opmerkingen [Av5]: Alle 4 runs staan niet in de excel die je stuurde michael? Die laatste kan weg lijkt mij.

De laatste shicheng run kan weg.

De 3 runs daarvoor zijn allen 2000 jaar. Als we de laatste 200 jaar behouden en Steeds voor de hele run de jaarlijkse oceaan en MSF files dan houden we nog 31.9 TB over.

→ Totale besparing voor deze runs: 42.1 TB

shicheng_data

run_henk_mixedbc	9.6 TB
X1_SAMOC_flux	6.3 TB

Met opmerkingen [Av6]: Kan weg volgens mij

De laatste beide runs kunnen weg

→ Besparing: 15.9 TB

cesm 1025 TB

cesm1.0.5

lowres (f19_g16) 775 TB

Preindustrial

b.PI_1pic_f19g16_NESSC_control_3000j	71 TB
b.PI_1pic_f19g16_NESSC_control_restart_2500_pi_vdc_overflows_tidal_on_3000j	9.5TB
b.PI_1pic_f19g16_NESSC_control_restart_2500_palaeo_vdc_overflows_tidal_off_3000j	62 TB
b.PI_2pic_f19g16_NESSC_control_1000j	21 TB
b.PI_4pic_f19g16_NESSC_control_2000j	39 TB
b.PI_4pic_f19g16_NESSC_control_restart_5000_palaeo_vdc_overflows_tidal_off_1000j	19 TB

Met opmerkingen [BM(7)]: Liefst jaarlijkse atmosfeer/oceaan velden. Maandelijks enkel laatste 200 jaar.

Met opmerkingen [BM(8R7)]: Idem voor andere PI runs. Het hebben van de jaarlijkse data kan nuttig zijn voor CS werk. Eventueel is enkel atmosfeer voldoende.

Met opmerkingen [HAvd(9R7)]: Mee eens, maar ik zou wel ook enkele jaarlijkse oceaan variabelen bewaren, bijv. MOC, T, S,...

Met opmerkingen [Av10R7]: Als mogelijk: laatste 500 jaar maandelijks velden, daarvoor jaarlijkse data

Als we bv voor de eerste run alleen de laatste 200 jaar behouden en voor de hele run de jaarlijkse atmosfeer, oceaan en MSF files dan houden we hiervan over: 22 TB

→ Totale besparing: 200 TB

Pliocene

b.PLIO_5Ma_Eoi280_f19g16_NESSC_control	1000j	19 TB
b.PLIO_5Ma_Eoi400_f19g16_NESSC_control	2000j	47 TB
b.PLIO_5Ma_Eoi560_f19g16_NESSC_control	1000j	28 TB
b.PLIO_5Ma_xpic_f19g16_NESSC_control	(volgens mij een testrun)	22 TB

Met opmerkingen [BM(11)]: Wordt nog actief mee gewerkt, liefst behouden. Testrun kan weg. Van oceaandata eventueel deels jaarlijkse velden behouden ipv maandelijkse (laatste 500 jaar maandelijks behouden?)

Met deze Pliocene data wordt actief mee gewerkt, als we de laatste 500 jaar bewaren en voor de hele run de jaarlijkse atmosfeer, oceaan en MSF files en de laatste testrun weg gooien dan houden we hiervan over: 35 TB

→ Besparing: 81 TB

Oligocene

b.EO_30Ma_paleomag_2pic_f19g16_NESSC_control_MOC	1200j	21 TB
b.EO_30Ma_paleomag_1pic_f19g16_NESSC_control	3000j	55 TB
b.EO_30Ma_paleomag_2pic_f19g16_NESSC_control	3000j	62 TB
b.EO_30Ma_paleomag_1pic_f19g16_NESSC_control_correct_veg_final	500j	7.1 TB
b.EO_30Ma_paleomag_2pic_f19g16_NESSC_control_correct_veg_final	500j	8.8 TB
b.EO_30Ma_paleomag_1pic_f19g16_NESSC_control_correct_veg_final_low_insolation		8.4 TB
b.EO_30Ma_paleomag_2pic_f19g16_NESSC_control_correct_veg_final_low_insolation		8.5 TB

Met opmerkingen [BM(12)]: Jaarlijkse atmosfeer/oceaandata behouden. Maandelijks van laatste 200 jaar.

