NCAR DASH Search and Linked Data: Investigation and Implementation using Schema.org

J. Robert Jones[1]
Prasil Mainali[1], Sophie Hou[2], Eric Nienhouse[3], Nathan Hook[3]

SIParCS Intern[1] – DSET[2] – SAGE[3]

Aug 1, 2018

Introduction

- DSET and DASH, What are they?
- Motivation of Project

Linked Data Overview

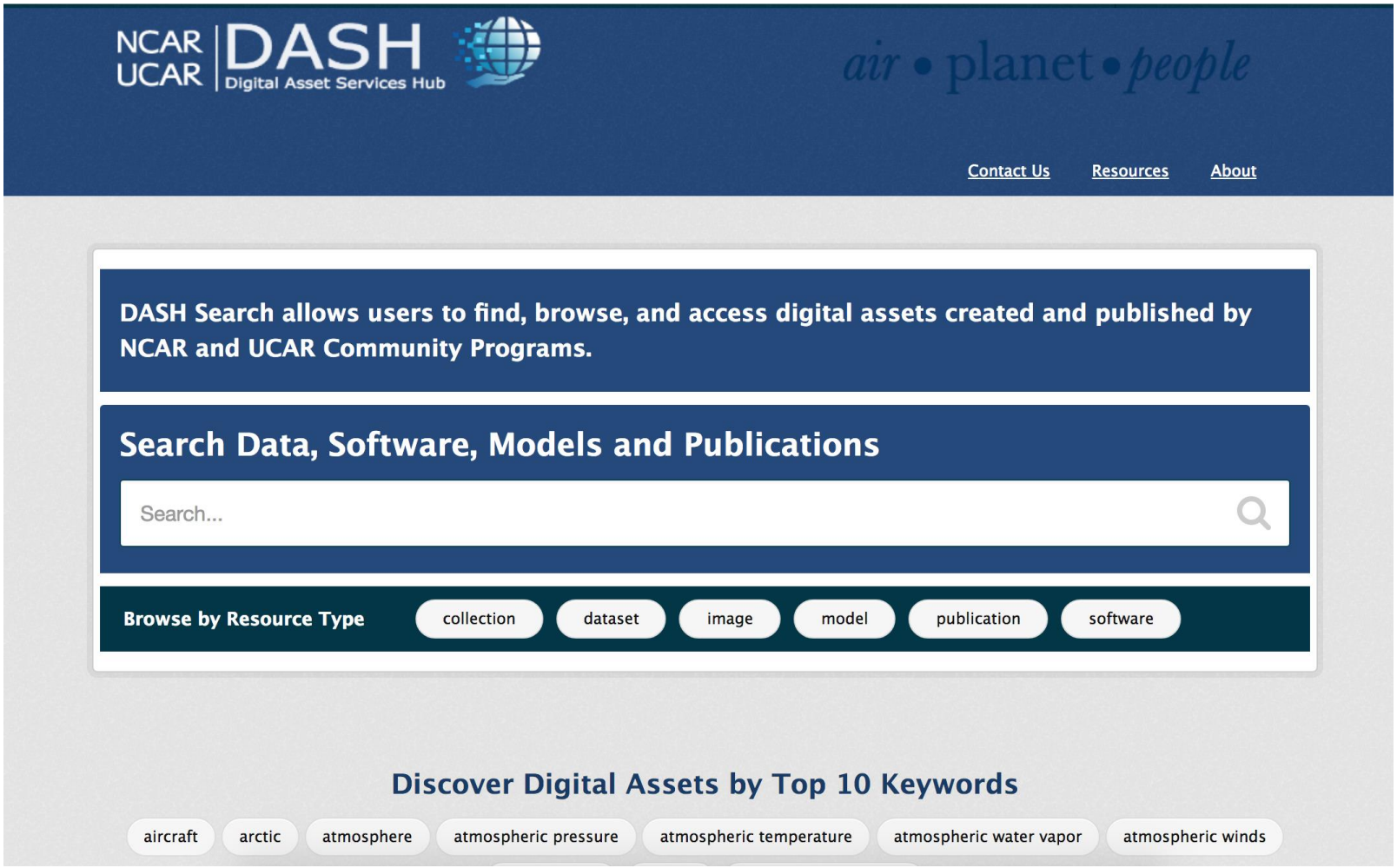
- Linked Data Model
- Schema.org

Implementation and Results

Benefits

- Project
- Internship

DASH

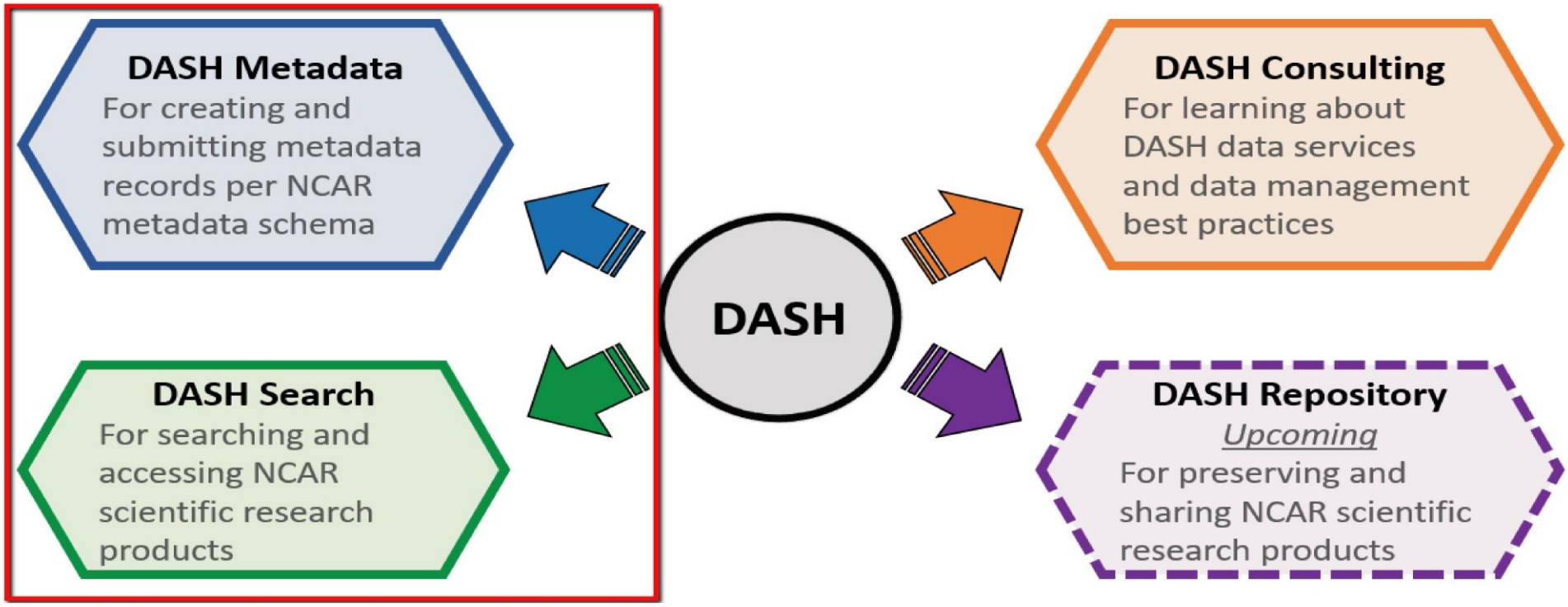


data.ucar.edu

DASH and DSET

