

VYTAUTO  
DIDŽIOJO  
UNIVERSITETAS  
M C M X X I I

# Atvirojosios duomenų prieigos ir tinkamo informacijos kodavimo svarba dirbtiniam intelektui ir duomenų kasinėjimui

Tomas Krilavičius ir Rokas Maliukevičius

PUBLISH

or

PERISH



**DEPENDENT SOURCE EXCEPTION**

Section 107(b)(1)(B) provides an exception to the general rule that the fair use doctrine applies to all copyrighted works. This exception applies to certain types of copying, including copying for the purpose of teaching, scholarship, or research, and copying for the purpose of creating new works for educational purposes. The exception is limited to copying that is done by or for an educational institution, and the copying must be for a non-commercial purpose.

**FAIR USE**

Section 107 of the Copyright Act provides a defense against a claim of copyright infringement. The fair use doctrine is a flexible, case-by-case doctrine that allows limited copying of copyrighted material without the permission of the copyright owner. The factors that courts consider in determining whether a use is fair include the purpose and character of the use, the nature of the copyrighted work, the amount and substantiality of the portion copied, and the effect of the copying on the market for the original work.

**SITUA  
ARE N**

**NOTE ON FAIR USE:** It is a public domain work if it is a work that is not protected by copyright law. This is because the copyright term has expired, or the work was never eligible for copyright protection.

**EXAMPLE FROM THE DOCTRINE:** The position is...

Pacituokite mane!!!

prašau



Kaip straipsnius  
padaryti randamus?



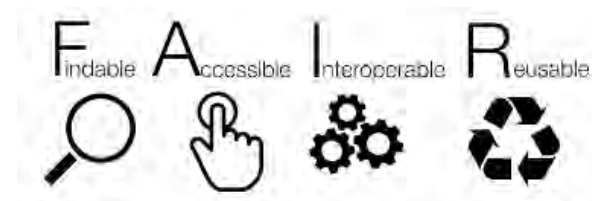
O kas mums gali padēti?

Leidējai



# FAIR Principles

# Compliance



## Findability

Resource and its metadata are easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.

- ✓ F1. Resource is uploaded to a public repository.
- ✓ F2. Metadata are assigned a globally unique and persistent identifier.



## Accessibility

Resource and metadata are stored for the long term such that they can be easily accessed and downloaded or locally used by humans and ideally also machines using standard communication protocols.

- ✓ A1. Resource is accessible for download or manipulation by humans and is ideally also machine readable.
- ✓ A2. Publications and data repositories have contingency plans to assure that metadata remain accessible, even when the resource or the repository are no longer available.



## Interoperability

Metadata should be ready to be exchanged, interpreted and combined in a (semi)automated way with other data sets by humans as well as computer systems.

- ✓ I1. Resource is uploaded to a repository that with other platforms.
- ✓ I2. Repository meta- data schema maps to CG Core metadata schema.
- ✓ I3. Metadata use standard vocabularies and



## Reusability

Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines and humans.

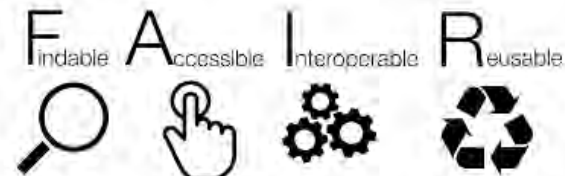
- ✓ R1. Metadata are released with a clear and license.
- ✓ R2. Metadata about data and datasets are r a plurality of accurate and relevant attribute





# FAIR Principles

# Compliance



## Findability

Resource and its metadata are easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.

- ✓ F1. Resource is uploaded to a public repository.
- ✓ F2. Metadata are assigned a globally unique and persistent

# if DATA is KING

# then METADATA is QUEEN

Provenance –  
Track Data  
and Reporting

## Reusability

Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines and humans.

- ✓ R1. Metadata are released with a clear and license.
- ✓ R2. Metadata about data and datasets are r a plurality of accurate and relevant attribute





Home

Search

## Services



### CLARIN portal

Get an example-based impression of what's currently available



### Depositing services

Store language resources in a sustainable repository at a CLARIN centre



### Virtual Language Observatory

Discover language resources using a faceted browser or a map



### Easy access to protected resources

Get easy access to protected resources, with your institutional username and password.



### Language Resource Switchboard

Explore and analyze language data with a wide variety of tools



### Language Resource Inventory

Submit and access information about language resources relevant to your research.



### Content Search (prototype)

Search different corpora with a single search engine



### Question:

Searching for an application? Would you like to assist your research? Feel free to contact us!



This XML file does not appear to have any style information associated

```

▼ <CMD xmlns="http://www.clarin.eu/cmd/" xmlns:xsi="http://www.w3
xsi:schemaLocation="http://www.clarin.eu/cmd/
http://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/pro
<!-- General Information Header -->
▼ <Header>
  <MdCreator>Dieter Van Uytvanck</MdCreator>
  <MdCreationDate>2009-11-18</MdCreationDate>
  <!-- A link to this very file -->
▼ <MdSelfLink>
  http://www.clarin.eu/cmd/example/example-md-instance.cmdi
</MdSelfLink>

