

How Helium Team's Stack Uses Minids and BDBags

May 24, 2018

Helium Team Presentation to KC7

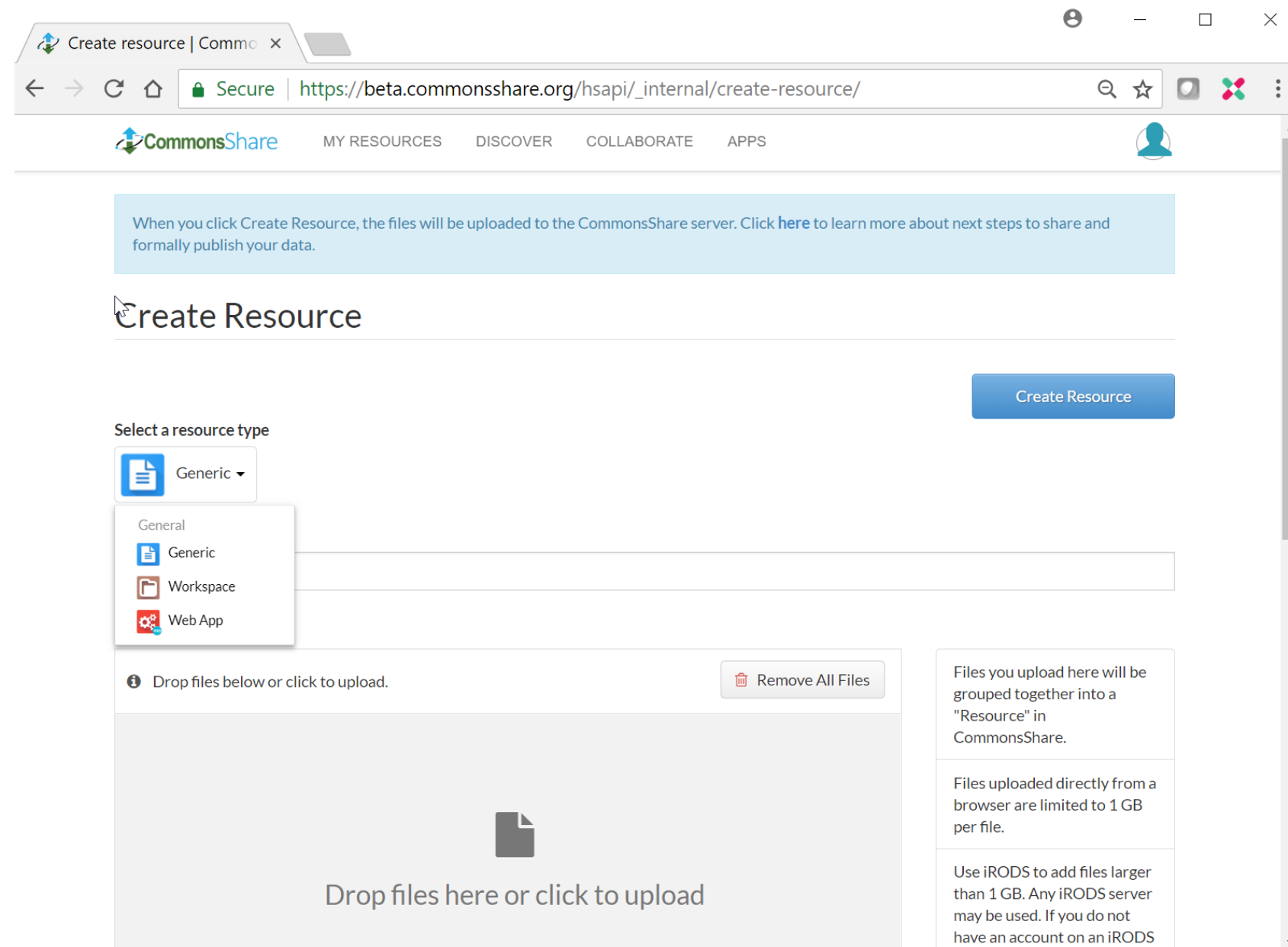
CommonsShare

- CommonsShare is Helium Team's KC5 web user interface to the NIH Data Commons
- <https://beta.commonshare.org/>
 - Supports “Sign In With Globus”

NOTE: All minids referenced herein and created in beta.commonshare.org are with test flag set to true indicating they are for testing purposes only.

