

NCAR DASH Search and Linked Data: Investigation and Implementation of Similar Assets

Prasil Mainali, Joshua Roberts Jones

Interns, SIParCS

Sophie Hou, Eric Nienhouse, Nathan Hook

Mentors, SIParCS

NCAR

August 1, 2018



Table of Contents

- Introduction and motivation
 - The Digital Asset Services Hub (DASH)
- Method
 - How DASH Works
 - “More Like This” Implementation
 - Solr: “More Like This”
 - Sample Solr MLT Response
- Results
- Future Work
- Lessons Learned
 - Using Agile Scrum
- Acknowledgements
- References

The Digital Asset Services Hub (DASH)

The screenshot displays the DASH search interface. On the left, there is a sidebar with a map and filter options. The main area shows search results for the query 'Temp'. The results are ordered by 'Best Match' and show 37 resources found. The first three results are highlighted with a dark border:

- Latnjajaure Site, CLIP 4–Temp 2 Community Data [Alatalo]**
This dataset contains CLIP 4–Temp 2 Community data from the Latnjajaure site, Sweden in 1995, 1996, 1997, 1999 & 2000. The Community Level Interaction Program (CLIP) data...
[dataset](#)
- Latnjajaure Site, CLIP 4–Temp 1 Community Data [Alatalo]**
This dataset contains CLIP 4–Temp 1 Community data from the Latnjajaure site, Sweden in 1995, 1996, 1997, 1999 & 2000. The Community Level Interaction Program (CLIP) data...
[dataset](#)
- Latnjajaure Site, CLIP 1–Temp 2 Community Data [Alatalo]**
This dataset contains CLIP 1–Temp 2 Community data from the Latnjajaure site, Sweden in 1995, 1996, 1997, 1999 & 2000. The Community Level Interaction Program (CLIP) data...
[dataset](#)

The sidebar includes a 'Filter by location' section with a map, a 'Refine by:' section, a 'Publication Year' filter, and a 'Keywords' section with a search box and a list of keywords: arctic (29), surface (18), earth science (6), and atmospheric tempera... (5).

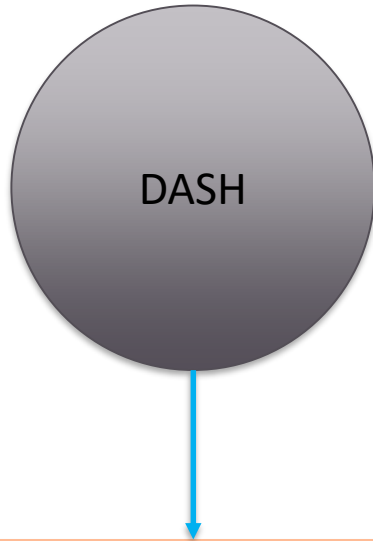
data.ucar.edu

The Digital Asset Services Hub (DASH)

The screenshot displays the DASH interface for a specific dataset. On the left, there is a 'Dataset extent' section with a map showing the location of the dataset in Sweden. The main content area is titled 'Resource' and contains the following information:

- Resource Title:** Latnjajaure Site, CLIP 4-Temp 2 Community Data [Alatalo]
- Description:** This dataset contains CLIP 4-Temp 2 Community data from the Latnjajaure site, Sweden in 1995, 1996, 1997, 1999 & 2000. The Community Level Interaction Program (CLIP) data comprises a block of poor, dry heath. For more information, please see the readme file.
- To Access Resource:** A section with a 'Go to Resource Homepage' button and a DOI link: <https://doi.org/doi:10.5065/D6QR4VBC>
- Questions? Email Resource Support Contact:** EOL Data Support, eol-datahelp@ucar.edu, UCAR/NCAR - Earth Observing Laboratory
- Temporal Range:** Begin: 1995-07-14T00:00:00Z, End: 2000-07-18T23:59:59Z
- Keywords:** A search bar containing the terms 'arctic' and 'surface'.
- Navigation Buttons:** Three blue buttons with white text and a plus icon: 'Scientific Information', 'Contact Information', and 'Citation Information'.

