T1 risk assessment

Scenario 1 - "Data loss at a T1"
Scenario 2 - "Partial loss of a T1"
Scenario 3 - "Procurement failure at a T1"
Scenario 4 - "Extended T1 outage"

PRELIMINARY - Daniele Bonacorsi - Last update: 22 Feb 2010

The worst scenario. It can happen all of a sudden (e.g. fire, tornado), or it can happen if Scenario 2 and 3 end up collapsing into this one.

Actions:

1. Assess the incident details. Talk to the affected T1 staff people. In such a condition, WLCG is already informed for sure, and is following up. Stay in the loop, and be sure you are in-sync. Strongly rely on CMS representatives at WLCG daily calls.

2. Profit of the WLCG support and infrastructure to follow-up with the T1 staff for a prompt solution of the problem. CMS Ops should focus instead on the following bullets.

3. The central Transfer Operator(s) centrally suspend all transfers in progress and most probably also any processing, aka freeze all the CMS workflows at the affected T1.

4. Quantify the general damage to CMS. Identify the possible consequences of a "what if this T1 will not come up again in months". It should be a joined {Data,Fac}Ops effort.

5. If the affected T1 **is FNAL**:

   - Each T1 hosting a non-custodial copy of the currently-not-accessible custodial data at FNAL gets this data 'promoted' to custodial.
   - Strategic decisions to be taken:
     - Strongly modifying the data distribution model, and/or using (a set of) T2's as a tapeless T1 for some period... might be forced decisions.

6. If the affected T1 **is not FNAL**:

   - DataOps **immediately** needs a backup T1
     - Must be chosen among the CMS ones
     - Computing Coordination and {Data,Fac}Ops propose which one(s)
       - (There are implications: see next bullet)

- The relevant CMS T1 contacts give green light to
  - Store new data (to always have the 1+2 RAW copies)
  - Store new MC (to allow T2s to be safe and dynamic)
  - Take the ownership of running prompt-skimming (now this is needed)
  - Not rolling back to the situation before the incident (i.e. this T1 will stably help to restore a datasets custodiality scheme at T1's)

7. In this extreme scenario, FNAL might be the obvious candidate. Consider that it could even not be possible, and harder the more we go deep into data-taking. Obvious questions to address (more can easily be found...):

   - ✦ Is disk/tape capacity at the back-up T1 sufficient?
   - ✦ Will the affected T1 immediately a smaller (or null) share...
     - Possible causing issues at other sites
   - ... or will it have to take more data later?
     - Higher bandwidth requirement at the post-problematic T1

8. In whatever data-taking time in which the incident occurred, due to this severity, DataOps must worry about 'old' data as well, and starts to discuss with Physics and assess priorities on what needs to be reproduced/ reprocessed/retransferred at/to the back-up T1.

9. In case you need to treat some data as "lost" to proceed, follow guidelines as from the "data loss" scenario:

   ➡ See the "Scenario 1" worksheet (bullets 9, 10, 11)

10. If/when the problem is fixed and the affected T1 site is back to operations:

    - ✦ The central Transfer Operator(s) runs a global `BlockDownloadVerify` to verify the overall status and accessibility of CMS data.
    - ✦ Most probably, in this conditions the affected T1 site will not negotiate with the Computing Coordination to eventually roll back data placement, and we will by default follow the general DataOps suggestion to not to attempt it.