The screenshot shows the CommonsShare website. At the top, there's a navigation bar with the CommonsShare logo, links for 'MY RESOURCES', 'DISCOVER', 'COLLABORATE', and 'APPS', and a 'SIGN IN WITH GLOBUS' button. Below the navigation bar, a banner area features the text 'Join the community to start sharing' and 'CommonsShare is an online collaboration environment for sharing data, workflows, and code.' with a 'Sign in now' button. The main content area has a large image of cells with the text 'Share your data and workflows with colleagues' and 'Upload, share, and access a broad set of data types and workflows in the Data Commons. Run analysis on the data with apps from the app store in a cloud-agnostic environment.' Below this is a 'How it works' section with four steps: 1. Create data (Collect your data using the same methods you use now), 2. Upload to CommonsShare (Upload your data files to CommonsShare through the web user interface), 3. Describe with metadata (Use CommonsShare's simple metadata entry forms to finish describing your data), and 4. Share with colleagues (You choose who has access to the data and workflows you have uploaded). Below the steps are four icons representing each step. The 'What you can do with CommonsShare' section lists several capabilities with checkmarks: Share your data and workflows with colleagues, Manage who has access to the content that you share, Share, access, visualize and reuse a broad set of data types and workflows, Use the web services API to program automated and client access, Publish data and workflows using FAIR-TLC principles (Findable, Accessible, Interoperable, Reusable, Traceable, Licensed, Connected), Discover and access data and workflows published by others, and Use apps from the app store to visualize, analyze and run workflows on data in a cloud-agnostic environment. To the right of this list is an image of a laptop displaying the CommonsShare interface. At the bottom, there's a footer with 'CONTACT US' (Email us at help@commonshare.org), 'FOLLOW' (Twitter, Facebook, LinkedIn icons), and 'OPEN SOURCE' (CommonsShare is Open Source. This is CommonsShare Version 0.9b). The very bottom of the page contains a small copyright notice: '© 2018 CUAHSI and University of North Carolina at Chapel Hill. This material is based upon work supported by the National Science Foundation (NSF) under awards 1148453, 1148090, 1459300, 1664018, 1664061, and 1664119, and National Institutes of Health (NIH) under awards OT3-OD025464-01S1 and OT3-HL142479-01. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or NIH. | Terms Of Use | Statement of Privacy | Site Map'.

CommonsShare Resources



- CommonsShare (CS)
 - In CommonsShare, ingested data becomes resources
 - One or more of any type of files can be added to Generic resources
 - CS Workspace resources can hold multiple other resources
 - Web App resources hold apps from the App Store
 - E.g. Jupyter notebooks

CommonsShare BDBag

- <https://beta.commonsshare.org/resource/80ce15a0357f42ad921a01543c3f3a2c/>

GUID 80ce15a0357f42ad921a01543c3f3a2c/

/data

bag-info.txt

bagit.txt

fetch.txt

manifest-md5.txt

manifest-sha256.txt

tagmanifest-md5.txt

tagmanifest-sha256.txt

fetch.txt

irods://test.commonsshare.org:1247/commonssharetestZone/home/bagsdata/80ce15a0357f42ad921a01543c3f3a2c/data/E-MTAB-5214.sdrf.csv 15348705 data/E-MTAB-5214.sdrf.csv

The screenshot shows the CommonsShare BDBag interface for the resource E-MTAB-5214. The browser address bar displays the URL <https://beta.commonsshare.org/resource/80ce15a0357f42ad921a01543c3f3a2c/>. The page title is "E-MTAB-5214 - RNA-seq from 53 human tissue samples from the Genotype-Tissue Expression (GTEx) Project". The authors listed are Ray Idaszak. The resource type is Generic. The abstract states: "E-MTAB-5214 - RNA-seq from 53 human tissue samples from the Genotype-Tissue Expression (GTEx) Project <https://www.ebi.ac.uk/arrayexpress/files/E-MTAB-5214/>". The subject is RNA and GTEx. The "How to cite" section provides the citation: Idaszak, R. (2018). E-MTAB-5214 - RNA-seq from 53 human tissue samples from the Genotype-Tissue Expression (GTEx) Project, CommonsShare, <http://minid.bdk.org/minid/landingpage/ark:/99999/fk4bk2m011>, <http://n2t.net/ark:/99999/fk4bk2m011>. The resource is shared under the Creative Commons Attribution CC BY license. The sharing status is Published. The content section shows a list of files, including "E-MTAB-5214.sdrf.csv" (14.6 MB, csv File).

CommonsShare and fetch.txt

- CommonsShare
 - In CommonsShare BDBags, all payload is always referenced in fetch.txt
 - “...a fully complete BDBag with all data present in its data directory is semantically equivalent to an empty BDBag with its remote dependencies specified in the fetch.txt file...”*
 - Practically speaking, our Helium+ NHLBI DC use cases involve multi-TB BDBag payloads, so the fetch.txt-only approach reasonably ensures that the BDBag will reasonably be servable by http requests
 - i.e. a BDBag using only fetch.txt and no local payload will likely mostly stay under a GB thus servable by http

