

What If We Reconsidered How We Ask Scientists to Share Their Data: When FAIR Meets Crowd-Sourcing and Nudge Theory

Mike Smit¹, Alexandru Ianta¹, and Arin MacNeill¹

¹Dalhousie University

November 28, 2022

Abstract

Journals, funding agencies, and researchers are more frequently expecting manuscripts to include links to shared research data. Effective data sharing requires that data be findable, accessible, interoperable, and reusable (FAIR), and is thus predicated on establishing a common understanding on how to communicate: data exchange standards, common data formats, controlled vocabularies, and a communal data repository. When conducting research, we still communicate in shorthand that is effective for everyone on the team who understands our context, but is lost when data is shared in the absence of that context. “Water temperature” means only one thing to my research team, yet can mean dozens of things outside of that context. Data sharing is thus an exercise in sharing not just the data, which is typically readily available, but also the context of that data, which requires additional effort. This effort is one of the barriers to sharing data. We’ll describe an alternative model for accepting data to a repository: the immediate ingestion of data regardless of its metadata quality, then behavioural nudges and crowd-sourcing features that ensure this data meets appropriate standards prior to publication. We’ll show a work-in-progress prototype software tool that supports this alternative model, capable of accepting and standardizing a research data set to use CF conventions and ISO 8601 dates.

What If We Reconsidered How We Ask Scientists to Share Their Data: FAIR, Crowd-Sourcing, and Nudge Theory

PRESENTER: **Mike Smit**
Dalhousie University

Alexandru Ianta, Arin MacNeill

INTRO

Sharing research data openly is good for everyone, but it can be an annoying process. Too often it combines the low-level detail of metadata standards with our universal love for entering data into multi-page web forms. We propose to invert the current data submission process: get the data first, and then use crowd-sourcing and persuasive technology, plus easy-to-use tools, to help researchers document it properly.

OUR APPROACH

1. Upload your data file (web form, or use right-click “Send to...” feature on desktop).
2. Tool support makes it easy to:
 - Map your own variable names to vocabularies (e.g. CF conventions)
 - Add metadata blocks (e.g. “add all the required Darwin Core fields”)
3. The submitter completes micro-tasks, each with a matching motivation:
 - Add a written description to show up in more search results!
 - Add 2 more fields to be issued a DOI!
 - Add your other authors so you can submit to Figshare!
4. All tasks can be crowd-sourced, with owner verification before publication.
5. Learn over time how researcher-supplied variables map to standard vocabularies, and automatically convert.
6. Release tool as FOSS to replace current data repository submission systems.

Strengthen your profile

Let's update your headline to match your current position

That way, people can easily find and connect with you

Not now

Update headline

Social media sites are brilliant at gently extracting your data.

We can reduce barriers to sharing data by improving the user experience of standardizing data, leveraging crowd-sourcing and gentle nudges.



Take a picture to **visit the prototype**. It is under active development so might break at any time! 😊

WORKING PROTOTYPE

Ocean Science

Data Integration Tool

Connect with your ORCID ID

Sign in

Sign up

Samples	Community Matching				
Date & Time(UTC)	2013-11-07 16:53	2013-11-07 17:23	2013-11-07 17:53	2013-11-07 18:23	2013-11-07 18:53
Latitude	4433.5220N	4433.5180N	4433.5180N	4433.5200N	4433.5180N
Longitude	6332.7210W	6332.7220W	6332.7250W	6332.7260W	6332.7240W
Avg Sea Surface Temp(°C)	9.04	9.02	9.01	9	9
sea_surface_temperature					
sea_surface_temperature					
sea_water_temperature					

Halifax Harbour Buoy Data

Source File: Halifax_Buoy_trial.csv

Uploaded: Mon Oct 21 17:51:57 GMT 2018

License

CC0

Item Type

Dataset

Categories

Physical Oceanography

Chemical Oceanography

Ecosystem Function

Keywords

Enter a value

Description

Enter a value

Authors

Enter a value

Figshare Publish

This set contains all the fields which are required for a document to be published on Figshare.

Title

Multiple Value

Authors

Support

Categories

Restricted Value

Keywords

Set

Description

Nested Values

Item Type

Multiple Value

Authors

Support

Item Type

Restricted Value

Date & Time(UTC)

2013-11-07 18:53

Search

Community Results

Date Time

ISO 8601

ISO 8601 Data elements and interchange formats – Information interchange – Representation of dates and times is an international standard covering the exchange of date- and time-related data. It was issued by the International Organization for Standardization (ISO) and was first published in 1988. The purpose of this standard is to provide an unambiguous and well-defined method of representing dates and times, so as to avoid misinterpretation of numeric representations of dates and times, particularly when data are transferred between countries with different conventions for writing numeric dates and times.

Submitted by 0000-0003-0278-0620

Confirm Match

Downloads

Original File

Integrated File

Publish

Figshare

CEAN
FRONTIER INSTITUTE

DALHOUSIE
UNIVERSITY