Data Rescue

Margaret Janz

University of Pennsylvania

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium

Part of the Scholarly Communication Commons

This work is licensed under a Creative Commons Attribution 4.0 License.

Repository Citation


This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Intro - my name is Margaret Janz, I am the scholarly communications and data curation librarian at University of Pennsylvania. I’m also a co-founding member of the Data Refuge project (along with Laurie Allen, Bethany Wiggin, and Patricia Kim)
History of Data Refuge in a nutshell: PPEH fellows were worried, we decided to hold a data rescue event, worked with EDGI and others to create workflows for copying data and nominating pages for the End of Term Harvest. Also Storytelling - which I'll talk about more in a hot minute.

Around 50 data rescue events that followed some version of our workflow then took place around the country. Great work was done by all who participated and we learned a LOT about "rescuing" data (gonna talk about that term too). And it raised a lot of questions for others.
Endangered Data Week was established by another group, predominately by Brandon Locke and Rachel Mattson. The focus of the week and the related events was on all kinds of data that might be at risk.
The idea of "rescuing" data was also not new when DR started. RDA has had a working group on data rescue since at least 2010. Rescuing data means a great many things to many people - as it should.
So what might be in need of rescue? Data. Audio. Video games..........literally anything. We rescue what matters to us - as individuals, as institutions, as communities, as ragtag groups of misfits. Figure out what's important to you - some of you probably already have thoughts on this and that's why you came to this session.
The other consideration when considering a data rescue, is why the data need rescuing. There was some misconceptions when DR started that we meant for it to imply that the stewards of govt data were bad at their jobs. They're EXCELLENT at their jobs. Those data are backed up beautifully in most cases. But they're backed up on federal servers, being managed by federal employees. The risk we were concerned with was about access.

The access to these data was at risk because of political factors, which is just one way data can be at risk. That applies to nongovt data, too. Org politics and leadership priorities can put data at risk in the same way.

Technical risks include file formats, software and hardware obsolescence, bit rot, regular rot of physical items, items made from unusual materials. A lack of documentation can also put data at risk; if no one know the data exists or has not context to make sense of the data it is hard to protect it.

How the data is used, and who it's used by can also make data vulnerable - as well as provide protections. The Data Refuge Stories project helps show broad usefulness of data by connecting it to everyday life and human stories.

Legal protections may exist for some data - like retention laws or policies. How or if those policies are enforced can be also put data in risky situations or protect them.
There may also be policies about data deletion, but those policies might not take into account all the users of those data.
Collaboration and partnerships are also key to data rescue. These are new ideas that need new workflows with a variety of stakeholders. We need new solutions that take each others' expertise and needs into account. We have to find new ways to work together. It's not easy to imagine, but I'm going to ask you to do that now.
Attendees of this session were asked to think about these questions and fill out the worksheet provided (with the same questions) with some data they are concerned about in mind. Then we shared some thoughts about rescuing data and help each other think of ways to go about rescuing the data in ways the writers had perhaps not considered. We can help each other so much!