

The Open Data Repository's Data Publisher. N. Stone¹, B. Lafuente², R. T. Downs², D. Blake³, T. Bristow³, M. Fonda³ and A. Pires², ¹Open Data Repository, Gray, ME, USA (nate.stone@opendatarepository.org), ²Geosciences, University of Arizona, Tucson, AZ, USA, ³NASA Ames Research Center, Mountain View, CA, USA.

Introduction: Traditionally, research data are published only as part of a conference or paper publication. The data that led to the publishable result often contains many intermediary files, input files for computer software, or raw data from scientific instruments that would normally never be published. Truly examining methodologies used by other researchers or learning from their mistakes is practically impossible.

Additionally, some research avenues never lead to a publishable result. These “negative” results are never shared resulting in duplicative research. Publishing negative results becomes possible when a low barrier to publication exists.

From experience with the RRUFF database [1], creating an easy-to-use, publicly searchable database, can quickly lead to many citations and significant value for researchers and the public.

The Open Data Repository's Data Publisher system lowers the investment requirement for data publication and makes it practical for smaller, distributed groups that may not traditionally have the budget or time to publish archives containing nearly any data they feel worthy of publication.

Objective: The Open Data Repository (ODR) Data Publisher software system aims to provide researchers with a simple tool to publish any or all data related to their research. Once published, unique, citable URLs can be generated for publication reference and all changes to the data are tracked over time so a living, yet citable, database can be maintained.

Data Publisher makes it possible to share data publicly, to invited collaborators, or internally so that research integrity can be maintained.

Serving individual researchers and small teams, the Data Publisher system will allow for easy maintenance, sharing and collaboration, and long term archiving of scientific data.

Technology and Capabilities: Data Publisher is a fully open source software platform created to run on a traditional LAMP (Linux, Apache, MySQL, PHP) server. Small to mid-size research teams will be able to easily administer an archive of their data using inexpensive, off-the-shelf server infrastructure and the easy-to-install software package.

Once installed, the software will optionally register with the ODR Hub to facilitate easier sharing with outside collaborators and ontology sharing for semantic web publication.

Users can easily create new data repositories in their local store for their different data sets. Creating a data repository is a drag-and-drop procedure facilitated through a web-based forms designer. No direct interaction with the underlying MySQL database is required. The user can select and add different field types to their data set and modify the design at any time to include new fields or file storage areas. The user can then start creating data records that describe individual pieces of data.

A plugin system allows data sets containing XY data to be graphed and plotted to the HTML output. Plugins for rendering chemical formulas and for displaying images in galleries are also bundled with the system. The plugin architecture is extensible and third-party developers can write plugins that will operate on data and produce custom displays or interaction.

As development continues, the software will include semantic web standards such as RDFa (Resource Description Framework in Attributes), OWL 2 (Web Ontology Language), or other technologies to facilitate semantic data queries and publication.

Current Status: Data Publisher has been in development for over two years and is undergoing beta testing by a number of research groups. The software is patterned after the RRUFF database (<http://www.rruf.info>) [1] and is currently testing a complete copy of that data to ensure functionality. The RRUFF database will be moved to the Data Publisher platform to provide the flexibility it needs to grow in the future. Additionally, NASA's CheMin [2] data from the Mars Science Laboratory is being published on the system as a proof of concept and **The Astrobiology Habitable Environments Database (AHED)** [3] will host data from a number of researchers in the astrobiology community.

References: [1] Downs R T (2006) *Program and Abstracts of the 19th General Meeting of the International Mineralogical Association in Kobe, Japan*. O03-13 [2] Blake D. et al. (2012) *Space Sci Rev*, 170, 341-399. [3] Lafuente B. et al. (2015) *AbSciCon 2015*, submitted.

Acknowledgments: We gratefully acknowledge the support for this study by the Science-Enabling Research Activity (SERA), and NASA NNX11AP82A, Mars Science Laboratory Investigations and University of Arizona Geosciences.