The Digital Asset Services Hub (DASH)



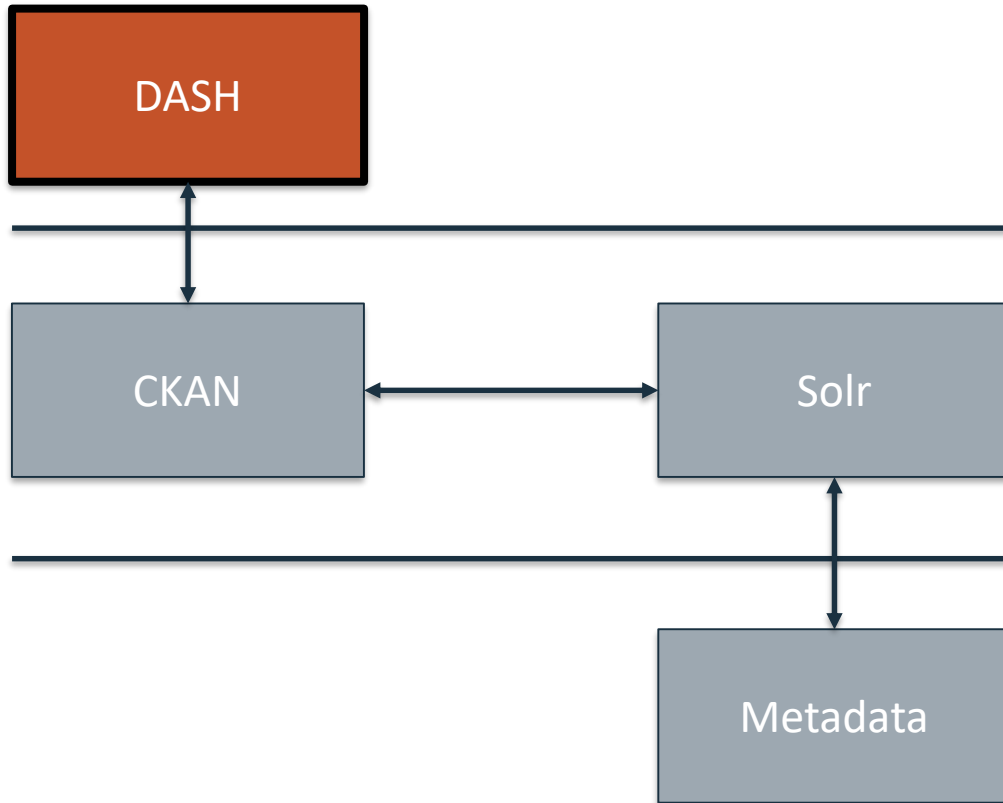
DASH Metadata

- NCAR Dialect Standard
- Bulk metadata ingest from GitHub repository
- Metadata Validation

Citation Information	
Author	Juha Alatalo Jrgen Pettersson
Publisher	UCAR/NCAR - Earth Observing Laboratory
Publication Date	2007-11-02T12:41:13
Digital Object Identifier (DOI)	https://doi.org/doi:10.5065/D6QR4VBC
Alternate Identifier	56.615
Resource Version	1.0
Topic Category	climatologyMeteorologyAtmosphere
Progress	completed
Metadata Date	2018-07-13T19:39:25Z
Metadata Record Identifier	edu.ucar.eol::56.615
Metadata Language	eng: USA
Suggested Citation	Juha Alatalo, Jrgen Pettersson. (2007). Latnjajaure Site, CLIP 4-Temp 2 Community Data [Alatalo]. 1.0. UCAR/NCAR - Earth Observing Laboratory. https://doi.org/doi:10.5065/D6QR4VBC . Accessed 30 July 2018.

