# 3.7 Practical experiences with WIS centres implementations

## Most common problems of WIS centers

During installation and set-up of software for several WIS centres (GISCs and DCPC) we were confronted with questions related to operational issues which we were not always able to give unequivocal answers to. I will try to summarize them below.

### How to populate 24 hour cache of globally distributed data

Several approaches are used or intended to be used

1. GTS way

Most of data is available (GTS RTH), let’s ask for what is missing (in old GTS way).

* 1. Pros
* Technically possible right now with help of routing catalogues and product catalogue (VolC1).
  1. Contras
* quite labor intensive (needs to be negotiated with various RTHs)
* 'old fashioned' - not sure how it will work in future

1. GISC to GISC cache synchronization
   1. Pros

* independent on GTS network and procedures
* involves only GISC (i.e. possibly easier to manage)
  1. Contras
* no common protocol defined yet (several approaches possible and partially used)
  + FTP push of regional data (partially GTS like approach too)
  + exposing cache over FTP or HTTP - pull what you need and whenever you want (may cause delays)
* not sure if GISCs really have all globally distributed data

#### General aspects of data synchronization

1. Identification of data to synchronize

* List other center's cache, compare and download (PULL)
  + requires common cache item representation
* Time critical data - store and forward (PUSH) to avoid delay
  + how are the timely critical data defined?

2. Data transfer

* HTTP or HTTPS (FTP is less efective and bad for fire-walls)
* rsync

3. Verification of cache completenes (“all globaly distributed data available”)

* Comparing product catalogue witch cache content
* Routinely produced data should have timing described
* How to check non-routinely issued data?

### How to harvest from other GISCs

* + 1. One set from each GISC
    2. All from two (three) nodes

In both cases it is necessary to handle duplicates (product descriptions) harvested from more than one center. Technically each GISC or DCPC should know its area of responsibility and include into its own set only data from this area but errors happen. Possible solutions:

* + - * "white list" - list of patters (with wild cards) of products for each other GISC or DCPC we are harvesting. This white list can be based on (http://www.wmo.int/pages/prog/www/WIS/centres/index\_en.php) and needs to be maintained by every GISC or DCPC which harvest MD from other center.
      * duplicate elimination - newer record is better (?)

### What level of redundancy is required

It is clear that WIS centres needs to be highly available and therefore need duplicate power supplies, redundant data storage etc. Some centres have HA clusters with each node in different room (or even building). Thanks to this, the systems reach service availability of 99.9% or higher. In addition some centres do have disaster recovery site - i.e. another identically configured system in distant location. This seems sufficient, the later (disaster recovery center) even more than sufficient. Requirement for GISC to GISC backup comes on top of it and without clear definition of what services are considered essential. Having cleaner definition of required service availability would help centres to plan their infrastructure.

### Metadata granularity - What is product?

Each known GTS message is currently defined as product. This is used as guideline for other bulletins or GTS messages in general. But people who define non-GTS products often don’t know what appropriate granularity for their product is. Some producers define each new “instance” as new product effectively regenerating their catalogue every day. Others struggle to define products where the granularity is configurable.

Many products are available through (web) services which allow selection by data type (parameter) and geographical location and produce (customized) data instances on demand. If such service is described as one product its abilities are not obvious from MD record (all observations could be described as one product).

## Recommended Text

The technical aspects of synchronization of globally distributed data (aka 24 hour cache) need to be discussed by appropriate task teams to be able to come with clearly defined procedure and protocols as soon as possible because the global data cache is essential functionality of WIS.

As the WIS grows and more product descriptions (metadata) are harvested into the DAR catalogue more conflicting entries (metadata records with the same ID) appear. Therefore it is recommended to develop procedure which either identifies such duplicate IDs in advance or procedure to automatically resolve such conflicts.

It is recommended to detail service level of GISCs and other WIS centres in quantitative terms (required availability of logical components) in order to allow effective planning of WIS center infrastructure. For example if the regular dissemination of data has highest required availability the center may consider having this functionality available in its disaster recovery center unlike other functions.

It is proposed to discuss product granularity to avoid flooding of DAR catalogue with too fine-grained records as well as to promote description of services which offer parameterized products.

## 