**DATA–CURATION-PUBLICATION**

**Introduction**

The ICDPs practices regarding data management and publication have distinct advantages and disadvantages.

ICDP is characterized by a highly individual, autonomous, and self-organized local project management style, which is in contrast to more platform-driven, centralized service approaches (e.g., IODP). The question arises for ICDP and associated clients: “How do we strike a balance between these two approaches?”, and with that the next pressing question is asked: “Should ICDP look for new ways to conduct its projects?”

ICDP’s current organizational framework defines the options for data management and publication procedures: Funding and support are closely related to the operational project phase, though data management and publication support services do not end at that point. Since ICDP has only loose controls on a project, (e.g., suggested start and end dates), multiple simultaneous projects can create logistical challenges and high pressure on OSG members. This situation can lead to disruptions in service for ICDP clients. In addition, little or no onsite presence of OSG staff during drilling and the dependence on cooperating external facilities to accomplish project research goals further complicate the conduct and execution of a typical ICDP project.

With 20 years of ICDP support for drilling projects and a growing pressure for project activities, there is an urgent need to define priorities for future ICDP modes of project support regarding data management and publication. Perhaps at the highest level, ICDP must determine whether to support as many projects as possible, at a possible cost of reduction in quality of support; or to focus on providing the highest possible quality of support, but for a limited number of projects.

During the workshop assignment on DATA MANAGEMENT, CURATION and PUBLICATION, a group of 24 individuals discussed and summarized their thoughts on defining future-binding directives and recommendations on these pressing questions, and how improving corresponding and existing ICDP practices, workflows and policies can be developed and realized regarding 1) Data and Sample Management in the Framework of ICDP; 2) Standards and Tools; and 3) Data Sharing and Publication.

The following Action Steps were identified and are summarized as follows:  **Subtopic 1:** Data and Sample Management in the Framework of ICDP (Discussion moderated by Dr. Jens Klump, AUS, CSIRO)

* Action #1: Develop **Common and Obligatory Data Management Plans** to be associated with all ICDP proposals. The plan should include **Data** and **Metadata** acquisition and dissemination, **Moratorium** agreements detailed in an **MoU**, **Open Access** policies, and best-practices/policies for data acquisition and data/sample **Curation** to accommodate the highly diverse workflows applied among different ICDP projects.
* Action #2: Establish a working group experienced in both data-publication- cyberinfrastructure and drilling and coring workflows and practices to **streamline a unified data curation**; include **use of mobile systems**, where applicable,to acquire and access data. This requires changes to the currently existing Drilling Information System (DIS) to accommodate the use of mobile devices (mobile phones, tablets, etc.). Leverage a cooperative network of international partners (e.g. CSDCO and LacCore/USA, CSIRO/AUS, Core Repositories in Germany, and elsewhere) for open source software development.
* Action#3: Enhance ICDP’s **Web Portal** to make access to metadata and research data easier. This important aspect for all ICDP data-related issues requires a more permanent ICDP staff base to manage and maintain currently existing data dissemination tools and add modifications as required.
* Action#4: Identify **suitable data formats** and **create** **machine and** **user interfaces** which allow connecting with other external data management systems (e.g. the evolving “Open Core Data” architecture of CSDCO/LacCore, IEDA, and IODP-JRSO; other IODP; other organizations) to acquire and merge data obtained during the initial field campaign, including **Visual Core Description** data.
* Action #5: Improve on **Project MoU’s** and **licenses** (incl. with other agencies) and **enforce** them through centralized oversight by ICDP while allowing adjustments to planning and execution for each specific ICDP project; establish clear **guidelines**, **milestones** and **deliverables** for the “What”, but not necessarily the “How”, of data -and sample management plans throughout the entire lifecycle of an ICDP project; the “What” is overseen and managed through a project-assigned ICDP staff scientist together with OSG’s data management team.
* Action #6: Discuss planning and project management in its initial phase with experienced PIs from other successful projects and thereby follow “Blueprint Projects” (e.g. COSC; Koyna); foster technical evolution of DIS/Web tools which serve all stakeholders (i.e. the entire science community with an interest in ICDP drilling data).
* Action #7: Data management plans must take into account national and international **Open Access** data policies. Post-workshop discussions with respective sub-topic leaders have already led to a stronger collaboration between ICDP and other respective scientific drilling organizations (i.e., CSDCO/LacCore and CSIRO) with the goal to match and merge data, tools and management plans.

