## WIS Monitoring Status Analysis (2016-03-22)

### What We Are Currently Doing

1. Monitoring Pilot participated **only by GISCs**
2. **Three** JSON files are published **daily**
	* monitor.json, centres.json and events.json
3. All monitoring stats in the JSON files are **snapshot values** at 00 UTC,
	* For examples: number\_of\_records\_at00UTC, number\_of\_products\_all.
4. All monitoring can be done **completely internally** without help from outside of the GISC's network.
	* For an example, the up and down status of the OAI-PMH service is measured internally. So it is possible that the service is down for external users but is still reported as up by monitoring.

### What Markus Has Demonstrated

(<http://perl-wismon.rhcloud.com/wis-mon/index.html>)

(<https://gisc-test.dwd.de/monitor/>)

1. It also **uses JSON files** internally
2. More of a **real-time** (every five minutes) monitoring instead of daily
3. Service **availabilities** are reported as seen by **externally** users/agents
	* It is achieved by sending HTTP queries from a 3rd party server (rhcloud.com in this case) to a GISC's service URL (e.g. OAI-PMH) and calculate the response time.
	* Status are reported for **all 15 GISCs**
	* The actual response time is reported instead of simple up or down. This allows the use of more colours for different values.
4. More **analysis** against **24h-Cache** of the metric reporting GISC
	* Arrival time interval between two bulletins from a centre
		+ Not sure exactly how this is reported at a point of time. For an example, if we are to report this value every 10 minutes, is it reported as an average of last 10 minutes?
		+ It also reports missing bulletins. But it is NOT clear how it is calculated (e.g. over what time period?)
	* Data volume size of a centre every 30 minute
	* Other metrics for each centre includes the follows:
		+ Missing metadata (daily updated)
		+ Potential corrupted BUFR (daily updated)
		+ A long list of bulletins and their last arrival time in the 24h-Cache

### Potential Integration Between Current Monitoring Scheme and Markus's Effort

1. **It is possible to add more 24-Cache analysis results (Markus item 4) in centres.json file.** For each centre in the GISC's area of responsibility (AOR), we can expand the existing file structure to include more metrics. For an example, the entry of a centre in current centres.json file is as follows:
* ...
{
 "centre": "New\_Caledonia",
 "count": 127,
 "volumesize": 63605
}
...
* It can be expanded and restructured into the follows:
* ...
{
 "centre": "New\_Caledonia",
 "metrics": {
 "count": 127,
 "volumesize": 63605,
 "data\_arrival\_interval": 0.15,
 "missing\_metadata": 5,
 "corrupted\_BUFR": 4
 }
}
...
	+ A few notes about above JSON fragment are:
		- The metric missing metadata is simplified by a total number instead of a list of individual records. (It is technically possible to list the record. But it could make the file quite heavy and difficult to read for human.
		- The potentially corrupted BUFR is also represented by a total number for simplicity.
		- The long list of every bulletin and its last received time **does not seem to fit** into existing monitoring scheme. This file is also quite large in size (over 4MB uncompressed).
1. **Using centres.json file, it is also possible to expand its semantics to include information for other GISCs** in addition to DCPC/NC. This would make it possible to transmit service availability metrics (Markus item 3) such as the follows:
* ...
{
 "centre": "Offenbach",
 "type": "GISC",
 "metrics": {
 "oaipmh": 1.08,
 "portal": 1.35,
 "SRU": 1.28
 }
}
...
* NOTE that the act of adding the new metrics to centres.json extends the purpose of this file to describe statistics about **any centre within the entire WIS network**, not just centres within the reporting GISC's AOR. The centres can be other GISCs or DCPCs/NCs that are outside of the GSIC's AOR.
1. The above changes can be done **without altering the current metric reporting frequency (daily)**, *though it could be very beneficial to increase the frequency to achieve semi-real-time effects if technically feasible*.
2. Both monitoring **do NOT require active involvement from DCPC and NC**, i.e. the metrics can all be calculated from the reporting GISC side using currently available information.
3. Service **availability** is likely to be more **useful** when measured **externally**. A GISC can easily measure other GISCs' availability externally. But **without using external servers, a GISC can only measure its own service availability internally**.
	* Since metrics are provided by more than just a single GISC, the full external availability can be derived from them.
	* It could be challenging (technically, financially, politically) to ask a GISC to perform the monitoring using both external and internal (required for Cache analysis) servers. It is much easier to allow each GISC to provide its own availability as an internal measure and defer to other aggregation programs (e.g. Common Dashboard) to provide the complete view of the network.
4. It could be advantageous for the monitoring ad-hoc task team to make provisions to allow anyone to deploy systems that produce or consume monitoring statistics as they see fit. The act of doing so will facilitate the creation of a **WIS monitoring "ecosystem"** with independently running systems managed in a **decentralised** manner. Some ideas of such provisions that may need to be defined are as follows (these provisions are backwards compatible with what we have now):
	* Systems that produce statistics are responsible for publishing these statistics as JSON files using HTTP.
	* Systems that consume statistics retrieve these JSON files by "getting" them (i.e. pull) from the statistics producer.
	* Statistics about the reporting centre itself (i.e. monitoring of itself) should be placed in a JSON file called monitor.json
	* Systems that publish statistics about OTHER centres (e.g. the dashboard Markus has put togeather) should be placed in a JSON file called centres.json.
5. A corollary of the above item is that **it would be better if the common dashboards built by JMA, CMA and INMET were kept separate**.

## Summary

1. The new metrics by Markus can be integrated into the current monitoring system, specifically centres.json file. The purpose of the file is expanded to allow metrics about any centres.
2. In theory, the current daily update frequency does not need to increase for new metrics. But it is recommended.
3. The use of external server should NOT be forced.
4. No active involvement from DCPC/NC is required. All metrics can be calculated from the GISC side. This is likely a good way to proceed further at this stage.
5. It is recommended to have multiple systems either produce and/or consume the monitoring metrics.