Libraries & research data

András Holl – Library, Hungarian Acad. Sci.

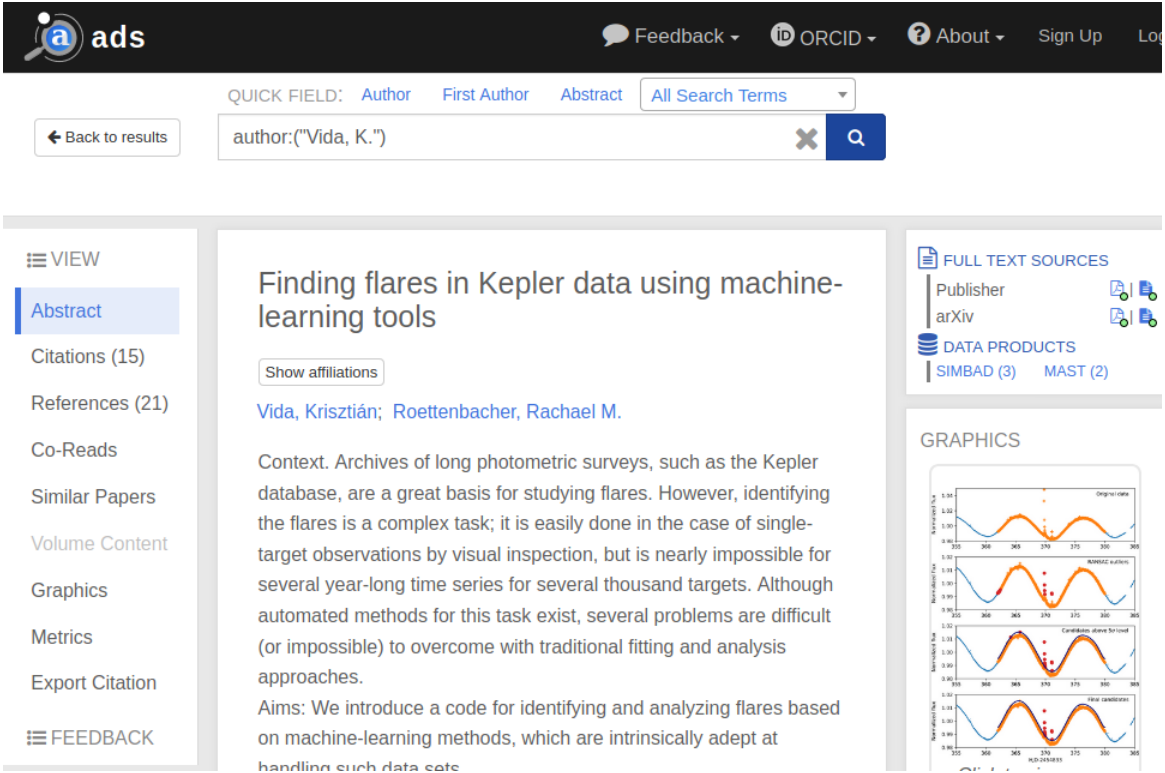
LIBER/SKS: Research Libraries and the EOSC

January 20th, 2021

The long tail of research data

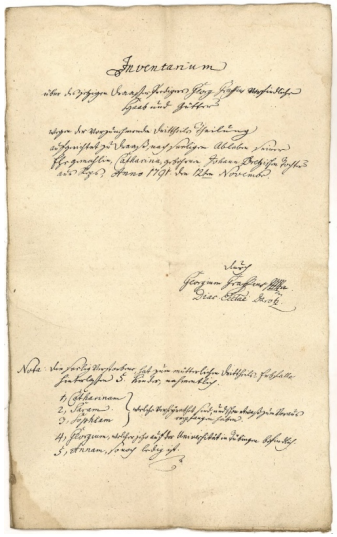
„Big Science" (a term of *de Solla Price*) can easily make research data FAIR. Large projects build pipelines to produce and process data – these pipelines could be modified to fit the FAIR principles. Space agencies and other large, well funded organisations do have the means.