Met opmerkingen [BM(13)]: Laatste 200 jaar atmosfeer/oceaan behouden is voldoende

De laatste twee runs zijn ook beiden 500 jaar (paste er niet meer naast)
Als we van deze runs ook alleen de laatste 200 jaar behouden en voor de hele run de jaarlijkse atmosfeer, oceaan en MSF files dan dan houden we hiervan over: 24.8 TB

→ Besparing: 146 TB

Eocene

b.EO_38Ma_paleomag_2pic_f19g16_NESSC_control_correct_veg_final	1000j	23TB
b.EO_38Ma_paleomag_4pic_f19g16_NESSC_control_correct_veg_final	600j	12TB
b.EO_38Ma_paleomag_4pic_f19g16_NESSC_control_correct_veg	1300j	24 TB
b.EO_38Ma_paleomag_2pic_f19g16_NESSC_control_correct_veg	2800j	49 TB
b.EO_38Ma_paleomag_4pic_f19g16_NESSC_control	4000j	76 TB
b.EO_38Ma_paleomag_2pic_f19g16_NESSC_control	3000j	78 TB
b.EO_38Ma_paleomag_2pic_f19g16_NESSC_control_correct_veg_final_low_insolation		9.4 TB
pliomip2_data_for_server_leeds		344 GB

Met opmerkingen [BM(14)]: Goede runs; data behouden

Met opmerkingen [BM(15)]: Verkeerde vegetatie, kunnen volgens mij weg.

Met opmerkingen [BM(16)]: Jaarlijkse data behouden, maandelijks enkel laatste 200 jaar.

Met opmerkingen [BM(17)]: Laatste 200 jaar atmosfeer/oceaan behouden

De laatste run (niet de leeds data) is 500 jaar (paste er niet meer naast)

Als we van deze runs ook alleen de laatste 200 jaar behouden en voor de hele run de jaarlijkse atmosfeer, oceaan en MSF files dan dan houden we hiervan over: 32 TB

→ Besparing: 239.4 TB

cesm1.1.2

121 TB

Met opmerkingen [HAvd(18)]: Zijn dit wel palaeo runs? Ik dacht dat we nooit cesm1.1.2 succesvol gebruikt hebben voor palaeo, dus misschien alleen tests?

lowres (f09_g16)

b.e10.B2000_CAM5.f09_g16.pd_control.4xco2.001	90j	3.2 TB
spinup_B_2000_cam5_f09_g16	1300j	59 TB
closeddrake_B_2000_cam5_f09_g16	600j	28 TB
closeditf_B_2000_cam5_f09_g16	500j	17 TB

Van deze data is het onduidelijk of we het moeten bewaren. Stel dat ik alleen de laatste 200 jaar bewaar en alleen jaarlijks dan hou ik voor de eerste run 100G over en voor de andere 3 runs ieder 3TB, totaal dus: 9.1 TB. Anna zei de laatste keer echter dat het allemaal wel weg kan.

→ Besparing: 98.1 TB (of 121 TB)

Alle onderstaande data in rood is allemaal van Henk, ik had deze tijdelijk hier neergezet maar kan allemaal weg. Dit scheelt ook weer zo'n 23 TB

rcp8.5_co2_f09_g16.002	3.8 TB
rcp8.5_co2_f09_g16	3.7 TB
rcp8.5_co2_f09_g16.003	6.2 TB
b.e10.B2000_CAM5.f09_g16.pd_control.2xco2.001	1.8 TB

cesm1.0.4. (hoort hier niet, kan weg) 7.8 TB

bevat o.a.

grl_fw_prace_f05_t12
spinup_pd_maxcores_f05_t12
spinup_pd_5692cores_f05_t12
spinup_pd_5692cores_f05_t12
spinup_f05_t12

→ Totale besparing: 919.5 TB!!

Er blijft dan over: 280.5 TB en we mogen 970 TB

Er staat echter nog data in de /projects directory die nog in ons archief moet. Zodra ik alle huidige data die weg kan weg gegooid heb zet ik die er bij. Ik laat weten waar ik dan nog op uit kom vwb de totale hoeveelheid data op ons archief.