



# Support for content URLs Datacite open hours November 2021

Ari :Lukkarinen



# Background

- CSC
  - State and university owned company providing IT services to research institutions and academia.
  - Lot of data management activities going on all the time.
- Research data management projects including, e.g.
  - EUDAT, EUDAT H2020, DICE, EOSC Nordic
  - National and EU level PID activities (ePIC, Datacite)
- EOSC Nordic - *Integrated Data Management Workflows*

# Content URLs and workflows

WORKFLOW (using DOIs):

- Locate, find and download a dataset
- Extract required data from transfer package
- Rerun the analysis defined by workflow script

Access to data through landing page is not enough !

If direct access to data would be possible, then:

- automation with scripts could be done,
- reproducibility of research results could be enhanced, and
- perhaps have an impact to quality of data.

Script could most likely be understood also by colleagues

## Vision - personal opinion

- PIDs should support machines
- Dataset (and PID) metadata attributes should contain ALL the necessary information required to recreate the dataset **automatically**.
- Other activities, that may be relevant
  - <https://the-turing-way.netlify.app/reproducible-research/compendia>
  - W3C DCAT vocabulary classes (DCAT:Dataset, DCAT:Distribution, DCAT:DataService) (<https://www.w3.org/TR/vocab-dcat-2/#dcat-scope>)



**Ari Lukkarinen**

Thank You !



[facebook.com/CSCfi](https://facebook.com/CSCfi)



[twitter.com/CSCfi](https://twitter.com/CSCfi)



[linkedin.com/company/csc--it-center-for-science](https://linkedin.com/company/csc--it-center-for-science)



[github.com/CSCfi](https://github.com/CSCfi)