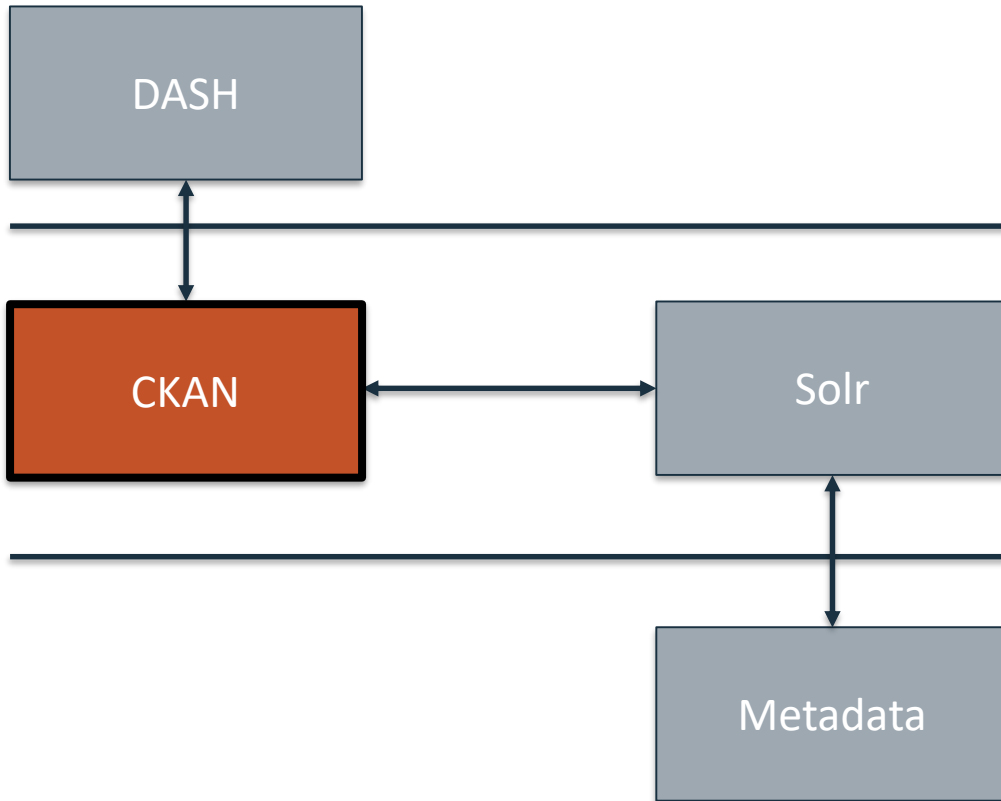
DASH uses metadata to build the product pages

How DASH works



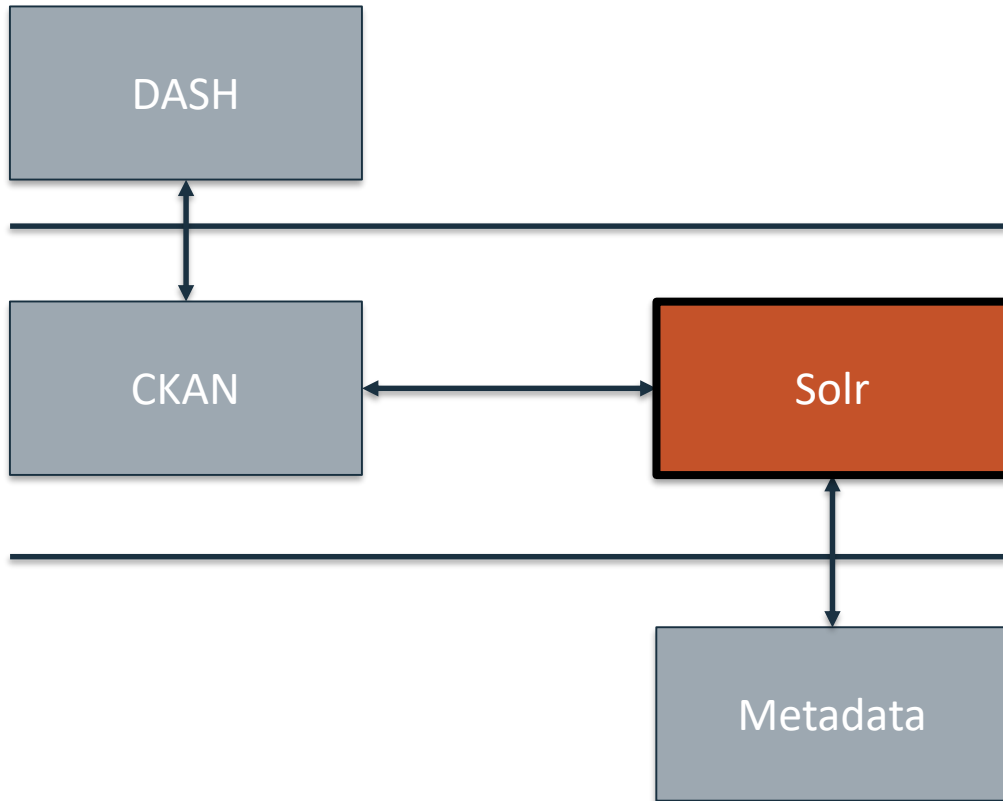
As the user performs a search from DASH. A query is made in CKAN.

