USGS Trusted Digital Repositories

ELT Champion: **Randy Orndorff, Acting Associate Director, Office of Science Quality and Integrity (OSQI)**

Sponsor: **USGS Data Policy Team (represented by Keith Kirk, Hydrologist, OSQI)**

Issue/Challenges: The U.S. Geological Survey (USGS) has developed a plan and associated policies to increase public access to scientific publications and digital scientific data resulting from research funded by the USGS. The plan, entitled "Public Access to Results of Federally Funded Research at the U.S. Geological Survey: Scholarly Publications and Digital Data" which is effective October 1, 2016, describes how the USGS will meet requirements established by Office of Science Technology and Policy (OSTP) and Office of Management and Budget (OMB). One requirement states that supporting digital data (those data used to support the scholarly conclusions and final project data) and final project data for USGS funded research are to be made available at no cost to the public, accessible from a USGS or USGS-approved Trusted Digital Repository (TDR), and discoverable by submitting the complete metadata record to the USGS Science Data Catalog (http://data.usgs.gov).

Moving forward, there are challenges to implementation:

1. The OSTP and OMB directives provide no additional funding to accomplish these ends and hence we must maximize efficiency to minimize impact (both time and cost) on our science centers and programs;
2. In the current USGS highly distributed landscape there is only one USGS server that meets TDR requirements (EROS Data Center) but many of USGS's 5,000 servers providing data and publications do not and will need to meet TDR requirements, what should be done about this;
3. There are short-term (startup) and long-term (staffing, maintenance and operation) costs that must be considered for every server that is--or is designated to become--a TDR;
4. Datasets required to be publicly accessible may range in size from kilobytes to multi-terabytes *(Storage of many such large datasets, cumulatively reaching into petabytes and requiring significant bandwidth to achieve desired I/O rates, present significant infrastructure and data management resource challenges)*;
5. How do we deal with existing servers that do not meet TDR requirements?

Background Material/Resources:

- USGS Web Site "Public Access to Results of Federally Funded Research at the U.S. Geological Survey" which includes a link to the 25-page USGS public access plan where

sections 3 and 9 are most relevant and also offers a Flash-based module summarizing the plan.
- Standards for USGS TDRs (USGS intranet FSP web site)

Charge/Expectations: Develop a consensus document recommending the best approach to meeting OSTP and OMB requirements for making USGS data publicly accessible from TDR(s). These ideas may include:

- Carefully weighing the cost and benefits, strengths and weaknesses of science center specific, mission area specific, or an enterprise solution for developing and maintaining TDRs ();
- A communication strategy to explain your preferred solution;
- Buy in – what is the benefit to the USGS of your preferred approach and how do you demonstrate this?
- Accountability – what does it look like?
- How should this be funded?
- Role of current and future 'trusted' external (USGS-approved but not managed) data repositories in terms of rationale, reliability, desirability, costs, and potential pitfalls.

Please think broadly about the options to make this document the best it can be.