Detailed below is some preliminary feedback that is yet to be discussed with the wider working group.

1) Data format in between DST and muDST. Although several analysis will never be happy with muDST, it is also true that having all tracks and calo clusters is, in some (probably many) cases, not needed. As an example, often isolation variables are calculated within a cone and what lies outside of this cone is not used at all and could be ignored. We would like your feedback on what is the minimal set of information you need for analysis that currently are on DST to understand whether some reduction in event size can be achieved. In particular, coming back to the "cone", we would like to have some idea of how this cone can be defined and a quantitative estimate of the "size" of it. Any other idea to save space is of course welcome.

This could certainly be very useful. It would allow, for example, to use more information about the tracks in the cone to perform isolation (i.e. in addition to the pt asymmetry etc. as provided by the current tools). This is something we're thinking about a bit in the Kshh analysis group for example. Unfortunately, in this case the “one size fits all” will not work, as different cone ranges have been required by different lines (often more than one cone per line). The cone angles currently used by the BnoC group range from 0.8 to 2.0.

2) Turbo will be the default in the upgrade (as is muDST in the Stripping now). How many analysis in your WG can not be moved to turbo? Keep in mind that point 1 and 2 are not orthogonal. We could envisage that in the future turbo will be able to provide different data formats in output.

This is slightly tricky to say since we have zero experience with using the Turbo at present. Also, most of our analyses use the Topo triggers in Hlt2. For the Topo to be written to Turbo you'd have to write out the rest of the reconstructed objects in the event as well so that you could then build the complete candidates offline. Would this be feasible? This would be great, but would the rate be way too high? Even streaming the Topo into 2/3/4-Body and even separating it in mass ranges would not work, as many 4 body decays often are triggered only by 2 or 3 body topos. We think that the only solution will be to keep the inclusive triggers into full DST format, but that will be a big burden in terms of bandwidth. Another alternative is that we'd have to add tons of exclusive lines (I guess akin to the current stripping lines) for the various analyses. But what about analyses that haven't been thought of yet?

3) Centralised ntuple production. Ntuples will never die and right now all users have to make them by running over the stripping output. Would you be happy if ntuples could be produced every couple of weeks centrally by the production team? The idea is to have analysis "trains" leaving at fixed time intervals and each "wagon" is an ntuple. The WGs have to provided the code to generate the ntuples and ask it to be added to the next "train". Ntuples can be downloaded once ready from the bookkeeping for offline analysis. The main advantage is that the resources of the users can be moved to the production team which could optimise the resource usage such as data replication, number of jobs on the grid, etc… If this scheme is adopted, in the future there will be a need for a dedicated liaison for each WG (similar to the MC and Stripping one in terms of workload).

From what we'd heard about centralised productions from other experiments we were initially not very enthusiastic. But this idea of each analysis still having their specific code for producing the ntuples but the production being centrally managed sounds rather sensible. This sounds good - taking away a lot of the pain of managing jobs from the

users. Our only concern here is that this is something that will have to be organised quite carefully. Code will have to be released in due time and validated by the analysts in order to avoid producing tuples for the same decay every 2 weeks. For analyses managed by a big group I think it will not be a big deal (manpower for the validation of tuples can be provided by large groups), but for analyses managed by 1 or 2 people, the chances that this work will be done “with the left hand and a blind eye” are quite large.

4) Event index. We want to understand if random event access is a viable option for the upgrade. The idea is that events are indexed according to trigger/stripping line they fired and everything is put into a sort of database. The users will make a query to the database which will return a list of events. Therefore users will read only the events they are interested in (I remind you that right now, one has to loop sequentially over a file, even if most of the events are not useful for the analysis) Most importantly, these events can be anywhere in the grid. The gain from this approach is a much faster access to the events and allows an optimisation of data replication. We would need input from the WG to understand what information should go into the index (e.g. Stripping lines, Trigger lines, multiplicities etc)

We are not 100% sure how this relates to 3). Is it envisaged that this would still be implemented if centralised NTuple production was introduced? In terms of the information that should go into index we think that stripping and trigger are certainly essential and that magnet polarity, TCK and run number would also be useful.