"*Acknowledgement of this funding was inadvertently omitted*": Enriching metadata of an institutional research repository by mining grey literature/technical reports for research funding-related information

Rationale

Funding acknowledgements in research papers are an important resource not only for studying the impact of funding on research, but they also present a utility for mapping the funding landscape, co-funding activities, the portfolios of certain funding agencies as well as the research performance of scientists and institutions. Even though the major literature databases Scopus and Web of Science have indexed funding acknowledgements for a number of years, the presence of funding acknowledgements varies in depth and lacks consistency in its usage and detail in research papers.

As it becomes increasingly important for research organizations to show the impact of their research to donor agencies, grey literature/technical reports can serve as a valuable source of information of an organization to complement the missing or incomplete funding acknowledgements in their research information management systems (RIMs). In addition, grey literature can serve as a means to bridge the delay between the research activities and their publication or even the possibility that some research may never be published.

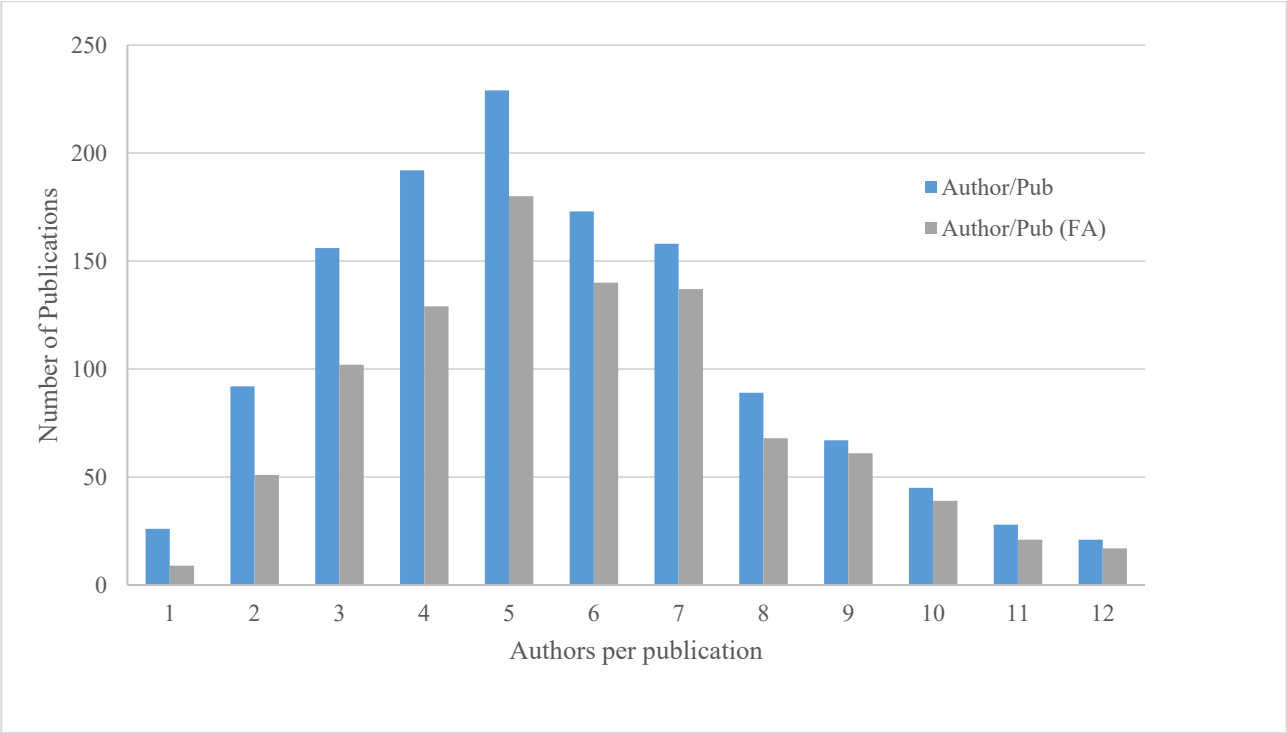
The purpose of this paper is to demonstrate how funding-related information can be extracted from grey literature through text mining and incorporated into the corresponding field of MARC 21 bibliographic records of an institutional research repository of an international agricultural research and development organization for further analysis of its funding and research trends. Enriching the metadata with additional bibliographic notes (funding information) provides additional information about the resource that can be parsed and utilized in a linked open data environment.