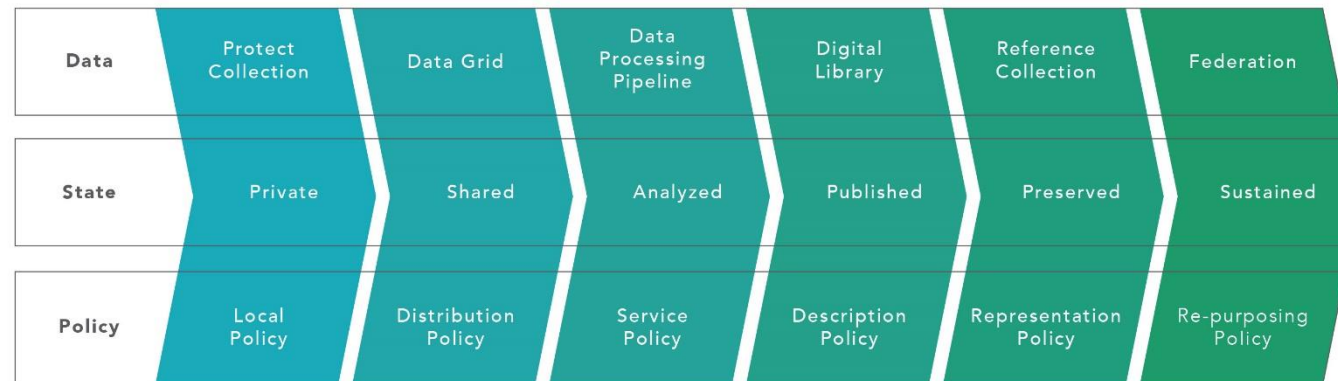
*K. Chard et al., “I’ll take that to go: Big data bags and minimal identifiers for exchange of large, complex datasets,” in Big Data (Big Data), 2016 IEEE International Conference on. IEEE, 2016, pp. 319– 328; <https://doi.org/10.1109/BigData.2016.7840618>.

CommonsShare: Support for Data Lifecycle



From Ingest to Institutional Repository

DATA LIFECYCLE



iRODS virtualizes the stages of the data lifecycle through policy evolution.

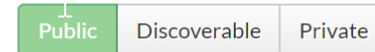
As data matures and reaches a broader community, data management policy must also evolve to meet these additional requirements.

Pre-minid Helium Team CommonsShare resources are mutable

- CommonsShare

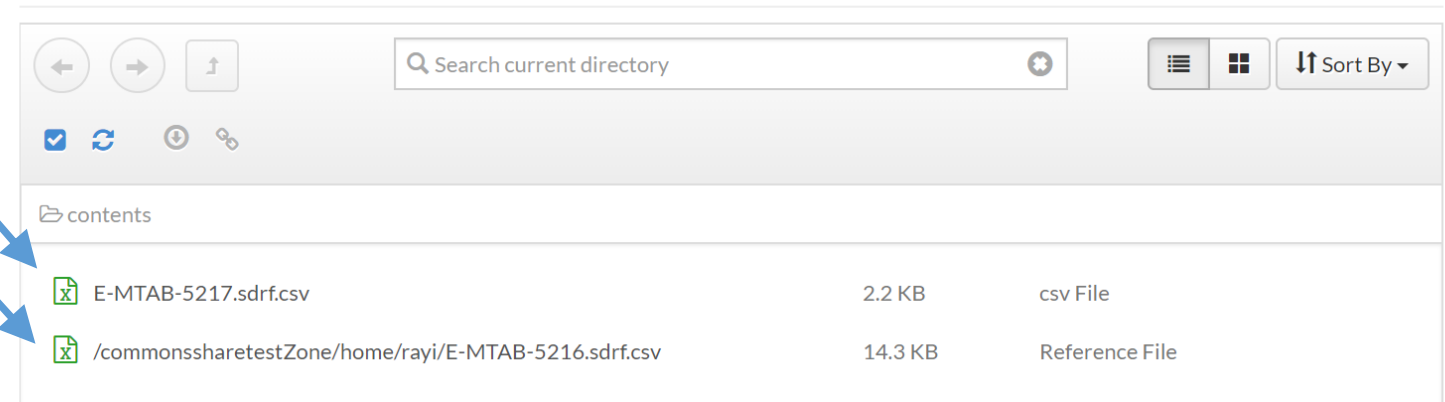
- In CommonsShare, resources are assigned a GUID at time of creation
- Resources are mutable until they are published with a minid
- The idea is in supporting the full data lifecycle, artifacts and metadata are added intermittently to resources until ready to be published
- The owner determines when to make a resource immutable so as to only create the ID provenance chain when ready and relevant
 - There will still be an audit trail, though primarily for handling sensitive data

Sharing status:



🔒 You are the owner of this resource.