How DASH works



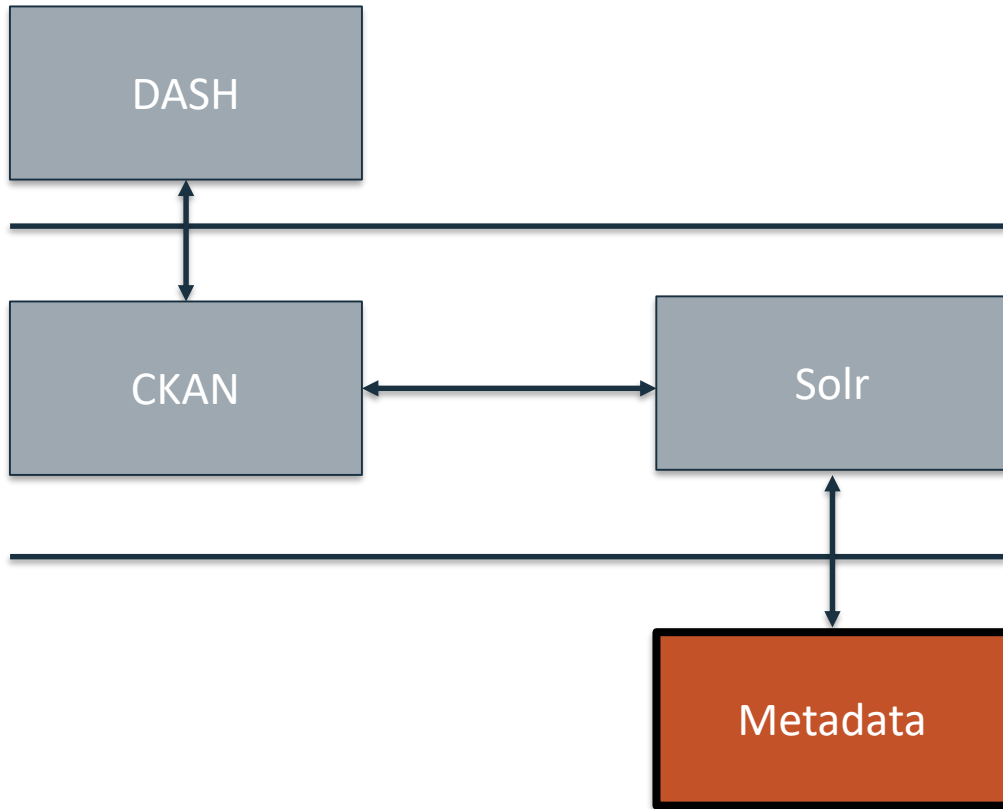
The query that is built in CKAN is now sent to Solr.