```

O toliau – techniškiau



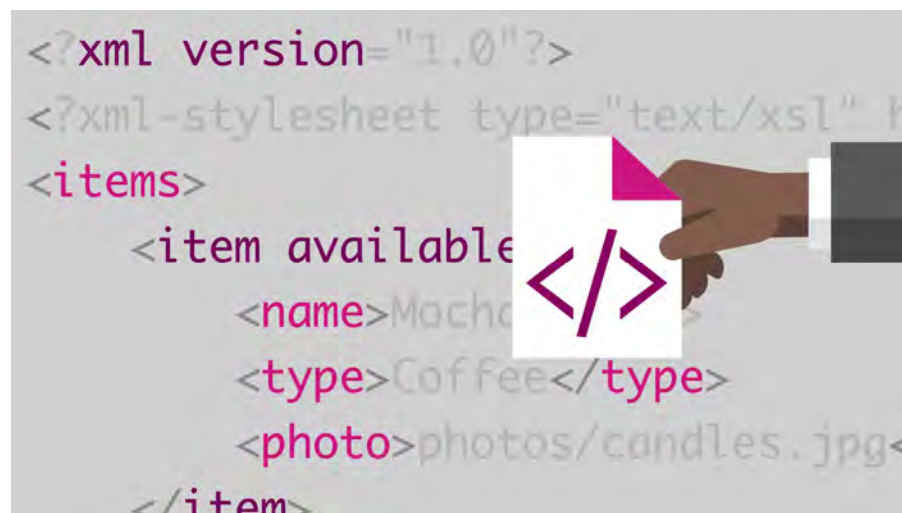
# Dirbtinis intelektas ir TDG su atvirais duomenimis

- “ 3,3 milijono santraukų
- “ 1000 žurnalų
- “ 1922–2018
- “ Kristalinė gardelė ir periodinė elementų lentelė
- “ TOP 3 daugiau nei 95 %
- “ Sunku atrasti prasmę
- “ DI dideliems kiekiams informacijos



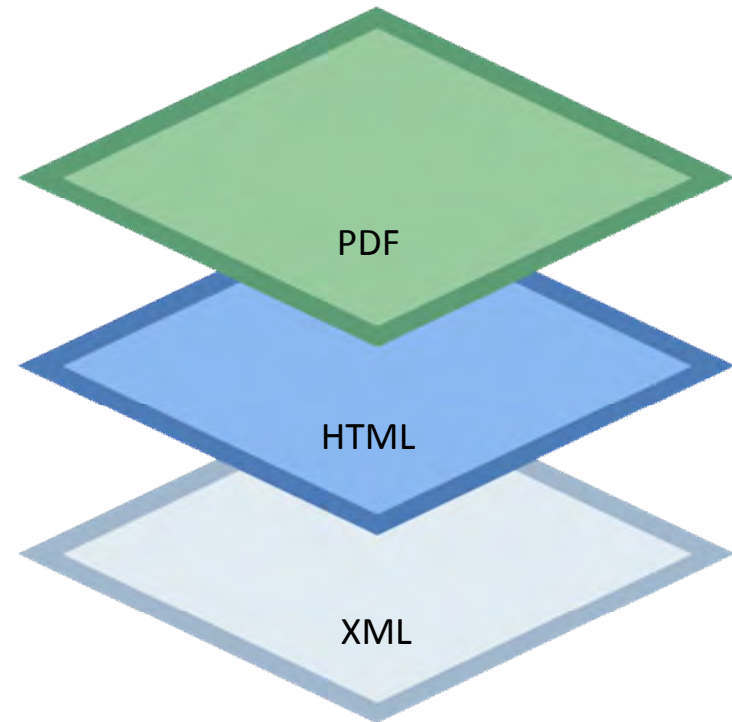
# Turinys

- “ Kodėl reikia XML?
  - “ Atrandamumas per duomenų bazines
  - “ Resursų taupymas
  - “ Atrandamumas internete
- “ Standartas straipsnių publikavimui
  - “ JATS
  - “ H&T XML vs. FT XML
- “ Ateitis ir tikslai
  - “ Reikalavimai
  - “ Siūlymai
  - “ Tikslai ir nauda



## Kodėl reikia XML?

- “ Leidžia nurodyti žodžio prasmę
- “ PDF HTML XML
- “ Žurnalo turinys – į duomenų bazes
- “ Įkėlimas rankomis vs. XML
- “ Tik PDF be XML



# Standartas straipsnių publikavimui

- “ Yra įvairių XML standartų
- “ NISO sukūrė JATS
- “ JATS – XML elementų ir atributų rinkinys
- “ PubMed, MEDLINE, Index Copernicus...
- “ Daug panaudojimo pavyzdžių



# Standartas straipsnių publikavimui

- ” Metaduomenys priskiriami per DOI
- ” Pilnesni metaduomenys – daugiau naudos
- ” H&T XML vs. FT XML
- ” Didžiulis potencialas
- ” Nauda ekonomikai ir visuomenei







## Ateitis ir tikslai

- “ „Plan S“ iniciatyva
- “ Tyrimai remiami visuomeninėmis lėšomis
- “ Privalomi ir atkakliai siūlomi reikalavimai
- “ Laikantis reikalavimų nauda – dvilypė
- “ XML padeda

Plan S

**Making full and  
immediate Open  
Access a reality**

The logo graphic for Plan S consists of several overlapping, semi-circular shapes in various shades of orange and red, arranged in a pattern that suggests a stylized 'S' or a series of concentric arcs.

Trumpai

Publikavimas -> Standartas -> XML -> Nauda ->

-> Didesnė nauda JATS -> Maksimali nauda FT XML ->

-> Kodėl jie privalomi -> Nauda leidėjams ir visuomenei

Jeff Beck iš PubMed:

***JATS is no longer one of the cool kids;  
it's just what you do if you have journal articles.***

<AČIŮ />