Discussions will take place on our discord channel
“how-to-better-describe-software-for-discovery-and-citation”
Where you can also find a link to the slides

Peter Teuben (UMD)
Alice Allen (UMD)
Bruce Berriman (Caltech)

How to better describe software for discovery and citation

“How to better describe software for discovery and citation”
Discussion Items  “how to better describe…”

1. Why this session? (Peter) [5 min]
2. Citing Montage (Bruce) [10 min]
3. Generating codemeta.json and CITATION.cff files from ASCL (Alice) [10 min]
4. Licensing metadata: B10-133 summary (Yan) [5 min]
5. New metadata items? (Peter) [5 min]
6. UAT to describe software (Peter/Katie) [5 min]
7. Use a niche science meeting to do a software census (Peter) [5 min]

8. AOB - open floor (All) [45 min]

Discussion in Discord after each item
1. Why this session?

- How can you - software writers - ensure your software is found and properly cited?
- Rely on registry in ASCL and/or Zenodo? [P9-103 Mavuram poster]
- Find in ADS?
- Rely on a noisy google?
- Spaans presentation [https://zenodo.org/record/2558482](https://zenodo.org/record/2558482) “making software findable….”
- Example: Citing Montage
2. Citing Montage

http://montage.ipac.caltech.edu

G. Bruce Berriman

(Caltech/IPAC-NExScI)
Using the CiteAs tool

https://citeas.org

All research products deserve credit.

Get the correct citation for diverse research products, from software and datasets to preprints and articles.

Paste a URL, DOI, arXiv ID, or any search term (e.g. software name/abbreviation)

http://montage.ipac.caltech.edu

Examples: http://yt-project.org  https://cran.r-project.org/web/packages/stringr  More examples
What It Returned...
Looking in the user input, we found a link to a webpage:

http://montage.ipac.caltech.edu

Looking in the webpage, we didn't find a link to a cite-as relation header.

Looking in the webpage, we didn't find a DOI.

Looking in the webpage, we didn't find an ArXiv page.
Looking in the webpage, we found a link to a

GitHub repository main page

https://github.com/Caltech-IPAC/Montage

Looking in the GitHub repository main page, we didn't find a DOI.

DOI API response

Looking in the GitHub repository main page, we didn't find

CodeMeta file

Looking in the GitHub repository main page, we didn't find a link to a

CITATION file

Looking in the GitHub repository main page, we found a link to a

README file

https://raw.githubusercontent.com/Caltech-IPAC/Montage/master/README.md
Looking in the README file, we didn't find a DOI.

DOI API response

Looking in the GitHub repository main page, we didn't find a link to a R DESCRIPTION file

Looking in the GitHub repository main page, we found a link to a

GitHub repository API response


Parsing the GitHub repository API response, we found

The citation metadata
... and so

What are the recommended best practices for software providers?
3. Creating software metadata files from ASCL entries

Alice Allen

Astrophysics Source Code Library (ascl.net)
Metadata files

codemeta.json

CITATION.cff
Metadata files on demand!

Add `/codemeta.json` or `/CITATION.cff` to ASCL entry URL

Works only for codes with ASCL IDs

Are a starting point; please edit as needed!
Example: codemeta.json

https://ascl.net/1010.051/codemeta.json
@context: "https://doi.org/10.5063/schema/codemeta-2.0"
@type: "SoftwareSourceCode"
name: "NEMO: A Stellar Dynamics Toolbox"
description: "NEMO is an extendible S...n that Barnes maintains."
identifier: "ascl:1010.051"
author: [...] 
citation: "https://ui.adsabs.harvard.edu/abs/1995ASPC...77..398T"
relatedLink: [...] 
repository: [...] 
referencePublication: [...] 
version: "PLACEHOLDER: Add version here"
license: "PLACEHOLDER: Add license...licenses/MIT.html) here"
@context: "https://doi.org/10.5063/schema/codemeta-2.0"
@type: "SoftwareSourceCode"
name: "NEMO: A Stellar Dynamics Toolbox"

description: "NEMO is an extendible S...n that Barnes maintains."
identifier: "ascl:1010.051"

author: [...]  


citation: "https://ui.adsabs.harvard.edu/abs/1995ASPC...77...398T"

relatedLink: [...]  

codeRepository: [...]  

referencePublication: [...]  

version: "PLACEHOLDER: Add version here"

license: "PLACEHOLDER: Add license.../licenses/MIT.html) here"
Example: CITATION.cff

https://ascl.net/1010.051/CITATION.cff
NEMO is an extendible Stellar Dynamics Toolbox, following an Open-model. It has various programs to create, integrate, analyze and visualize N-like systems, following the pipe and filter architecture. In addition there a
Why have these?
(not an exhaustive list)

*Preferred citation* information lets people know how to cite your software

Ingestion into other systems (CiteAs, Zenodo, etc.)