How DASH works



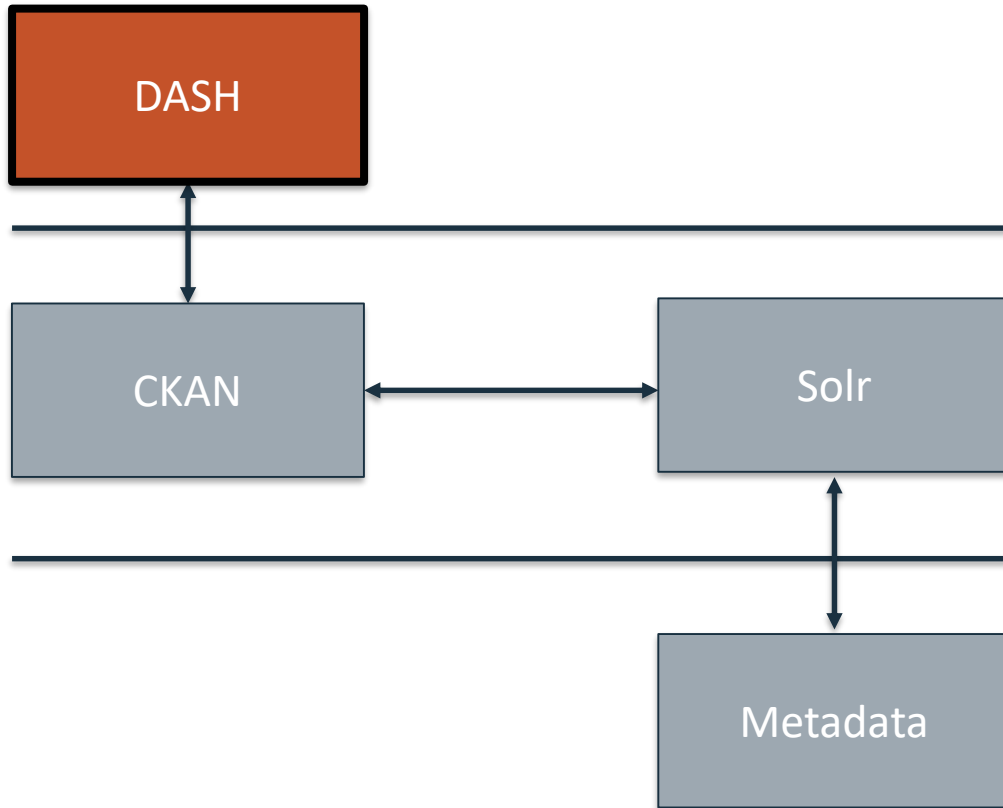
Now, Solr searches the metadata to find what the query asked for.

How DASH works



Solr performs a search for digital assets based on their metadata.

How DASH works



Once the metadata is retrieved by Solr, CKAN then builds the dataset page on DASH.

The Digital Asset Services Hub (DASH)

Filter by location [Clear](#)

Temp

37 resources found Order by: Best Match

Latnjaure Site, CLIP 4-Temp 2 Community Data [Alatalo]
This dataset contains CLIP 4-Temp 2 Community data from the Latnjaure site, Sweden in 1995, 1996, 1997

How can the users find similar assets?

to

Keywords

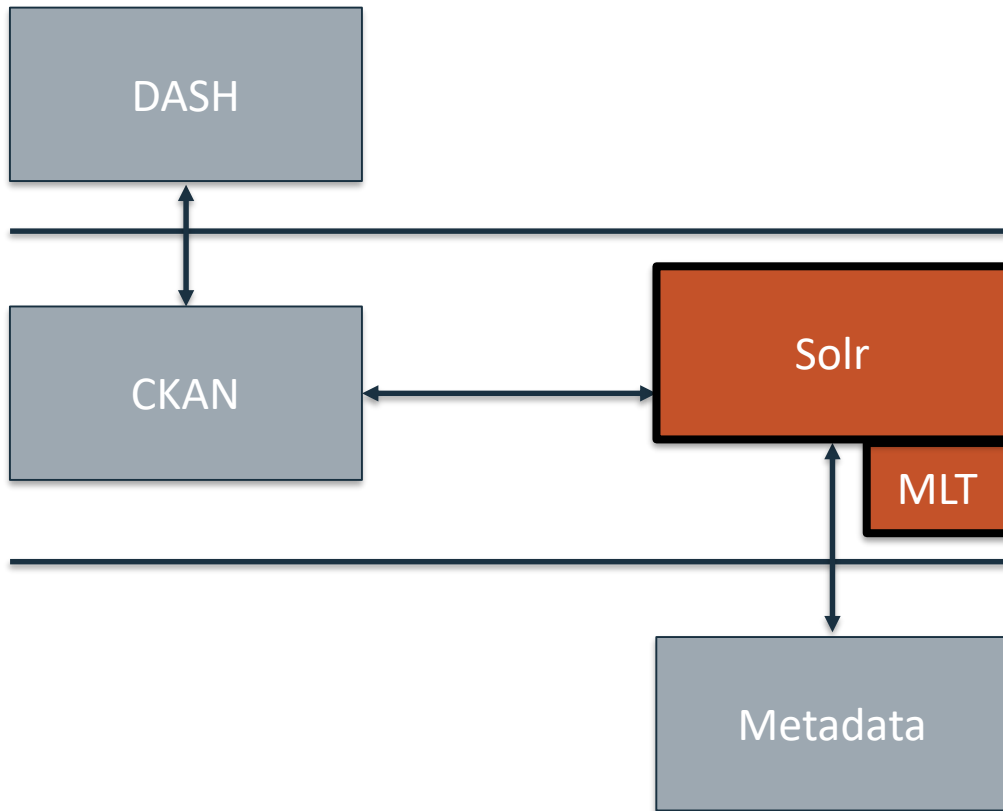
Search keywords...