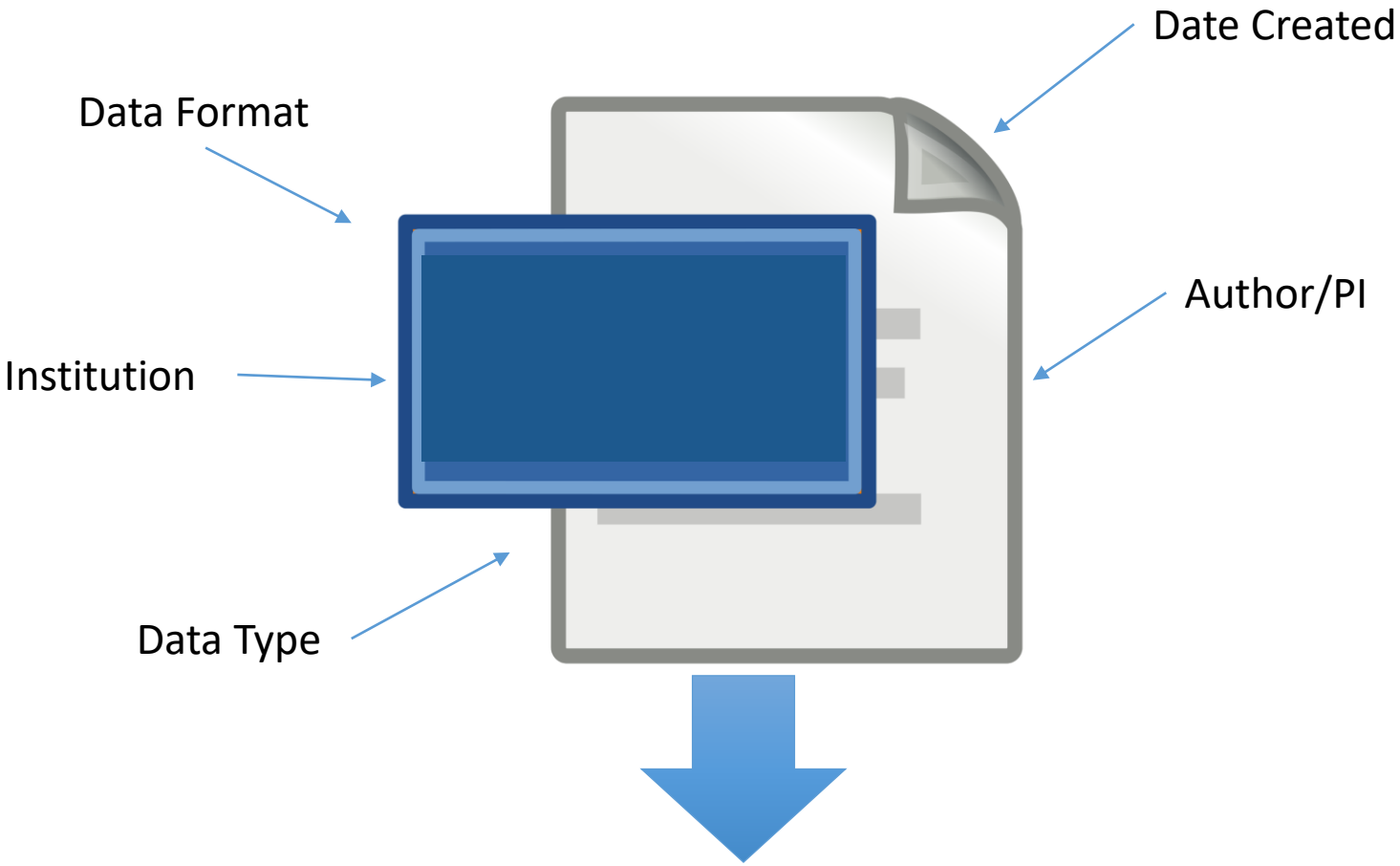
DASH – Digital Asset Services Hub

DSET Team – Data Stewardship Engineering Team



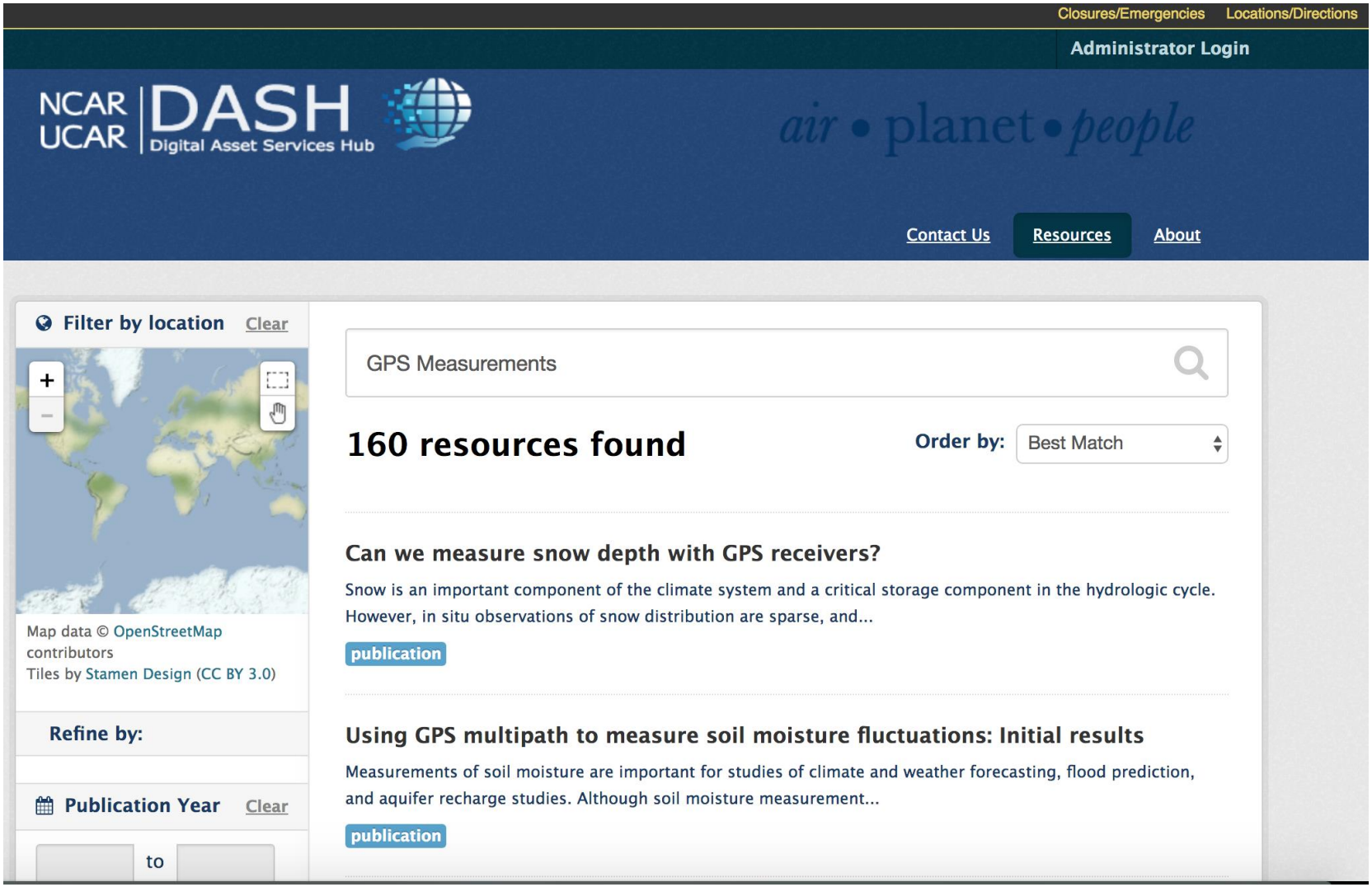
To provide efficient access to digital assets with excellent user experience and to improve coordination across NCAR