Content

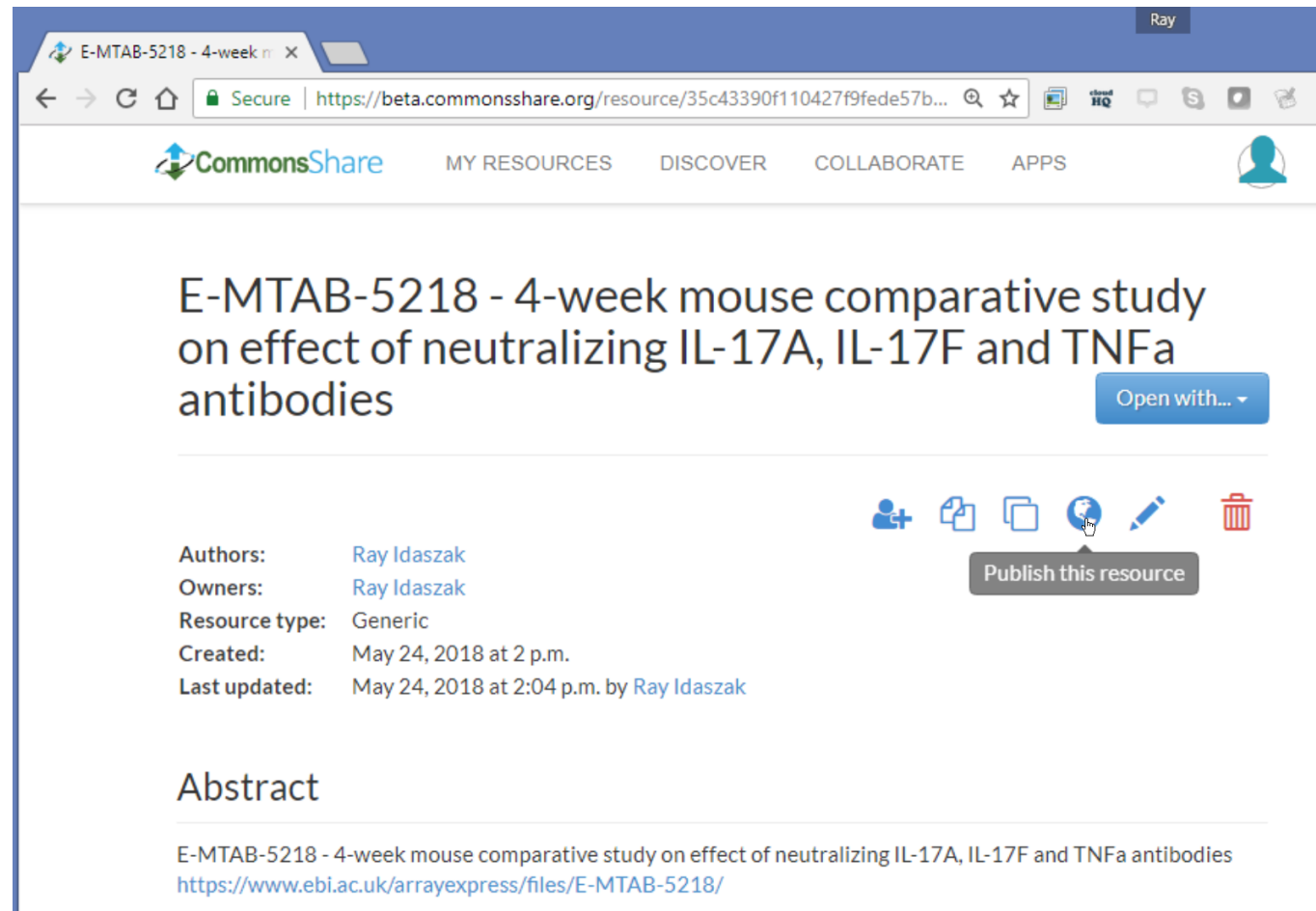


CommonsShare Identifiers

- Helium Framework Services and Interop: KC2-- Global Identifiers and metadata
 - “We have extended our identifier management layer to comply with the BD2K Minimum Viable Identifier (Minid) proposed by the Argon team. Team Helium also provides PrefixCommons, which supports robust curation of identifier prefixes for standardizing identifier resolution. The content within PrefixCommons is synchronized with N2T.net and Identifiers.org. Since Minids build on ARK, *we will use N2T.net as the primary resolver to resolve identifiers globally*, however we will also utilize identifiers.org as the secondary resolver as we believe that it is important to have more than one resolver in case of outage or other issues. We have incorporated the code and services developed by the Argon team to mint and register minids, and are curating all prefixes from the three data stewards within PrefixCommons to support resolution of all identifiers within any stack.”
 - https://docs.google.com/document/d/1R0Ucl6R6h0AjqwqOPr1f8_IW_6ZP8SBITB_25jGb41U/edit?usp=sharing