- arctic (29)
- surface (18)
- earth science (6)
- atmospheric tempera... (5)

Latnjaure Site, CLIP 1-Temp 2 Community Data [Alatalo]
This dataset contains CLIP 1-Temp 2 Community data from the Latnjaure site, Sweden in 1995, 1996, 1997, 1999 & 2000. The Community Level Interaction Program (CLIP) data...
dataset

Latnjaure Site, CLIP 3-Temp 1 Community Data [Alatalo]
This dataset contains CLIP 3-Temp 1 Community data from the Latnjaure site, Sweden in 1995, 1996, 1997, 1999 & 2000. The Community Level Interaction Program (CLIP) data...
dataset

“More Like This” Implementation



Solr uses its “More Like This” function in order to find similar assets based on metadata.

Solr: "More Like This"

Sample Solr Query

```
select?q="Data Page Identifier"&mlt=true&mlt.fl=title,tags,description&fl=*
```


"q="

Uses the given
data to base
similar assets
off of

Solr: "More Like This"

Sample Solr Query

```
select?q="Data Page Identifier"&mlt=true&mlt.fl=title,tags,description&fl=*
```



"mlt=true"
Activates the
"More Like
This" function

Solr: "More Like This"

Sample Solr Query

```
select?q="Data Page Identifier"&mlt=true&mlt.fl=title,tags,description&fl=*
```




"mlt.fl="
Looks for the
given field to
return similar
assets

Solr: "More Like This"

Sample Solr Query

```
select?q="Data Page Identifier"&mlt=true&mlt.fl=title,tags,description&fl=*
```



"mlt=true"
Activates the
"More Like
This" function

Solr: "More Like This"

Sample Solr Query

```
select?q="Data Page Identifier"&mlt=true&mlt.fl=title,tags,description&fl=*
```

"q="

Uses the given data to base similar assets off of

"mlt=true"
Activates the "More Like This" function

"mlt.fl="

Looks for the given field to return similar assets

"fl="

returns all the metadata fields for similar asset datasets


Sample Solr Response


Once the query is processed, Solr returns the results back to CKAN in the form of a Python dictionary.

```
{'0d710677fb90774f48eac31a36bd0c6e': {'start': 0, 'numFound': 1909, 'docs': [{'name': 'council-climate-ncar-iss-rass-virtual-temperatures-ascii', 'title': 'Council Climate NCAR ISS RASS Virtual Temperatures (ASCII)'}, {'name': 'ncar-eol-sodar-rass-data-sodar-rass-site', 'title': 'NCAR/EOL Sodar-RASS Data (Sodar-RASS Site)'}, {'name': 'fable-1999-ncar-atd-tennessee-site-915mhz-wind-profiler-winds-and-rass-data', 'title': 'FABLE 1999 NCAR/ATD Tennessee Site 915MHz Wind Profiler Winds and RASS Data'}, {'name': 'ncar-eol-1290mhz-wind-profiler-nima-winds-and-rass-data-west-profiler-site', 'title': 'NCAR/EOL 1290MHz Wind Profiler NIMA Winds and RASS Data (West Profiler Site)'}, {'name': 'mapr-wind-profiler-and-rass-data-ncar-eol', 'title': 'MAPR Wind Profiler and RASS Data [NCAR/EOL]'}]}}
```

This non user-friendly Python dictionary is formatted to be visualized in DASH.

