Mail Archive Q & A

Q1.- What's the estimated total size of the emails to archive? We just need the order of magnitude.

A1. The current archive is around 14Gb when in mbox form. See <http://www.ietf.org/mail-archive/text>

Q2.- What's the estimated maximum size of an mbox file or a Maildir message file?

A2. Most of the individual mbox files are small - for an example of a very active list, see
<http://www.ietf.org/mail-archive/text/rtcweb/>
Larger files are rare, but they exist. The largest of the files in that tree is just under 50Mb.

Q3.- Is it required to allow subthread queries? (a subthread query looks inside a message and all its descendants in that thread but not its ancestors).

A3. This should be possible, at least as a combination of "messages in this thread" AND "messages after this datetime"

Q4.- What's the Mailman version that the mailing lists use? If there are several different versions, what's the oldest one?

A4. The lists hosted at ietf.org are handled by Mailman 2.1.12 currently.

We have no insight into the lists hosted externally beyond the artifacts they may have added to the copy of the message that ends up in the archives (see the requirement to archive messages by subscribing to a list in section 2.2 of draft-sparks-genarea-mailarch).

Q5.- Will the archive system has access to the mailman files to check for user permissions information or it will need to run in different machine? If not, can the system has a component that is run in the mailman machine/s to provide that information as a REST Api or something similar?

A5. Permission checking will be needed for the (very small number of) private mail lists. And the archive will interact with the IETF's datatracker system, not mailman when it needs credentials. Some integration with the datatracker will be required.

Q6.- About the export feature, does it need to include external mailing lists or just the local mailman ones?

A6. Archives are maintained locally, even if the mail list is hosted somewhere other than ietf.org.