DASH Metadata



ASSET: dataset, publication, model

DASH Search

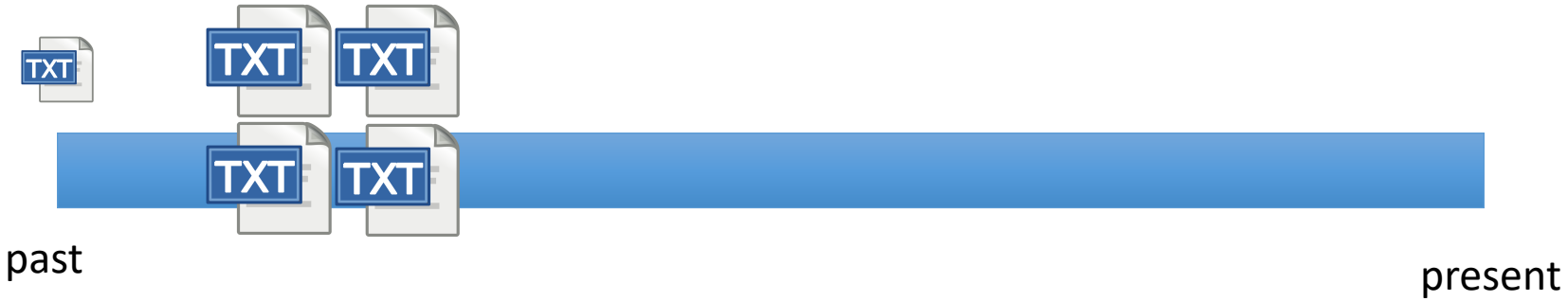

Closures/Emergencies
Locations/Directions
Administrator Login
NCAR UCAR
DASH
Digital Asset Services Hub
air • planet • people
Contact Us
Resources
About
Filter by location
Clear
Map data © OpenStreetMap contributors
Tiles by Stamen Design (CC BY 3.0)
Refine by:
Publication Year
Clear
to
GPS Measurements
160 resources found
Order by:
Best Match
Can we measure snow depth with GPS receivers?
Snow is an important component of the climate system and a critical storage component in the hydrologic cycle. However, in situ observations of snow distribution are sparse, and...
publication
Using GPS multipath to measure soil moisture fluctuations: Initial results
Measurements of soil moisture are important for studies of climate and weather forecasting, flood prediction, and aquifer recharge studies. Although soil moisture measurement...
publication

Motivation

As science progresses in the digital age, scientific data and digital assets are being created at a higher rate