With uptake, searchable Ring for/of software
Outstanding issues

Competing formats

Uptake: Getting people to use a standard format

Ingestion into other systems (CiteAs, Zenodo, etc.)
4. Licensing metadata

Yan Grange
(Science Data Center, ASTRON)
Jutta Schnabel, Thomas Jürges, Mattias Füßling, Nuria Lorente
4. Licensing metadata (Yan Grange)

1. Attendance was very high (>120 people interested in software licensing)!
2. It is widely accepted that nowadays source code licenses are a must.
3. Main conclusion is that the main issue is not whether or not to choose a license, but that what license to pick can be a complex discussion because it depends also on external constraints.
4. In the community, permissive licenses tend to be more used than non-permissive ones.
5. One should keep in mind that relicensing is very tricky.
6. There is a clear wish from the community to share best practices, learn from each other and exchange guidelines and knowledge.
Potential for follow-up discussions (next year’s ADASS?):

1. Data licensing
2. Dealing with contributions
3. establishing and enforcing guidelines in your work environment
4. collaboration on a community-based approach

https://jschnabel.pages.in2p3.fr/licensing-bof/session/
5. Expand/deepen codemeta file with “API” information

Peter Teubken
University of Maryland
How deep should we go?

1. How deep should we go? cf. the old (now defunct) code.google.com/archive
2. Keywords describing the API and its one liners - great for searching, for example which functions or programs in a package deal with deconvolution
3. An example of one level deep is automated in NEMO’s `mktasklist` script
4. Is this something for schema.org https://schema.org/SoftwareApplication

```json
"version": "4.0.1",
"license": "PLACEHOLDER: Add license (e.g. https://spdx.org/licenses/MIT.html) here",
"tasklist": [
    { "name": "snapplot", "description": "plot particle positions from a snapshot file" },
    { "name": "snapprint", "description": "tabulate a snapshot" }
],
```
6. Unified Astronomy Thesaurus (UAT)

Peter Teuben

University of Maryland
1. Thesaurus: (noun)
   A list of words in groups of synonyms and related concepts
   A registry of terms you can use
   a. Maintained in github
   b. example
3. Can we use it to describe software via keywords and thus improve discovery?
   a. Codemeta keywords?
4. Do we need more coverage for software and algorithms in UAT?
5. Examples (or see next slide on Software Sensus)
6. Astronomy & Computing

   Galaxy classification systems ; Convolutional neural networks ; Galaxy structure
   Computational astronomy ; Computational methods
   Milky Way, galaxies, stellar dynamics, stellar populations
Keywords

**AAS: (from UAT)**
Galaxy classification systems; Convolutional neural networks; Galaxy structure
Computational astronomy; Computational methods
Milky Way, galaxies, stellar dynamics, stellar populations

**A&C: (from ???)**
Photoionisation modelling; Scientific visualisation; Cloudy; Planetary nebula; (PN)Novae
Galaxies: photometry; Methods: data analysis; Machine learning; Techniques: image processing; Galaxies: GeneralCatalogs
FITS; File formats; Standards; World coordinate system
7. Software census at a niche science meeting?

Peter Teuben
University of Maryland
7. Software census - at a niche science meeting?

1. Find a nice meeting with well defined science goals and do a software census
   a. Have all the domain experts in a room
   b. [https://extragalactic-milkyways.org/](https://extragalactic-milkyways.org/)
   c. … Stellar Dynamics & Stellar Populations …
   d. This meeting is in two stages, even more ideal for this idea, since they reconvene next year (Dec 2020, Nov 2021)
   e.

2.
8. AOB / Open Floor

1. Expand this BoF into a post-ADASS White Paper?
   a. Should this include licensing from B10-133?
2. IVOA next week
3. 