An example from Big Science: data from the Kepler satellite stored at Mikulski Archives for Space Telescopes.


ads
Feedback
ORCID
About
Sign Up
QUICK FIELD: Author First Author Abstract All Search Terms
Back to results
author:("Vida, K.")
VIEW
Abstract
Citations (15)
References (21)
Co-Reads
Similar Papers
Volume Content
Graphics
Metrics
Export Citation
FEEDBACK
Finding flares in Kepler data using machine-learning tools
Show affiliations
Vida, Krisztián; Roettenbacher, Rachael M.
Context. Archives of long photometric surveys, such as the Kepler database, are a great basis for studying flares. However, identifying the flares is a complex task; it is easily done in the case of single-target observations by visual inspection, but is nearly impossible for several year-long time series for several thousand targets. Although automated methods for this task exist, several problems are difficult (or impossible) to overcome with traditional fitting and analysis approaches.
Aims: We introduce a code for identifying and analyzing flares based on machine-learning methods, which are intrinsically adept at
FULL TEXT SOURCES
Publisher
arXiv
DATA PRODUCTS
SIMBAD (3) MAST (2)
GRAPHICS

On the other hand, small projects will face a big challenge.

Individual scientists or small teams, in fields and in institutions not well founded, produce research data too – this is the long tail. Such data is important! (Though they might look different…)



```
# IBVS 6223-t3.txt
#Author: TERRELL, D. and NELSON, R.H.
#Telescope: 25 cm f6.3 Schmidt-Cassegrain
#Detector: SBIG ST-7XE CCD camera
#IBVSdataKey: photometry
#Object: Z Dra
# Caption: Z Dra - Photometric Data
# Filters: Johnson V,B

# Filter: V
# HJD-2400000   magnitude
53432.57290 10.519000
53432.57570 10.551000
53432.57850 10.547000
```

Gulyás Pál: Magyar írók élete és munkái. Cédulatár

Készült az Informatikai és Hírközlési Minisztérium támogatásával

Gyorsválasztó: P R S T U V W X Y Z

Cédulák száma:
27138

p (4470)
Paál - Pajor
Pajor - Pál
Pál - Palásthy
Pálfi - Pálfi
Pálfy - Pallós
Pállya - Palotás
Palotás - Pannuhi
Panus - Papp Ferenc
Papp Ferenc - Papp Jeno
Pap Jeno - Pap Mihály
Papp - Pápay
Pápai - Parragh
Parragh - Pásztói
Pásztor - Patak
Patak - Páter
Páter - Paul
Paul - Pável
Pável - Pécs
Péch - Pecz
Pecz - Pekár
Pekár - Perci

Összesített lista

2701.

Petrides Jenő

2735.

Petrók Gyula

Közig.példatár.(M.$68

2768.

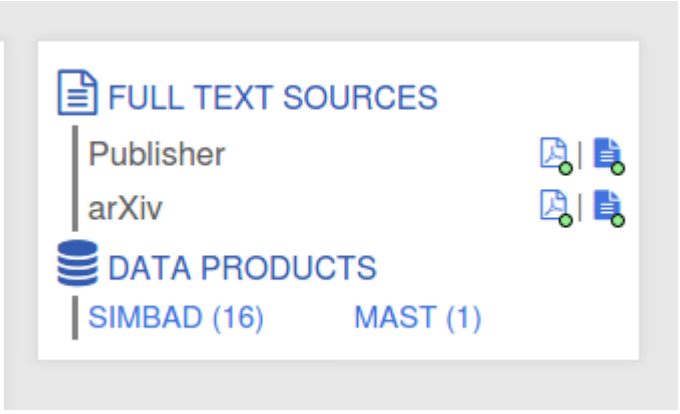
Petrovics János

Making research data FAIR – the library way

Libraries have long experience of making documents discoverable using metadata, identifiers and classification systems. They have a long history of helping researchers. This expertise could be utilised for making sporadic data packages more FAIR.

– Ask your librarian for advice!

– The publication connection



- Data can be found through the related publication

– Discovery level metadata, Dublin Core descriptions


```
−<resource xsi:schemaLocation="http://datacite.org/schema/kernel-4 http://schema.datacite.org
/meta/kernel-4.1/metadata.xsd">
  <identifier identifierType="DOI">10.18426/obm.36vn3g36r3m0</identifier>
 −<creators>
  −<creator>
     <creatorName>OBMQuery dinpi.openbiomaps.org</creatorName>
   </creator>
  </creators>
 −<titles>
  −<title>
     Database query for Violet click beetle from Duna-Ipoly National Park species occurrence
     database (doi: 10.18426/obm.23mj4em6bca0)
   </title>
  </titles>
  <publisher>Duna-Ipoly National Park</publisher>
  <publicationYear>2017</publicationYear>
 −<subjects>
    <subject xml:lang="eng">OpenBioMaps</subject>
    <subject xml:lang="eng">species occurrences</subject>
    <subject xml:lang="eng">Limoniscus violaceus</subject>
    <subject xml:lang="eng">Violet click beetle</subject>
    <subject xml:lang="hun">kék pattanó</subject>
  </subjects>
 −<contributors>
  −<contributor contributorType="DataCollector">
     <contributorName>Bérces, Sándor</contributorName>
```