CommonsShare minting a minid

- Once a CommonsShare mutable resource with GUID is ready to be published as immutable with a minid, the user selects the “Publish this resource” icon on the resource landing page



The screenshot shows a web browser window with the CommonsShare interface. The title of the resource is "E-MTAB-5218 - 4-week mouse comparative study on effect of neutralizing IL-17A, IL-17F and TNFa antibodies". The user "Ray" is logged in. The page includes a navigation bar with "MY RESOURCES", "DISCOVER", "COLLABORATE", and "APPS". Below the title, there is a button "Open with...". A row of icons for sharing and actions is visible, including a globe icon with a tooltip that says "Publish this resource". The metadata section lists the author as Ray Idaszak, the owner as Ray Idaszak, the resource type as Generic, the creation date as May 24, 2018 at 2 p.m., and the last updated date as May 24, 2018 at 2:04 p.m. by Ray Idaszak. The abstract section is partially visible at the bottom.

E-MTAB-5218 - 4-week mouse comparative study on effect of neutralizing IL-17A, IL-17F and TNFa antibodies

Open with...

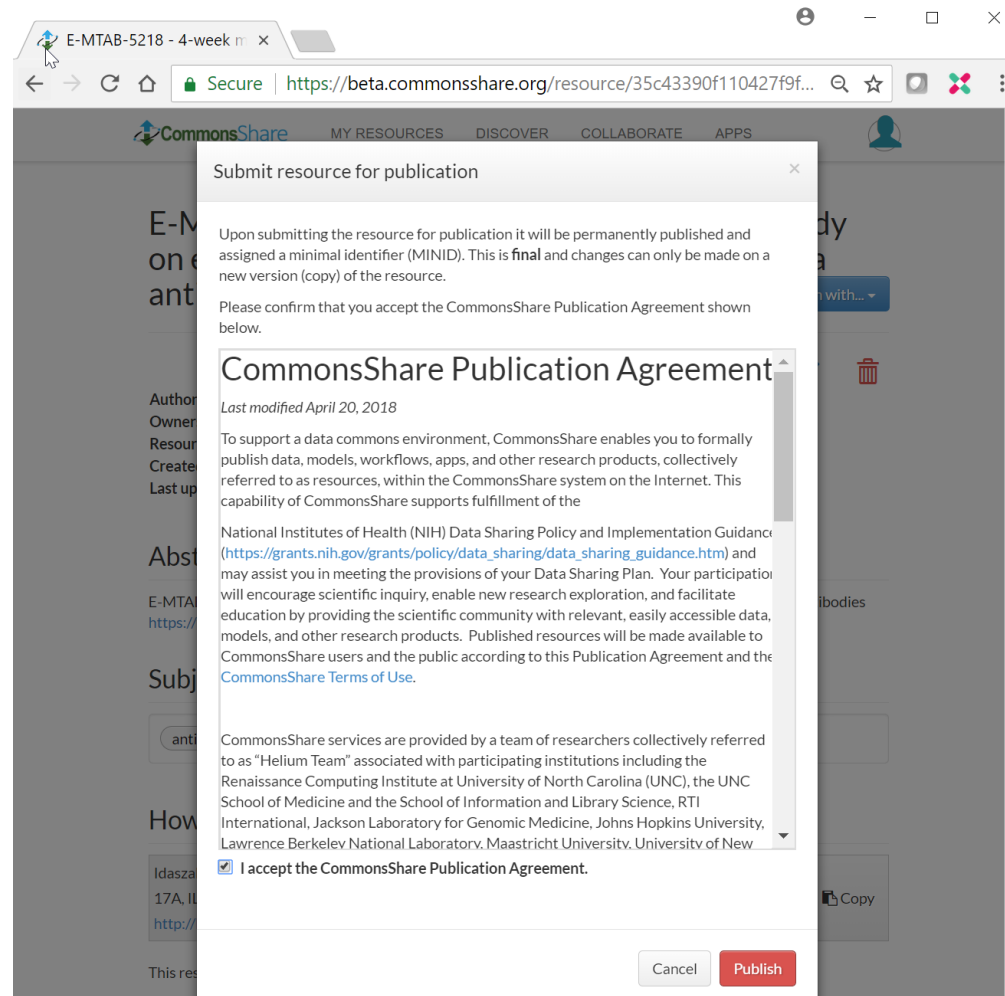
Authors: Ray Idaszak
Owners: Ray Idaszak
Resource type: Generic
Created: May 24, 2018 at 2 p.m.
Last updated: May 24, 2018 at 2:04 p.m. by Ray Idaszak

Abstract

E-MTAB-5218 - 4-week mouse comparative study on effect of neutralizing IL-17A, IL-17F and TNFa antibodies
<https://www.ebi.ac.uk/arrayexpress/files/E-MTAB-5218/>