Results

Similar Assets 


Council Climate NCAR ISS RASS Virtual Temperatures (ASCII) 

This dataset contains Radio Acoustic Sounding System (RASS) virtual temperatures collected at the Council, Seward Peninsula, Tundra Site (C2) during the Arctic Transitions in the Land-Atmosphere System (ATLAS) program 1999 summer field campaign. ...

dataset


NCAR/EOL Sodar-RASS Data (Sodar-RASS Site)

This dataset contains final, non-quality controlled Sodar-RASS data collected by the NCAR/EOL Integrated Sounding System (ISS) Sodar-RASS system that was located at the southeast end of the Vale do Cobrao during the Perdigao field project in ...

dataset 

FABLE 1999 NCAR/ATD Tennessee Site 915MHz Wind Profiler Winds and RASS Data

This dataset contains 915 MHz wind profiler data collected by the NCAR/ATD Integrated Sounding System (ISS) located at the Walker Branch Watershead in Oak Ridge, Tennessee during the Forest Atmosphere Boundary Layer Experiment (FABLE) in 1999. This ...

dataset 

NCAR/EOL 1290MHz Wind Profiler NIMA Winds and RASS Data (West Profiler Site)

This dataset contains quality-controlled 1290 MHz wind profiler data collected by the NCAR/EOL Integrated Sounding System (ISS) that was located west of the double ridge at the West Wind Profiler Site during the Perdigao field project in 2017. This ...

dataset

MAPR Wind Profiler and RASS Data [NCAR/EOL]

This dataset contains Multiple Antenna Profiler Radar (MAPR) wind profiler data collected by the NCAR/EOL Integrated Sounding System (ISS) that was located at the "Homestead Profiling Site" near Elmwood, OK during the International H2O ...

dataset

Future Work

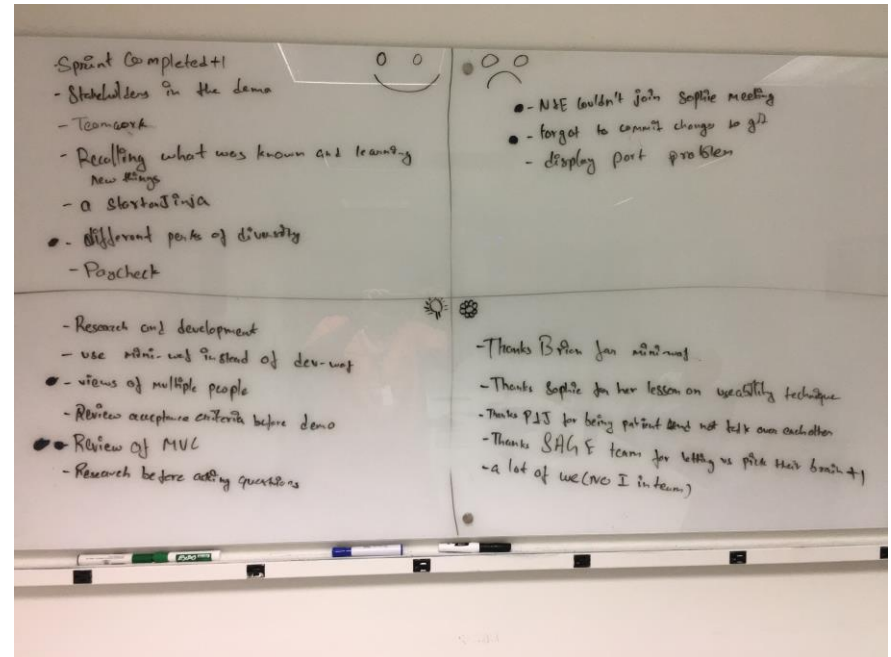
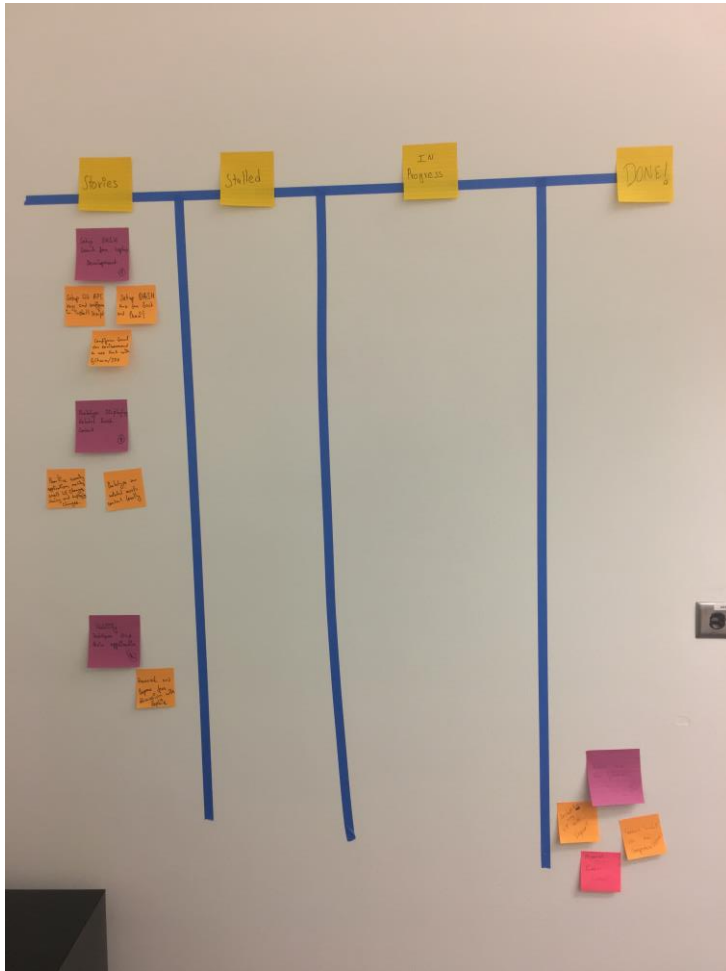
- More “More Like This” fields to narrow down similar assets.
- Add click counter to calculate how often our product is being used.
- Add a feature that lets the user search for “More Like This” based on their desired fields.
- Usability testing

