CR réunion CAN – jeudi 1er décembre 2022 à 14h @R117

Présent·e·s :
Lucie Etienne (CIRI)
Estelle Devêvre (CYTO)
Hervé Gilquin (PSMN)
Sandrine Hugues (IGFL)
Carine Rey (CIRI)
Laurent Modolo (LBMC)
Olivier Gandrillon (LBMC)
Jacques Brocard (PLATIM)
AnaMaria Kriss (RDP)
Mourad Bekhouche (LBTI)
Sergio Sarnataro (Spatial-Cell-ID)

Excusés
Olivier Ali (RDP)
Nicolas Goudemand (IGFL)
Jocelyn Turpin (IVPC)

OG proposes that we switch to english so that it's easier for SS to follow the discussion.

We then discuss the question of holding comodal meetings as raised by Olivier Ali (RDP). OG is not in favor of it since he's afraid physical attendance will decrease drastically if there is a video possibility. LM also raises the difficulty of balancing the discussion between real life and video, while JB blatantly calls comodal presentations "a pain to organize"! LE wonders whether Olivier Ali is absent just today – in which case minutes are redacted for this purpose – or will be absent each time. In such case, may be the RDP would choose another representative? Even though he lives in Grenoble, Olivier Ali was present last time, so that it is believed that he may still represent the RDP. OG finally rules against holding CAN meetings in a comodal mode.

SH will present the results of the poll she helped format and implement, and obtained from 83 respondents. This is to be compared with the ~ 1000 people present in the labs of the SFR Bioscience who may have responded… 30% are respondents from the RDP; 75% responded for themselves while 25% are team leaders or representatives.

The pool can be found from the CAN website at:
https://www.sfr-biosciences.fr/la-sfr/conseil-analyse-numerique/sondage-data-storage-and-transfer/view

Most people "feel comfortable" with their data but not all of them; general knowledge about Data Management Plans (DMPs) is scarce so that it would be interesting to train our colleagues on that

matter. Most people say their data storage needs will increase in the long term.

It appears that storage is performed almost 100% locally – computer and HD mostly – in addition to elsewhere for 50% – via a shared server or CBP/PSMN. 50% of the respondents indicate a data flux between 1 and 100 Gb/month; 10% more than 100 Gb! Most data are files bigger than 1 Gb which "need to be stored indefinitely". The need for accessing one's data is evaluated as 75% "Hot" and 50% "Cold".

50% of the respondents know they're not doing the best towards data management and 15% lost data during their career. 50% admit they need better storage solutions (while 15% estimate they don't). However, 50% also admit they do not have knowledge of data storage solutions within their own institute. 85% need to transfer data, mostly from SFR plateaus. 50% would use HDs or FileSender for that purpose.

LM remarks that huge amount of data may not travel through the network? HG says that there should be no difficulty since the ENS network is being boosted from a mere 10 Gb/sec to 40 Gb/sec. Also, one "useful" Pb (petabyte=1000 terabyte) now costs ~ 50 k€ (= 50 €/Tb) only.

60% of the respondents have difficulty transferring their and as many would like a universal solution. SS asks whether it would be possible to limit storage to non-redundant files only?

60% of respondents share data with colleagues via Filesender or ftpserver, but 25% use Wetransfer/Google Drive or Slack.

LM indicates that ANR now makes it compulsory to mention DMPs in grant proposals.
OG would like to publicize the existing solutions for data storage. SH says that it can be listed on the CAN web page, which OG feels would not be enough. Let's have 1-hour seminars instead, to inform everyone interested. LM adds that we may also indicate requirements such as the necessity to transfer data to public archives. For cold data storage, OG reminds everyone that the iRODs solution has been developed locally, @LaDoua actually: https://irods.org/ It is required to write a DMP to access it though.

HG comments that PSMN should not be used for cold storage and that a physical link between PSMN and IN2P3 / iRODS exists. So that fast data transfer is ensured between these two facilities.

AK asks for advice about data structure so that their availability is actually useful? HG says DMPs will help with that while OG disagrees as data storage structuration will not help with identifying conditions of data realization (dates, sample, protocol, etc). Even in the presence of an accompanying electronic lab book. The priority here is to help anyone with their own data storing and accessing.

MB says the long-term need for storage is set by the possibility to run new analysis on previously acquired data. ED realizes that public data storage should be enough for that purpose. Depending on how long data are stored for. LM estimates that it's probably as long as possible :) Also, he adds that every piece of data produced by public institutions employees (CNRS, INSERM, etc) are considered public anyway, unless embargoed.

OG notices that everybody seems OK with 1-hour presentations to a general SFR audience about data storage. Content will be debated during the next meeting, early 2023. HG and AK will be added to the CAN mailing list can@listes.ens-lyon.fr.

LM remarks that we did not even talk about biomedical data that should be confidential and/or anonymized! MB wonders whether there should be some sort of reviewing before data are stored in specific databases for biomedical data?

CAN meeting ends at 3.15 pm. JB has written the present report. Edited by the CAN members.