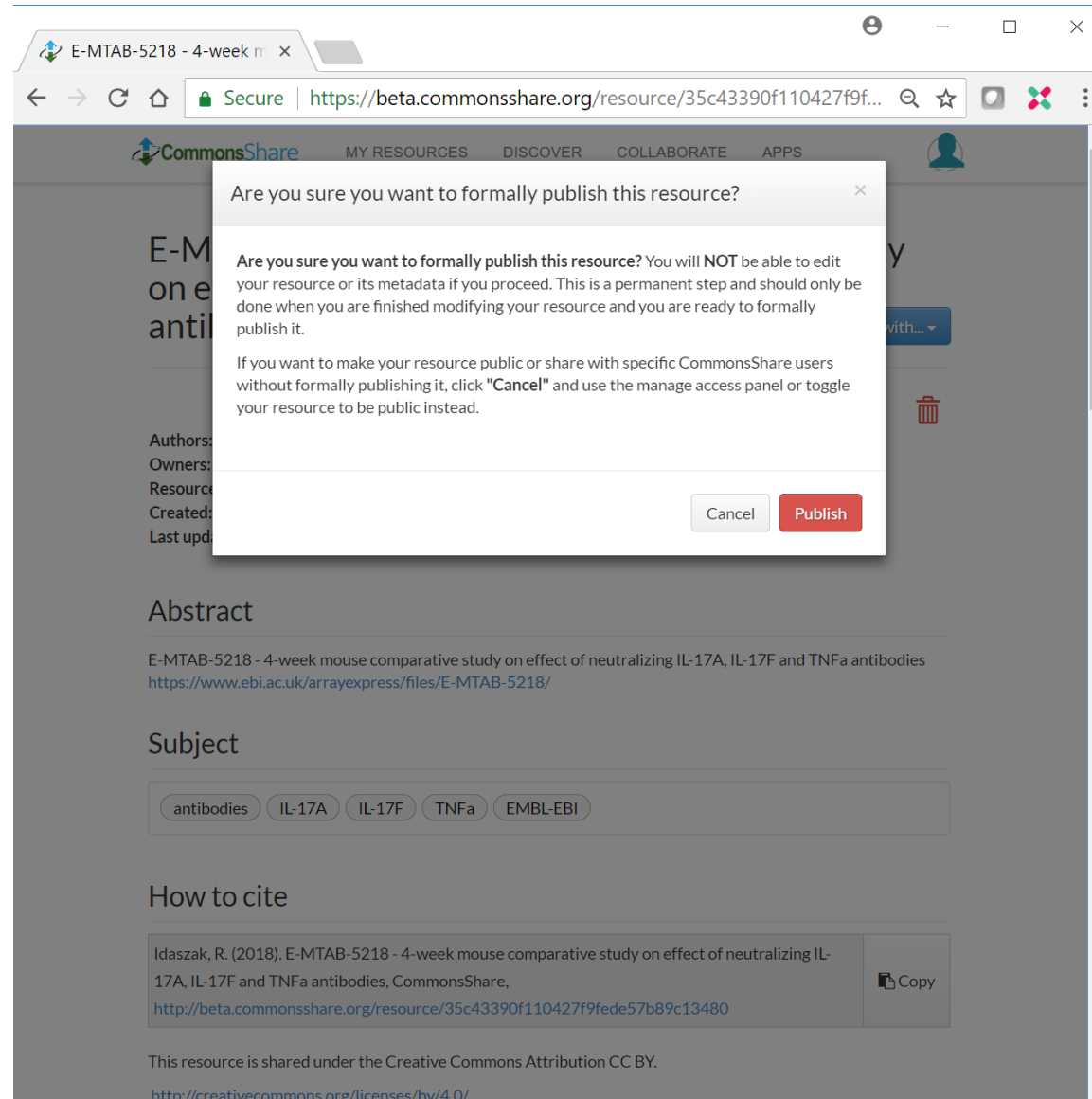
CommonsShare minting a minid

- Upon selecting “Publish this resource” a CommonsShare Publication Agreement is presented for the user to accept as they Publish.
 - Text is here: <https://www.commonsshare.org/publication-agreement/>



CommonShare minting a minid

- ...along with one more “Are you sure” message.