- **Data quality and certification**

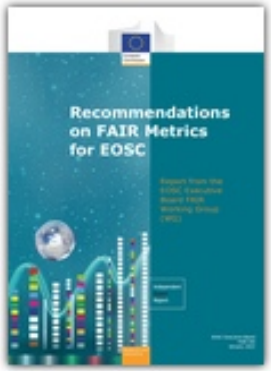
  - **The referee of the publication must see the data!**

  - **The library might consider a CoreTrustSeal certification for the repository.**

– Data standards

EOSC FAIR WG Metrics & Certification documents

- ***Recommendations on FAIR metrics for EOSC DOI: 10.2777/70791***
- ***Recommendations on certifying services required to enable FAIR within EOSC DOI: 10.2777/127253***



– Technical issues

- **Keep it simple!**
  (TXT, XML, JSON, PDF, etc.)

Text
CompressedTubeContactMeasurements_figure.pdf
Download (1MB) | Előzetes bemutató

Text
CompressedTubeContactMeasurements_1.csv
Download (37kB)

- Legal or re-use issues

  - sensitive areas, indirect checks
  - re-use licence

– Grey literature – documentation

- **Important library task – archive the documentation!**
  (Technical documentation, DMPs, project reports, etc.)

– PIDs

- CHECK
- USE
- PROVIDE

My library (Library and Information Centre, Hungarian Academy of Sciences) is a member of both CrossRef and DataCite.

– FAIR is a journey

(Kumano Kodo by David West)

Training and support

Which organisational unit will be responsible for the research data? The library is an option.

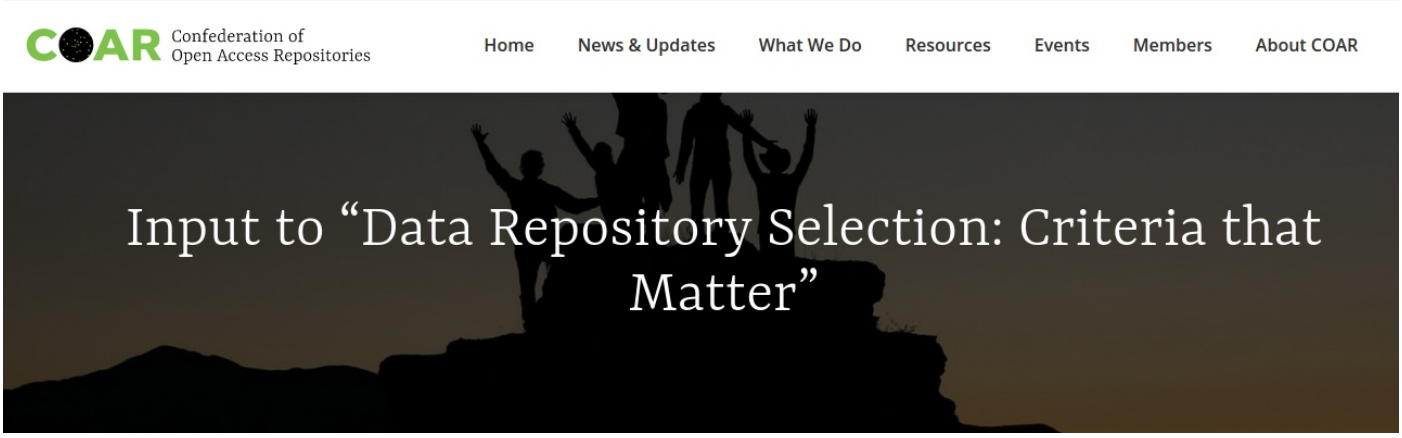
Who will train the data stewards, and who will train and advise the researchers? The library is an option.

Broad participation

If we set the threshold of FAIRness too high, we would loose datasets and communities. If it is too low, we would loose functionality. That is the reason for the slogan "FAIR is a journey". Try to make your data discoverable and accessible. Then You might improve your data handling practices.

Publishers try to dictate the terms of archival – and libraries seem to have a different view.

See [COAR's statement] on the issue.



Questions?

Thank You!