Open Access to GTS strategy document

TASK: Develop an OpenGTS strategy document reflecting OCG-9 discussions, concerns and interests. Including possible national sensitivities regarding data distributions, “junk” data, clarity of mission, how does openGTS compare to “best copy” data sets, nodes for data insertion and capacity development.

* **Introduction**

During 2017, prior to JCOMM V, the JCOMM Observations Coordination Group (OCG) initiated a pilot project whose goal was to improve the way data is getting onto and off of the GTS. The goal was to make this process easier for users who were neither operational forecast nor national weather service users.

As described in more detail below, the pilot project was a success in a few key ways: 1) An improved workflow, based upon common data standards (NetCDF) and leveraging existing tools (ERDDAP data platform) was developed that provides an easier mechanisms for data providers to put their data onto the GTS; 2) Similarly, leveraging the ERDDAP data platform, the pilot project was also able to provide interoperable access to the near real-time observations that were pulled from the GTS; 3) the project was presented at several meeting venues, each time to enthusiastic response and requests for participation from data provides and users alike.

However, because the GTS is a well-structured and feeds operational processes, concerns were raised about the project. These concerns are discussed below in the “issues” section. The combination of user enthusiasm and operational concern brought forth the recommendation, at the 9th JCOMM Observations Coordination Group meeting, that a more defined strategy be developed to ensure that any further expansion of the Open Access to GTS workflow meet with the strict requirements of the GTS and operational community.

* **Background**

**PIlot Project Description**

A pilot project to take well understood data (ie, physical ocean data) from known platform types and inject the data in near real-time onto the GTS for distribution globally. Retrieve this data and other “essential” data[[1]](#footnote-0) from the GTS and make them available through interoperable web services for public access.

The goal of the pilot project is to prototype a workflow providing ocean data producers a simpler method of distributing real time data through the GTS infrastructure and ocean data consumers a simpler method of accessing that data.

**Pilot project requirements for data ingestion onto GTS**

* Data must be well understood physical ocean data (ie, temperature, salinity, meteo-marine data)
* Platform must be a well understood type and associated with known BUFR templates with which to encode the data for distribution onto the GTS
* Each platform must have a WMO ID assigned by authorized entity (i.e., JCOMMOPS)
* Platform observation data must be in recognized file format and be associated with complete metadata

**Pilot project process for ingestion of data onto GTS**

* Platform observations and metadata to be available in known format (netcdf, csv, xml)
  + Metadata requires WMO ID to be assigned to platform
* Data and metadata will then be ingested into ERDDAP server/framework
* Participating GTS institution (NDBC) will use the defined ERDDAP API to access the platform observation data and metadata
* Participating institution will encode the data and metadata using the appropriate BUFR template
* Participating institution will inject encoded BUFR message onto GTS
* This process will be automated for daily or more frequent data ingestion, as required