Agile Scrum

The screenshot displays a Jira Agile Scrum board for the project "DSET SIParCS". The main view is the "Backlog" for "DSET Sprint 9", which is active and runs from 16/Jul/18 5:00 PM to 20/Jul/18 2:00 PM. The sprint contains three issues: DSETSP-42 (2 points), DSETSP-43 (5 points), and DSETSP-44 (2 points). Below the sprint, a larger backlog of 9 issues is visible, including tasks like "Review DASH Search Solr Configuration", "Map NCAR Dialect Minimum Concepts to Schema.org", and "Add Schema.org Metadata to DASH Search Resource Page".

The right-hand panel provides a detailed view of issue "DSETSP-42". The task is "Map DSET Metadata to schema.org's CreativeWork Metadata" with an estimate of 2. The description includes the text "types to schema.org metadata types." and a comment from Nathan Hook: "Here is the link to the dset metadata definitions and their iso mappings:". Another comment from Joshua Jones provides a Google Docs link: "https://docs.google.com/spreadsheets/d/1qD".

Agile Scrum



Acknowledgements

- Rich Loft
- AJ Lauer
- Jenna Preston
- Elliot Foust
- Mentors
 - Eric Nienhouse
 - Nathan Hook
 - Sophie Hou
- Project Partner
 - Joshua Jones
- DSET
- SAGE Team
 - Brian Bonnlander
 - Christy Grant
 - Saquib Aziz Khan
- SIParCS Interns



References

- David Read. "More Like This in CKAN (DGU Extension).", (2012), GitHubGist, gist.github.com/davidread/2217684.
- "Overview¶." Overview - CKAN 2.8.1 Documentation, docs.ckan.org/en/ckan-2.3.4/.
- "Solr Tutorial." Asf - Revision 1831250: /Openoffice/Trunk/Main/Extras/Source/Gallery/Arrows, svn.apache.org/repos/asf/lucene/solr/branches/branch-1.4/site/tutorial.html.