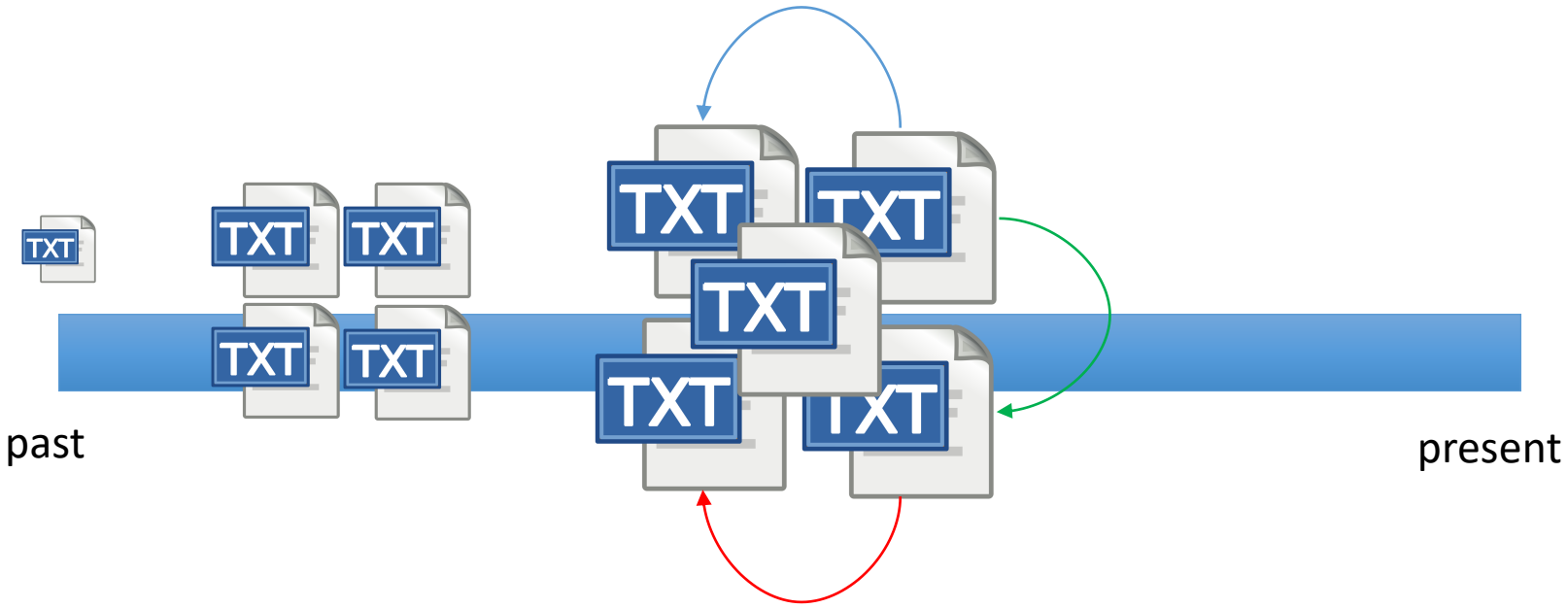
Motivation

These datasets and digital assets continue to grow in size



Motivation

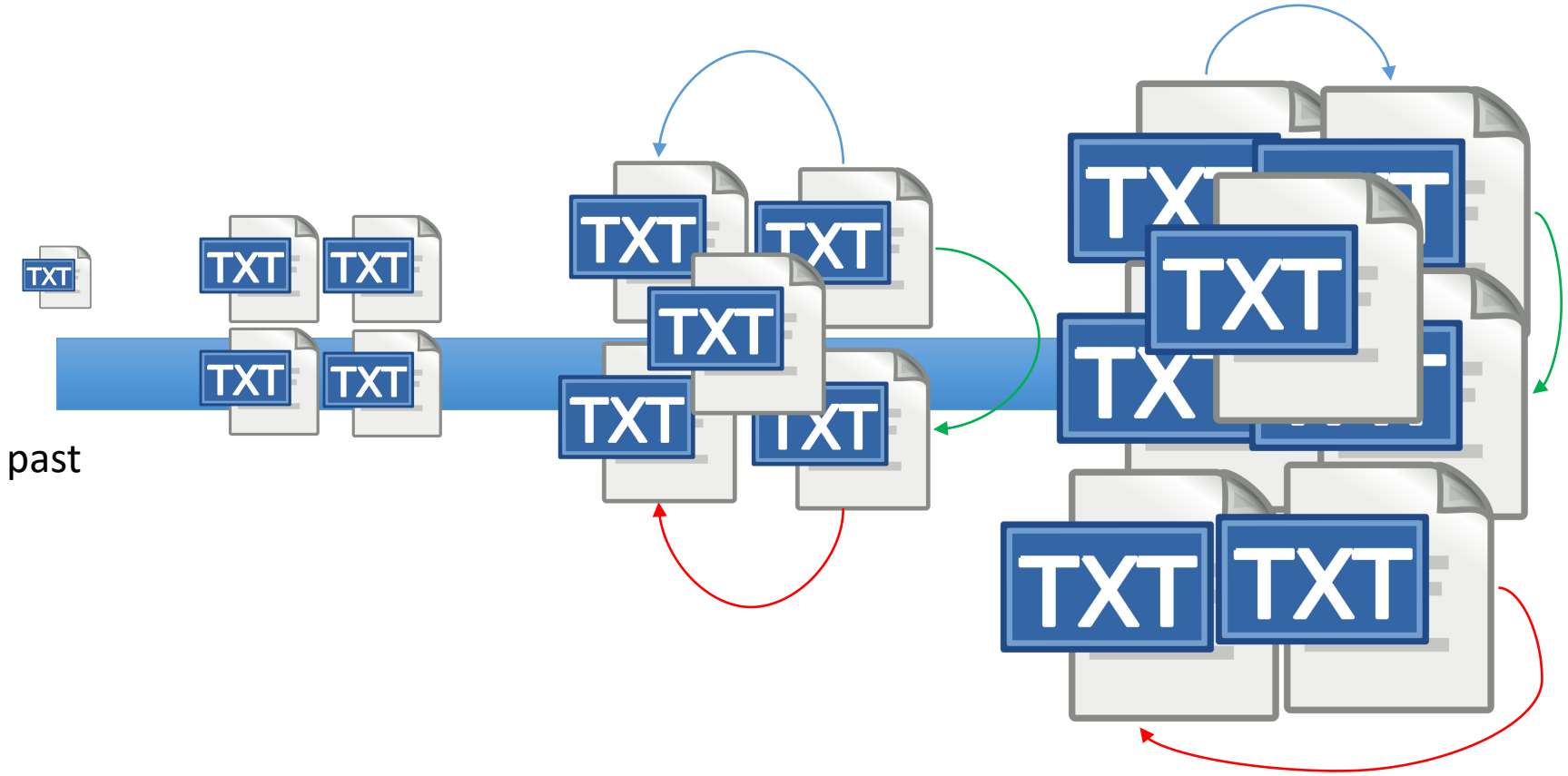
With more digital assets scientist and users started to group these assets using metadata