CommonsShare minting a minid

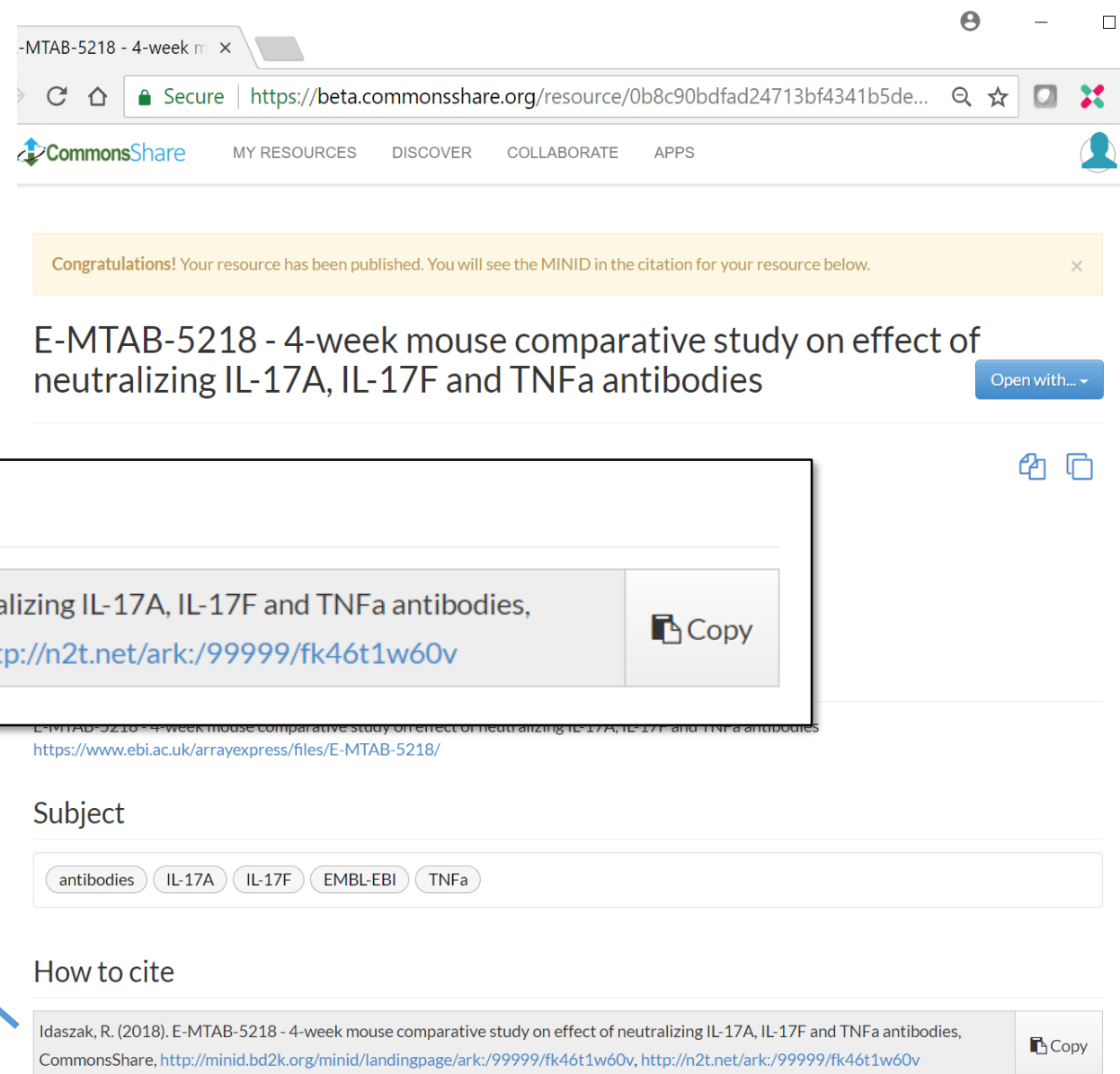
- Upon Publication, the “How to cite” on the resource landing page changes from the GUID to the minid and N2T URIs

The screenshot shows the CommonsShare interface with a modal dialog box open. The dialog asks: "Are you sure you want to formally publish this resource?" and provides instructions: "You will NOT be able to edit your resource or its metadata if you proceed. This is a permanent step and should only be done when you are finished modifying your resource and you are ready to formally publish it. If you want to make your resource public or share with specific CommonsShare users without formally publishing it, click 'Cancel' and use the manage access panel or toggle your resource to be public instead." The dialog has "Cancel" and "Publish" buttons. In the background, the resource page for "E-MTAB-5218 - 4-week mouse comparative study on effect of neutralizing IL-17A, IL-17F and TNFa antibodies" is visible. The "How to cite" section shows a citation using a GUID: <http://beta.commonsshare.org/resource/35c43390f110427f9fede57b89c13480>. A blue arrow points from this GUID to the minid URL in the next screenshot.

The screenshot shows the CommonsShare interface after successful publication. A yellow banner at the top says: "Congratulations! Your resource has been published. You will see the MINID in the citation for your resource below." The resource page for "E-MTAB-5218 - 4-week mouse comparative study on effect of neutralizing IL-17A, IL-17F and TNFa antibodies" is displayed. The "How to cite" section now shows a citation using minid and N2T URIs: "Idaszak, R. (2018). E-MTAB-5218 - 4-week mouse comparative study on effect of neutralizing IL-17A, IL-17F and TNFa antibodies, CommonsShare, <http://minid.bd2k.org/minid/landingpage/ark:/99999/fk46t1w60v>, <http://n2t.net/ark:/99999/fk46t1w60v>". A blue arrow points from the GUID in the previous screenshot to the minid URL in this one.