**Pilot project process for access data from GTS**

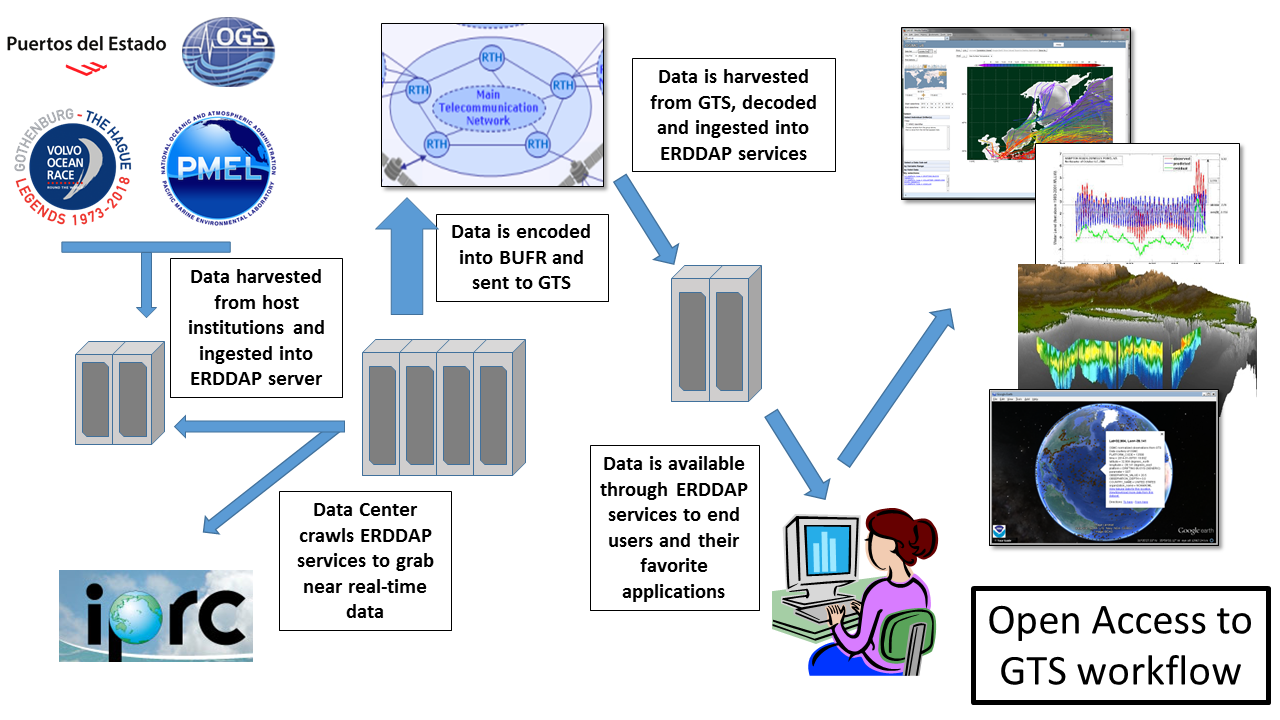
* The OSMC project already pulls data hourly from the GTS at the National Data Buoy Center and makes this data available through an ERDDAP framework
  + The OSMC project will retrieve the newly inserted platform data from the GTS, decode the BUFR data with the appropriate decoder and then add the data and metadata to the OSMC database
  + OSMC project will develop web-accessible tools and visualizations to illustrate the platform data in context with the other real time observations available from the GTS
* **Issues**

**Data ingestion**

* Data duplication
  + - Potential Recommendation: Ensure only one National Center is harvesting data for insertion onto GTS
* Data provenance
  + - Potential Recommendation: Connection to JCOMMOPS metadata services to provide as much metadata as possible
    - Question: What additional metadata should be recommended to be inserted into BUFR messages, if any?
* Data quality
  + - Potential Recommendation: Utilize verified QC procedures (QARTOD, Coriolis, ??) prior to providing data for insertion onto GTS.
    - Potential Recommendation: Add QC status/info/process to BUFR metadata template
* Participating institutions
  + - Potential Recommendation: Data center should be doing the transition/encoding to BUFR because BUFR tables change frequently
    - Potential Recommendation: With WMO, identify data centers will to participate using the defined Workflow
      * IMOS, Marine Institute of Ireland, BODC, US NDBC
* National ownership of data
  + - Question: Can participating data centers user headers that properly describe nation collecting observations, if it is from a nation that is not the same as national center?
* For “new” data platforms
  + - Potential Recommendation: Ensure proper data template used
    - Question: What to recommend if BUFR template doesn’t exist (ie, USV)?
* GTS routing headers used
  + - GTS routing headers are important to:
      * Establish responsibility of data that is put onto GTS
      * Provide some information about what is collecting the data
        + I would prefer this is in BUFR metadata rather than contextually contatind in a header such as “IOB”
      * Credit the nation providing the data
    - Question: Who is the authority on what the headers should be? WMO? Each National Center?
    - Question: Should this process have it’s own GTS routing header????

**Data Access**

* Data Provenance
  + - Again, the more medata provided the better.
    - JCOMMOPS and their global metadata would have to be a strong partner in this effort.
* RT vs Delayed mode data
  + - How to inform user that high-quality, delayed mode data now exists and these real-time observations should no longer be used
* **Next Steps**
  + Develop, in partnership with JCOMM, a Unmanned Surface Vehicle BUFR template to encode Saildrone data
  + Continue to place high-value Saildrone data onto GTS
    - Implement automated QC for Saildrone meteorological data
  + Identify National Data Centers willing to participate
  + GO-SHIP as possible partner for Open Access to GTS
  + Tide Gauge platforms as possible partner for Open Access to GTS



1. All available in situ observations from the marine environment as well as upper air observations are regarded as “essential” by the resolution 40 and thus they can be exchanged without charge and with no conditions on use [↑](#footnote-ref-0)