Linked-Data: Best practices for connecting assets on the web

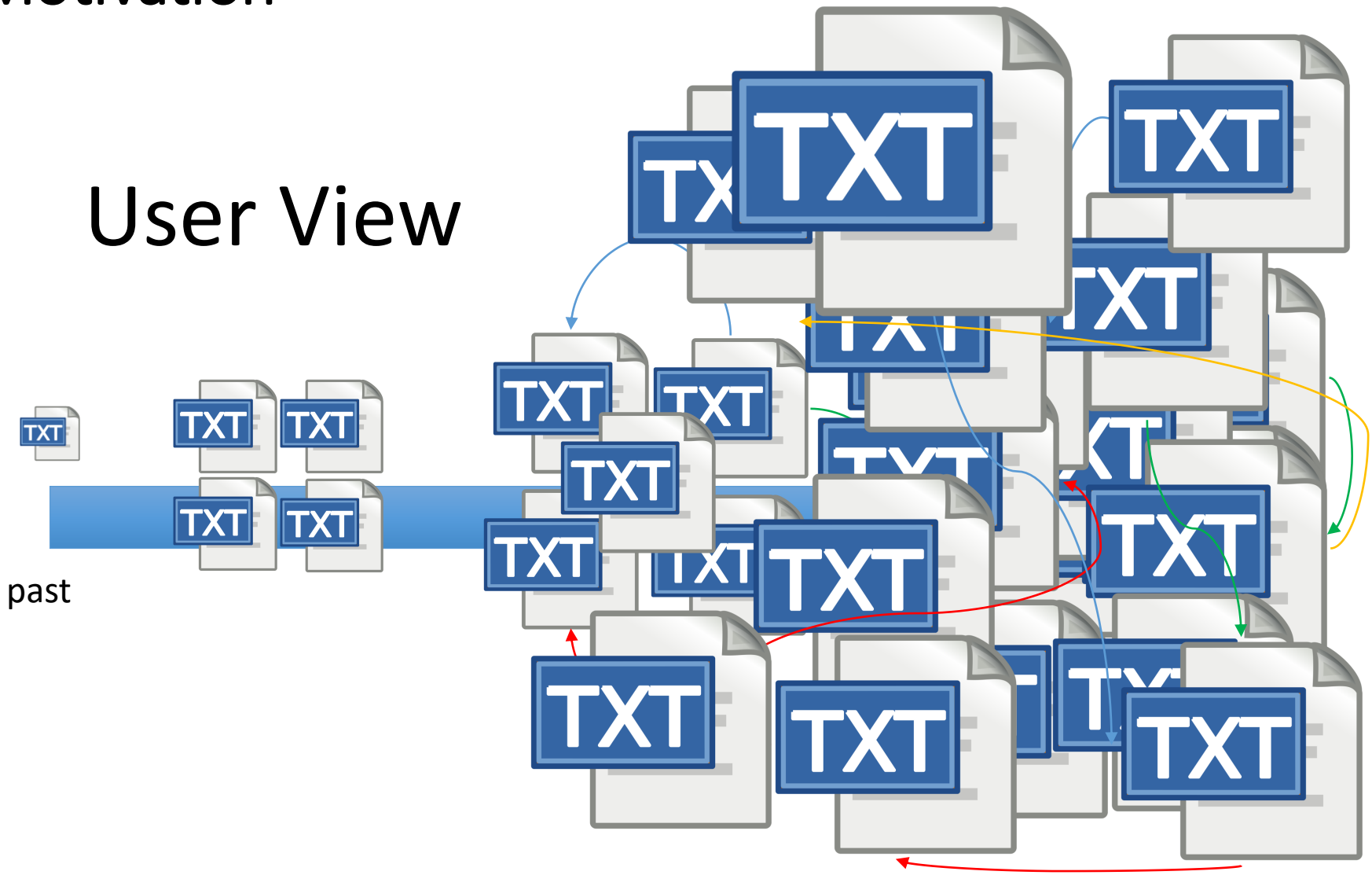
Motivation

This has led to the present with hundreds of links made between digital using different vocabularies and descriptors



