

i06.01(02) **Future WIS strategy**

(input to OPAG-ISS ad-hoc ... just my ideas)

Jeremy Tandy, Met Office

Status of WIS :: Metadata

- WIS DAR Catalogue comprises approx 150,000 metadata records
- Metadata records conform to ISO 19115:2003, encoded in ISO/TS 19139 XML format complying with WMO Core Metadata profile
- ‘Quality’ of metadata is mixed; guidance to metadata authors for 60+ categories of common data available soon
- Very ‘fine-grained’ metadata; dominated by “GTS” bulletins
- WIS intended to support all WMO Programmes; reality is that only WWW is well represented
- “one metadata record per TTAii+timestep” is damaging the search experience; e.g. when searching for temperature products, a user is deluged with bulletins
- With the retirement of WMO No. 9 Vol C1, the primary function of WIS appears to have become management of operational bulletins; e.g. notification of changes between operational centers
- WMO Core Metadata Profile regulation helps WIS Centers operate the WIS _system_ - it does not describe how to organize data nor make data easier to discover

Opportunities and threats :: Technology trends

- Increasing use of the Web as a data sharing platform; use of HTTP URIs (URLs) for identifiers, use of HTTP verbs for resource interaction (RESTful web services) – see W3C Data Activity
- Commercial infrastructure providers offer stable, secure and cost-effective ‘cloud’ hosting of virtualized computing resources (Infrastructure as a Service: IaaS)
- Commercial infrastructure providers offer ‘content distribution networks’ (CDN); a globally distributed network of proxy servers deployed in multiple data centers to serve content to end-users with high availability and high performance
- Search engines (Google, Bing, Yahoo etc.) remain the common entry point for consumer discovery of information; Spatial Data Infrastructures and Metadata Catalogues are used within expert communities
- Major search engines collaborate in development of schema.org to define structured mark-up for description of data resources on the web; search engines use [static] HTML pages annotated with schema.org markup to help index data published on the web
- DCAT becoming most widely used vocabulary for describing datasets and catalogue resources within open data portals; GeoDCAT-AP established as a geospatial application profile of DCAT for data portals within Europe.

- Faceted search is commonplace for exposing content in [large] datasets, e.g. World Ozone and Ultraviolet Radiation Centre uses facets of dataset, station, instrument, start/end date; open source implementations widely used, e.g. Apache SOLR, ElasticSearch
- OpenSearch API is widely used to enable searching within remote data holdings and forms a core element of the Open Geospatial Consortium (OGC) Catalogue Service (CSW) specification; OGC also provide standardized extensions for geographic and temporal search
- Data volumes created in numerical weather prediction (and other environmental sciences) continue to grow at a considerably faster rate than the performance of telecommunications networks; in the near future it will no longer be feasible to share data by sending it to remote users, instead users will need to select the subset that meets their needs (“data reduction”) or execute their queries and algorithms in close proximity with the data (“data gravity”)
- Infrastructure-level messaging services, e.g. RabbitMQ, and messaging protocols, e.g. AMQP, have become the de facto mechanism to share event driven notifications and alerts; permitted message sizes easily exceed the size of typical TDCF bulletins which are typically < 5KB; proven scalability far exceeds traffic volumes of the GTS, e.g. Twitter
- Applications such as Dropbox and open source ownCloud indicate that file distribution services have become commoditized and are no longer the domain of specialized applications such as GTS message switching
- Use of third-party identification services, e.g. using protocols such as OAuth2, to establish the identity of a user interacting with a web service is commonplace.

■ ■ ■