If it's important to rescue at-risk cultural and scientific data, shouldn't we also preserve the capabilities for doing so?

IPDRC

Initiative for the Preservation of Data Rescue Capabilities

Not only are important data collections at risk of loss, but so also can be the necessary technical and organizational tools for the "rescue" process. Perhaps it's a university researcher on the verge of retirement who once recovered arcane datasets. Or a government project for which funding has suddenly lapsed, and staff are relocated. Or a commercial entity that had done similar work, now closing. Rare equipment, software, skills and related documentation may be discarded. Future data rescue efforts can be helped by <u>not having to re-invent those wheels</u>.

A great project, with many lessons to be learned is the now-famous "McMoons" effort for the *Lunar Orbiter Image Recovery Project* (see http://tinyurl.com/mcmoons)[1]. Dedicated people, led by Mr. Dennis Wingo, did great work to resurrect rare capabilities and then rescue the needed information. But that work may not have been possible had not an earlier project manager, Ms. Nancy Evans, *on her own initiative*, saved vital equipment in her garage. Note that Mr. Wingo, was yet again a good example for us by passing the equipment along to Library of Congress to ensure its preservation. *We're proud to say that he has endorsed the goals of the IPDRC movement.*

Suggested Goals and Activities

<u>Collecting details</u> of rare capabilities and projects across government, academia, and industry, is a critical role. A central repository for this sort of information will be very helpful. It should make it easy for people planning projects to search across a broad spectrum of sources for prior efforts with matters in common to their own. It should also allow for and encourage submission of information about other data rescue projects/capabilities. One important category would be *"CAROL"* – capabilities at risk of loss—which should receive special attention.

<u>Staying in touch</u> and making sure IPDRC has contact points with key custodians will help ensure against surprise disappearance. Also, reaching out to other departments and institutions to ensure the description is as complete as possible. Working with related interest groups such as those at CODATA, RDA, NDSA, etc., and using social media platforms would surely be helpful in promoting IPDRC goals.

<u>Finding homes</u> for important resources at risk of loss. By posting such situations online and staying in touch with institutions that may have an interest, hopefully new stewards will be

---

[1] There's also a good summary at https://en.wikipedia.org/wiki/Lunar_Orbiter_Image_Recovery_Project

found. In some cases the new homes may have staff and funding for things like upgrading/documenting old data rescue software and equipment.

File format documentation in many instances, will be a key precursor to a rescue capability. There are very helpful online sources for that sort of information. Even so, some types of files have *fallen through the cracks*, as it were. Years later, stashes of data in that format are realized to have significant value. IPDRC would reach out to authors of the software that created the files, or to others who were later forced to crack the format. It would be best to work in cooperation with existing file-format registries. We have some further discussion on this and a real-life example in the Appendix.

Questions about Further Goals

Husbanding *CAROL*[2]? When new homes for at-risk capabilities can't be found, should IPDRC itself provide such housing and stewardship?

Perform some DR projects? First efforts should be to help the institution requiring the DR work to find any pre-existing and otherwise appropriate capabilities. Best case is that the found entity can assist the requestor in replicating the capability, or perhaps assist in performing the project. But in cases where that's not realistic[3], (and if the answer to the "husbanding" question above is "yes") should the IPDRC organization itself offer to help with the DR work?

Suggested Long Range Activities

Ongoing inventory outreach. Continuing and improving the collection of details and making them available to all.

Elder-Geek outreach. Looking for key players over time from government, academia and business and finding ways to pass their skills, tools and experience along.

Young-Turk outreach. Recruiting staff with modern know-how, talent and fresh perspective, who also appreciate our early-digital legacy and will gradually assume IPDRC leadership.

Cooperation with institutions around the world with resonating interests.

Please Help Move This Forward

If you have any suggestions regarding goals, tactics, possible funding sources or other ways of promoting this group, they would be greatly appreciated. Thanks for your time.

CODATA Data at Risk Task Group - IPDRC committee
[committee founder: Chris Muller – archive.raider@gmail.com]

.

---

[2] Sorry, addicted to puns and such.
[3] Note that commercial entities generally provide the best and most cost-effective data recovery solutions. But in the case of scientific and cultural information in older/arcane formats, the marketplace may not have created or retained the specific capabilities.

Appendix

(Some supporting details moved here to un-clutter the main document)

The File Format Issue. Of course, it can be to the benefit of a software company to keep certain matters close to the vest. If they upgrade their products for new environments, or even go out of business, the descriptions of the older file structures are often not available to the public. If it represents a large enough marketplace, third parties will often hack their way through the files and create conversion tools. But some <u>fall through the cracks</u> for a variety of reasons.

Some situations: (a) the originating company may have provided free file upgrade services, reducing market justification for third-party conversion tools. Later, even though the file format documentation is of less value to them, they still don't make it public; (b) unique conversion tools may have been developed by another company. They must have either "hacked" the format or gotten a copy of the creator's documentation. If it's no longer of proprietary importance to them, perhaps they will pass it along to IPDRC.

An example of the (a) situation involved statistical software SPSS. File formats prior to the mid-1980s were quite different than current ones, and current versions cannot import those files. At the time, SPSS helped clients to migrate the files and no third-party conversion tools had been created. Decades later, after a number of "old professor stashes" on tape reels with valuable data cropped up, someone was asked to convert them. Neither SPSS (now a branch of IBM) nor the primary original author could provide any documentation on that older file format[4].

Several other examples are available if desired.

In situations where data conversion capabilities don't exist, the work of creating those can be greatly reduced by ensuring documentation is preserved.

Perhaps IPDRC can play a role, working with existing file format registries to help prevent more items from falling through the cracks.

---

[4] Somebody spent many hours hacking away on the older format. The file description, along with the resulting conversion program (and many others) are expected to be donated to IPDRC once formed.