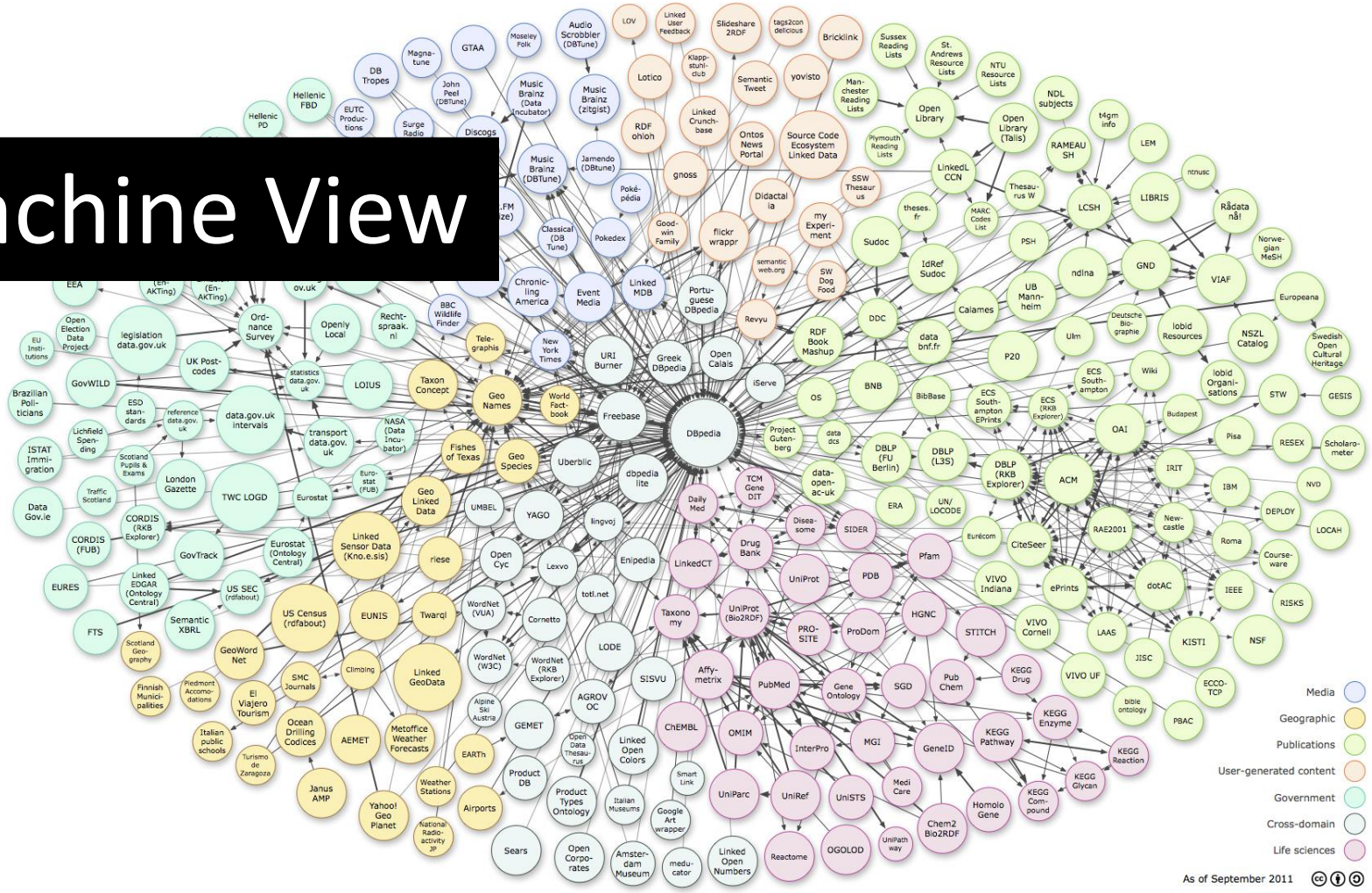
Motivation

User View



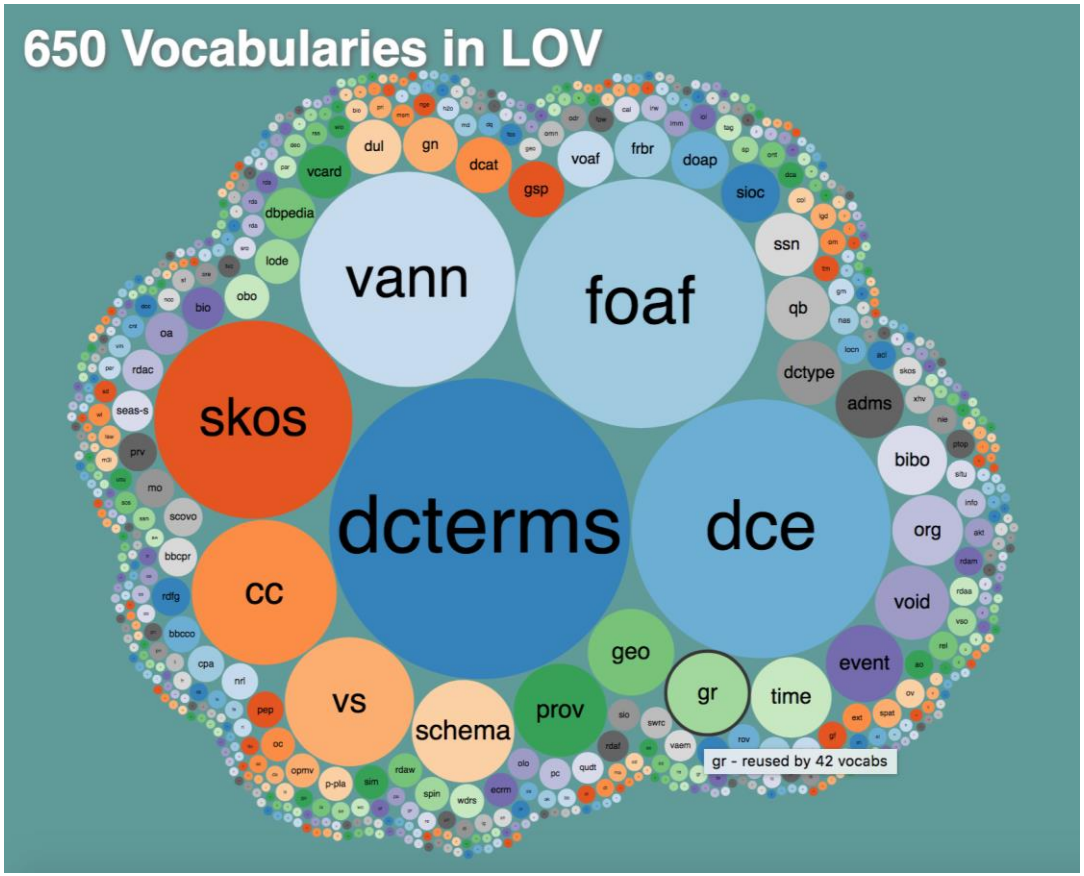
Linked-Data Model

Machine View



Linked-Data Model

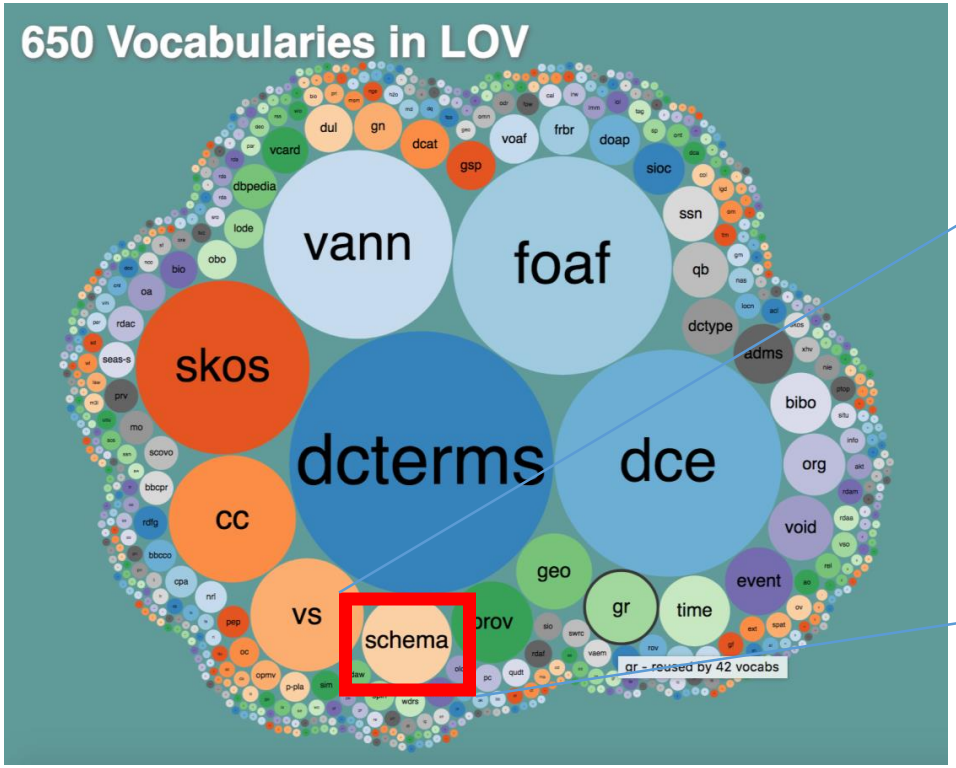
There are many different linked open vocabularies **(LOV)** created that can be used to link data with metadata requiring different markups

