

Unstandardised standards – free use of meta data and data retrieval, on the experience of mining Polish digital humanities sets.

Piotr Malak

UMK w Toruniu

Collection

- digital documents, representing traditional and digital publications available via Digital Libraries Federation: <http://fbc.pionier.net.pl>
- 1,714,571 records
- meta data compatible with Dublin Core,
- sociological analysis: ca. 25 % of authors are women

Dublin Core: type

```
<dc:title xml:lang="pl"><![CDATA[Brudnopisy listów  
Eugenii Brończykowej do Czerwonego Krzyża w  
Moskwie i Ministra Spraw Zagranicznych, Wiaczesława  
Mołotowa w sprawie jej syna [nazwa red.].]]></dc:title>
```

```
<dc:type xml:lang="pl"><![CDATA[Brudnopisy listów  
Eugenii Brończykowej do Czerwonego Krzyża w  
Moskwie i Ministra Spraw Zagranicznych,  
Wiaczesława Mołotowa w sprawie jej syna [nazwa  
red.].]]></dc:type>
```

Dublin Core: type

- `<dc:type xml:lang="pl"><![CDATA[info:eu-repo/semantics/article]]></dc:type>`

`<dc:type xml:lang="pl"><![CDATA[33 x 21 cm]]></dc:type>`

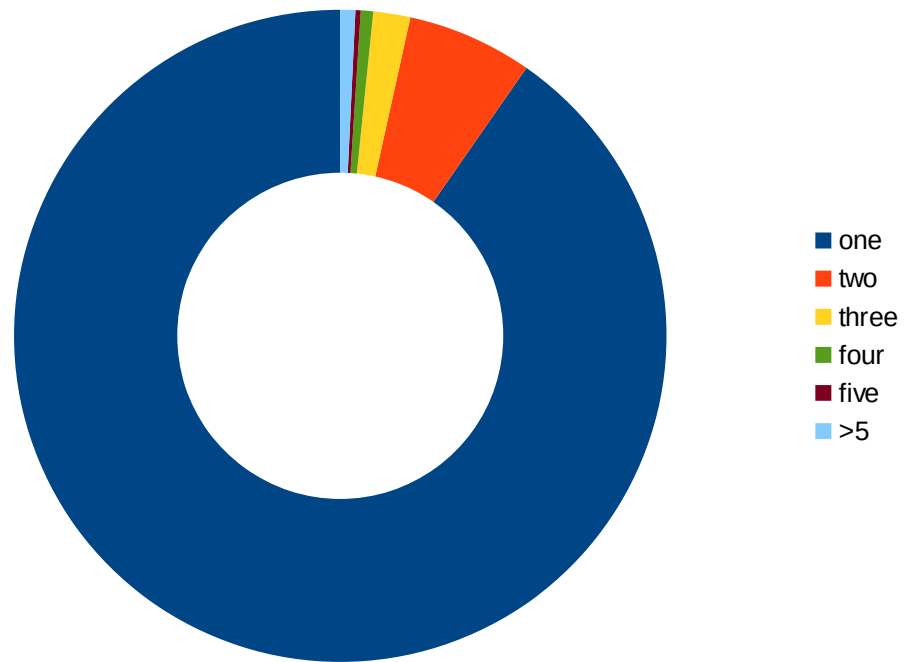
- `<dc:type><![CDATA[Rybicki, Leszek]]></dc:type>`
- `<dc:type><![CDATA[23.07.1989]]></dc:type>`

Authors

- single 365365
- two 25147
- three 7393
- four 2512
- five 1044
- more 3020

- Max: 11 coauthors

Authors



Authors of humanities

Authors pairs and affiliations	number	fraction %
Both affillieted	23438	37,14%
Only one with af.	23530	37,29%
Both without aff.	16140	25,58%

Authors

- Names:
 - KINGA, TAMARA, EWA, Krystyna, DANUTA, Cvetanka, Anna, Zofia, Teresa, Małgorzata, Safa, Ghada, Parisa, Diana, Yehya, Thushara, Krystyna, Renata, Edyta, Janina

Dates

- Dates:

YYYY-MM-DD

DD-MM-YYYY

YYYY.MM.DD

DD.MM.YYYY

YYYY-YYYY

-YYYY

YYYY

Dates

- Dates:
 - „po YYYY”, (after YYYY)
 - „przed”, („before”)
 - „post ante”,
 - „ok.”
 - „ca ...”

Conclusions

- there are standards and they are widely used
- details make the difference
- for Digital Humanities purposes we need uniform data
- there is need to make data operators (librarian, indexers etc.) aware of the value of uniformed data input
- data access points can do some data normalisation (eg. convert YYYY.MM.DD to DD.MM.YYYY) but not all (like set proper type of publication).

Acknowledgements

The research in the frame of grant
NCN 2013/11/B/HS2/03048

*Badanie struktury i dynamiki rozwoju
cyfrowych zasobów wiedzy przy pomocy
metod wizualizacji*

Thank you for your
attention

Piotr Malak

piomk@umk.pl