A data publication workflow tool

Deborah Smith
Ajinkya Kulkarni
Introduction and Motivation

• This is a centralized data publication workflow tool for managing dataset publication at GHRC from acceptance of data from a provider through to publication to the end user.

• It is adapted from a tool (SAuS) originally developed at ORNL DAAC and shared with GHRC. We adapted the tool to meet our needs and upgraded it to work in a single coherent system based on Drupal 8.

• This tool provides a centralized system for Data Providers, the GHRC Data Management Group and the GHRC DAAC Managers. It automates many the steps required for dataset publication to reduce redundancy and repetition of effort.
* adapted from ORNL DAAC’s Swimlanes
Key Components of the Data Workflow Tool

- Various data entry forms for data providers and GHRC data managers
- Views/dashboards for data managers and data providers
- Tracking and viewing of dataset workflow progress at each step
- Autofill information between forms
- Email notifications and email aggregation at one central place
- Integration with GCMD Keywords List (auto fetch/update keywords)
- REST API interface to GHRC catalog database
- Management of users and user permissions
- Management of DOI and link management (edit, validate, submit, etc.)
For the Data Provider

**Forms**
- Archival Interest Form
- Questionnaire Form

**Views**
- Check publication status of own data
- Edit/clone own records

For the GHRC Data Management Group

- Metadata Records
- User Guides
- Instruments
- Platforms
- Organizations
- Projects
- Publications Tracking

- Track publication status of all datasets
- View email communications
- Quickly link to various entries and submissions
Step 1: A data provider submits a request for submitting data to GHRC for archival
Step 2: If accepted for the archival at GHRC, the data provider fills out questionnaire form
Step 3: GHRC personnel are then automatically assigned to publication workflow tasks and email notifications are sent out.
Step 4: GHRC Data Management Group lead creates an initial metadata record.
Step 5: A software developer works on configuring ingest/archival software
Step 6: A science team works on the documentation and completes the metadata record
Step 7: An ingest coordinator uses the tracker to check the progress and decides when to publish the dataset.
GHRC DAAC Managers can use the central view to track the progress of all datasets currently in the workflow process

Filters:

<table>
<thead>
<tr>
<th>Published status</th>
<th>- Any -</th>
<th>Filtered by</th>
<th>- Any -</th>
<th>Filtered by</th>
<th>- Any -</th>
<th>Filtered by</th>
<th>- Any -</th>
<th>Filtered by</th>
<th>- Any -</th>
<th>Filtered by</th>
<th>- Any -</th>
<th>Filtered by</th>
<th>- Any -</th>
<th>Filtered by</th>
<th>- Any -</th>
<th>Filtered by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Producer (PI)</th>
<th>Authored on</th>
<th>Initiated</th>
<th>Metadata</th>
<th>Processing</th>
<th>Catalog</th>
<th>Documentation</th>
<th>Published</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hurricane and Severe Storm Sentinel (HS3) Hurricane Imaging Radiometer (HIRAD)</td>
<td>Cecil Daniel</td>
<td>2016-09-13</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hurricane and Severe Storm Sentinel (HS3) High-altitude Imaging Wind and Rain Airborne Profiler (HIWRAP)</td>
<td>Gerald Heymsfield</td>
<td>2016-09-13</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Hurricane and Severe Storm Sentinel (HS3) Global Hawk High Altitude MMIC Sounding Radiometer (HAMS)</td>
<td>Bjorn Lambrecht</td>
<td>2016-09-13</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Email Communications

A centralized email logging system allows for Data Management Group and DAAC Managers to view the external as well as internal communications for any particular dataset.
We make use of Drupal taxonomy to manage various GCMD keywords, tags, and to autofill values at a centralized place.

* retrieved from http://gcmd.nasa.gov/learn/keyword_list.html
Demo and Release Schedule

- Development version available - [https://ghrcdev.nsstc.nasa.gov/data-publication/](https://ghrcdev.nsstc.nasa.gov/data-publication/)
- Demo User Login Information
  - User: ghrc-demo-user
  - Password: ghrc-demo-user

- Once logged in, check out the Archival Interest Form and the PI Questionnaire Form.
- Both links to these forms located on [https://ghrcdev.nsstc.nasa.gov/data-publication/](https://ghrcdev.nsstc.nasa.gov/data-publication/)

Anticipated Release Schedule

- Test version – 10/30/2016
- Operational version – 3/31/2017
Discussion

1. Has any terminology been used in this form that you do not understand? If so, please explain.
2. Do you feel any pertinent dataset information is missing from this form? If so, please explain.