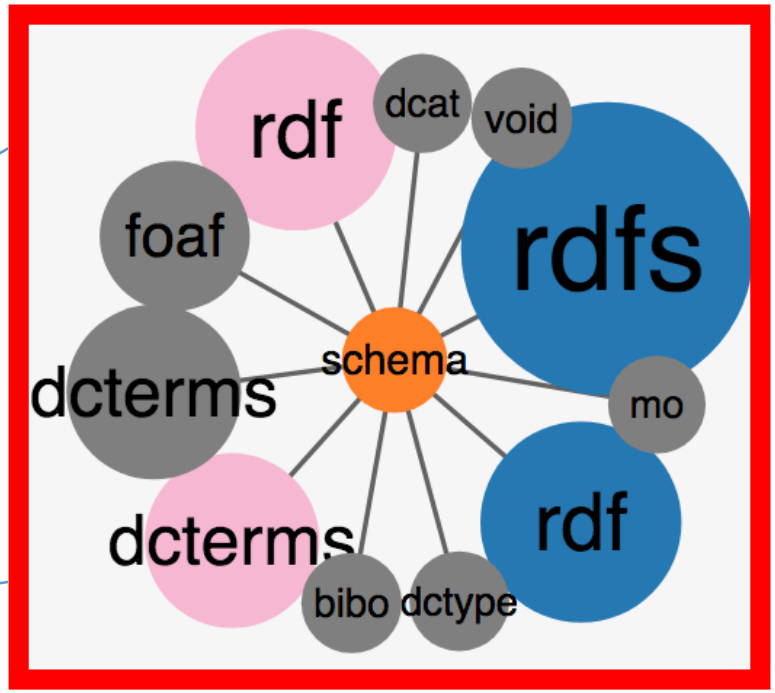


Linked-Data Model

We choose the recently developed Schema.org for DASH implementation

Schema.org builds off of previously created LOV





Why Schema.org?

Schema.org has many properties that makes it desirable as a linked open vocabulary:

- implemented underneath webpage in HTML source code
- can be read by all most search engines to increase search accuracy and discoverability
- multiple ways to add the Schema.org markup to the webpage
- easy to understand linkable vocabulary

Why Schema.org?

Three primary methods for webpages and digital assets to include a Schema.org markup

1. JSON- LD: linked data descriptors placed in a json script formatted block
   can be read dynamically
   can be placed anywhere on the page

2. RDFa: data descriptors are placed in annotated HTML tag attributes.
   Needs to be in both the head and body sections of HTML

3. Microdata: data descriptors placed with the data's HTML content.
   Needs to be in the body

Implementation

NCAR metadata and Schema.org use two different vocabularies to define the same metadata field

This required a mapping from NCAR metadata dialect to Schema.org's

Base Level Data Describers				
NCAR Dialect		Schema.org Dialect		Field Definition
Metadata Record ID	------>	identifier	------>	Persistant ID given to an asset
ISO Asset Type	------>	additionalType	------>	Type/Format of asset
Metadata Point of Conact	------>	Creator	------>	Individual/Group responsible for metadata
Resource Support Contact	------>	accountablePerson	------>	Individual/Group responsible for scientific asset
Metadata Date	------>	dateCreated	------>	Date of data collection creation
Publication Date	------>	datePublished	------>	Data of data publication
Author	------>	author	------>	Name of princicple investigator for asset
title	------>	name	------>	Name of the asset
Landing Page URL	------>	sameAs	------>	URL of asset source
Description	------>	description	------>	Description of asset

Results – JSON LD

```
<script type="/application/ld+json">
{
        "name": "Offshore propagation of coastal precipitation",
        "creator": "Li, Yanping", "Carbone, Richard”
        "sameAs": "http://n2t.net/ark:/85065/d7gm88sb",
        "author": "Li, Yanping", "Carbone, Richard ”
        “dateCreated": "2018-01-24T17:55:27.693543",
        "accountablePerson":  "UCAR/NCAR - Library” ,
        "additionalType": "publication",
        "@context": "http://schema.org",
        "datePublished": "2015-12-01T00:00:00",
        "@type": "CreativeWork",
        "description": null}
</script>
```

Project Benefits

Scientists or NCAR labs wanting to increase asset visibility can include more data in the initial metadata collection that is included in the Schema.org vocabulary.

Assets in DASH will automatically:

- Create the correctly formatted attached Schema.org block
- Increase in visibility to general search engines
- Be connected to other assets in the future with relationship descriptors (i.e. subset, parent, etc)

Internship Benefits

experienced firsthand how research is done at large national lab

renewed interest in Computer Science

rise in computer science skills

networking opportunities throughout the summer

seminars built to help career development

collaborations made

personal growth

always five minutes away from a hiking trail with awesome geology

Acknowledgments

Rich Loft for support of SIParCS Program

AJ Lauer, Jenna Preston, Eliott Foust

Mentors: Sophie Hou, Eric Nienhouse, Nathan Hook

Project partner: Prasil Mainali

CISL

Sage Team

SIParCS Interns