**Subtopic 2:** Standards and Tools (Discussion moderated by Dr. Anders Noren, CSDCO/LacCore, USA)

* Action#1: Define, create and maintain **Accredited Repositories** with standardized workflows, data, and data formats, and thereby following best-practices from well-established repositories in Germany (MARUM, SPANDAU, GEOMAR) and elsewhere (KOCHI/Japan; GCR and LacCore/USA).
* Action#2: Establish and maintain sufficient **Staff Resources** to support ongoing projects, thereby ensuring proper implementation of standards and use of tools. Personnel and contractors from partner institutions (e.g. CSDCO/LacCore and others) could be entrained to support projects as needed to provide sufficient staff capacity at times of greatest demand for support. A “shared contract” personnel base was discussed in the aftermath of this workshop as means to optimize sharing of personnel resources across different organizations (e.g. GFZ, US CSDCO).
* Action#3: Sufficient **Instrumentation Resources** to support ongoing projects (following the same recommendations and conclusions as above).
* Action #4: Define a **Repository Accreditation Process** to establish federated repositories around the world that follow a commonly accepted sample material storage, processing workflows and distribution service to the scientific community. It was recommended that there should be criteria for evaluating the suitability of such a repository for ICDP projects, perhaps analog to [re3data.org](http://re3data.org) criteria established for accreditating data repositories.
* Action #5: Depending on the science community’s call for “Full Service” (e.g. containerized labs deployed to the drill site for full core processing on site; dedicated repository service; and others — the group recommended against this approach) vs. “Lightweight Service” (e.g. variations on the status quo — the group embraced this approach), ICDP must provide and accommodate sufficient **staff resources**, perhaps utilizing staff from other organizations in periods of greatest demand,for technical and scientific service and support of each ICDP project on a more permanent basis if possible. As consistent personnel perform this work routinely, a strong **institutional memory** is created, which can lead to a greater degree of consistency among the various and diverse ICDP projects and ease of data integration across projects.
* Action #6: Maintain an **Instrumentation Pool** on a shared-cost base (as recommended for “personnel”), which pertains to lending, exchanging, and possibly purchasing certain instruments of common usage for typical ICDP projects (e.g., borehole logging tools, and/or core images/MSCL scanners in mobile operation modes). Since instrumentation often requires special care and knowledge of workflows and trouble-shooting, this aspect of extending a global ICDP-service also warrant the expansion of a trained personnel pool.
* Action #7: Define a solid and explicitly dedicated budget within the project for data and sample management and publication procedures, i.e. ICDP funding that is reserved and not accessible during drilling operations.
 **Subtopic 3:** Data Sharing and Publication (Discussion moderated by Dr. Kirsten Elger, GFZ, GER)
* Action#1: Follow a well-established **data publication workflow,** which defines clearly **Basic Data** among the science community prior to project start.
* Action#2: Make an **Operational Report (OP) mandatory** within a well-defined operational moratorium period which includes core opening, initial measurements, and lithological descriptions. This must take place *before* the science moratorium (i.e. the actual scientific evaluation) starts, associated with a sampling party and/or post-drilling workshop. **Basic Data** sets should be added as a mandatory supplementary publication to expected research publications. Both are feeding an overall **ICDP** **Metadata** **Catalogue.**
* Action#3: Establish **IGSN Registration** as one component of best practice. However, the allocating agent and methods used to register and track the IGSNs must be allowed to vary. The DIS workflow and GFZ allocating agent is a readily-used option for ICDP projects without existing workflows and allocating agents in place.
* Action#4: Clearly define “protection plans and licenses for data generation and origin”, i.e. **OPEN ACCESS** policies and practices to control the access to and usage of data.
* Action#5: Develop a system to integrate new datasets and publications generated during the post-drilling phase of a specific ICDP project with the already existing project database, publication list and **ICDP** **Metadata** **Catalogue**.
* Action #6: Discuss, refine and test the **IGSN granularity** (incl. sub-samples, pieces etc.) throughout upcoming ICDP projects, which use the DIS as auto-assigning IGSN tool during the data entry process.
* Action #7: Overcome conflicting requirements between MoU–based restricted access to the **Basic Data** during the moratorium period and recent requests by Copernicus to make them available when publishing the initial paper in the Scientific Drilling Journal (following the “COPDESS Statement of Commitment”). It has been noted that variable and sometimes conflicting “rules and regulations” exist regarding the data sharing and time of data protection even within the ICDP structure. This requires immediate attention to avoid confusion among the scientific community. An adjustment of the MoU will be necessary to overcome such conflicting levels of understanding. A possible solution could be to publish the **Basic Data** as data publication synchronous to the publication of the initial article in Scientific Drilling, hereby applying an embargo period (which correlates with the Moratorium) to sensitive data (see COSC-1 example, <http://doi.org/10.1594/GFZ.SDDB.ICDP.5054.2015>).
* Action #8: Tie the publication of the OP Report to a funding pay-back obligation if not delivered in time. Conversely, in the case of fulfilling this requirement, ICDP commits to sponsor the funding of a post-drilling workshop (WS) as progress incentive and bonus to the entire project portfolio.
* Action #9: Extend ICDP’s services, such that a harmonized data acquisition and sample registration schema across the various database systems of various institutions (e.g., CSDCO/LacCore ”Open Core Data”; AWI/Pangaea; IODP/LIMS, etc.) is established and maintained.
* Action #10: Define a **Feedback System** with and for Science Team Members (STMs) to funnel their post-drilling research and publications into ICDP’s web portal information dissemination system; the question here is what stimulus and incentive can be created to provide this information to ICDP beyond the active drilling phase of the ICDP project

Suggested **Action Steps** are in alignment with “ingredients” recommended by Fran Berman (UCSD) for successful data management procedures:

* Economic support (funding)
* Policy
* Political support and public acceptance
* Stakeholder support and responsibility (social norms)

**Costs** for such **Action Steps** include:

* Technical developments (Web Portal; Mobile Apps)
* Human Resources (web data manager; staff scientist: additional programmers/software developers)
* extra (yet worthwhile) ICDP funding for post-drilling WSs
* same applies for data sharing procedures and corresponding hard/software developments