CommonsShare minting a minid

- The resource then becomes Published with a minid and N2T identifier
 - All editing functions and icons are removed from the published resource's landing page ensuring its immutability
- As stated earlier, Helium team uses N2T.net as the primary resolver to resolve identifiers globally



The screenshot displays the CommonsShare web interface. At the top, a browser window shows the URL <https://beta.commonsshare.org/resource/0b8c90bdfad24713bf4341b5de...>. Below the navigation bar, a yellow banner reads: "Congratulations! Your resource has been published. You will see the MINID in the citation for your resource below." The resource title is "E-MTAB-5218 - 4-week mouse comparative study on effect of neutralizing IL-17A, IL-17F and TNFa antibodies". A blue button labeled "Open with..." is visible. On the right, there are icons for sharing and copying. Below the title, a "How to cite" section is highlighted with a blue box. It contains the citation: "Idaszak, R. (2018). E-MTAB-5218 - 4-week mouse comparative study on effect of neutralizing IL-17A, IL-17F and TNFa antibodies, CommonsShare, <http://minid.bd2k.org/minid/landingpage/ark:/99999/fk46t1w60v>, <http://n2t.net/ark:/99999/fk46t1w60v>". A "Copy" button is next to the citation. Below this, the "Subject" section shows tags: "antibodies", "IL-17A", "IL-17F", "EMBL-EBI", and "TNFa". Another "How to cite" section is visible at the bottom, identical to the one above.

How to cite

Idaszak, R. (2018). E-MTAB-5218 - 4-week mouse comparative study on effect of neutralizing IL-17A, IL-17F and TNFa antibodies, CommonsShare, <http://minid.bd2k.org/minid/landingpage/ark:/99999/fk46t1w60v>, <http://n2t.net/ark:/99999/fk46t1w60v>

Copy

Subject

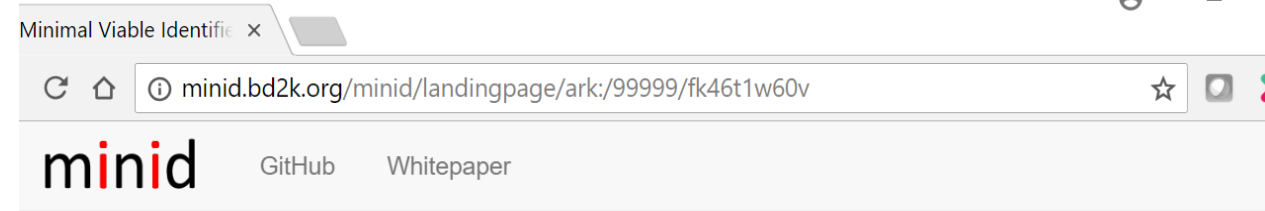
antibodies IL-17A IL-17F EMBL-EBI TNFa

How to cite

Idaszak, R. (2018). E-MTAB-5218 - 4-week mouse comparative study on effect of neutralizing IL-17A, IL-17F and TNFa antibodies, CommonsShare, <http://minid.bd2k.org/minid/landingpage/ark:/99999/fk46t1w60v>, <http://n2t.net/ark:/99999/fk46t1w60v>

Copy

CommonsShare minid landing page



Link back to CommonsShare resource landing page
{Note: in discussion after this presentation, it was pointed out a new field needs to be created on the minid landing page for this type of location reference and will consequently be pursued}

BDBag .zip'ed contents

Identifier:
[minid:fk46t1w60v](#)

Created:
2018-05-25 03:51:41.911331

Creator:
CommonsShare ([None](#))

Checksum:
TEST-6456fe1b10a0edf1e3f51c3d292c1397282637a08bfdbb0b008cc715fa0c69da

Status:
ACTIVE

Locations:
<http://beta.commonsshare.org/resource/0b8c90bdfad24713bf4341b5de6dd83a>
http://beta.commonsshare.org/django_irods/download/bags/0b8c90bdfad24713bf4341b5de6dd83a.zip

Titles:
MINID for E-MTAB-5218 - 4-week mouse comparative study on effect of neutralizing IL-17A, IL-17F and TNFa antibodies



CommonsShare Minid & BDBag Future Work

- Helium Team Bring Your Own Data (BYOD)
 - Support for Open Science Framework, Dropbox, and additional support for Globus
 - Support for ingest from other iRODS zones
- Extended metadata considerations
 - CommonsShare supports entering extended metadata directly on the resource landing page
 - Aligning metadata models
- Emerging CommonsBag considerations