Originality/Related Work

There has been great interest in the area of acknowledgment analysis to assess funding acknowledgements (FA) to track research outputs, managing funding portfolios, and evaluate impact of grants, for instance, by examining the amount of funding in a particular field, the type of funding agency by particular fields or at research program funding levels. A number of studies have shown that using multiple data sources to derive performance measures provides for a robust analysis of research impact and improves the validity and reliability of conclusions, for instance, by tracing references of research publications as cited in other artefacts such as policies, regulations, or expert panel reports (Drew et al., 2016), by combining grants administration data, publication and citation data, inventions, and patent data into a single performance management system (Haak et al., 2012), or by using metadata in the grant records to validate the research results and monitoring research trends (Freyman, Byrnes, & Alexander, 2016).

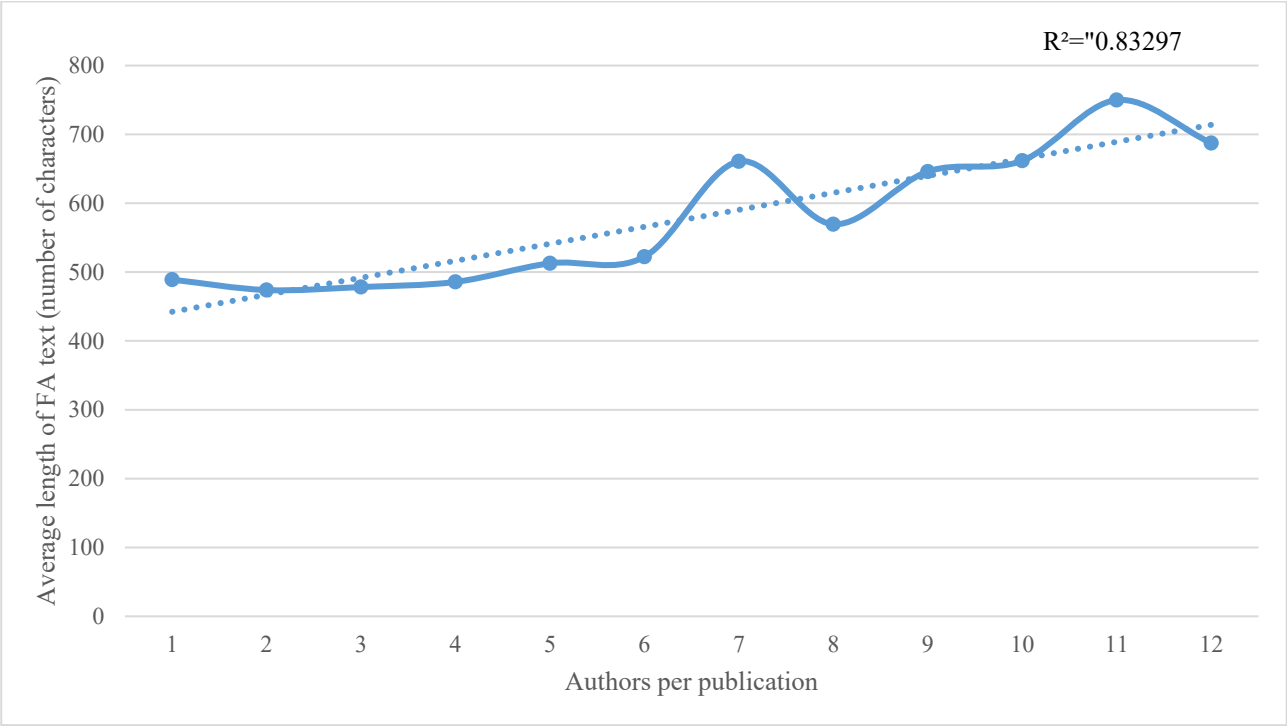
The relevance of tracking funding acknowledgements also becomes clear when we look at the publications of the World Agroforestry Center, an international agricultural research institute of the CGIAR. In **Error! Not a valid bookmark self-reference.** we have included cases of publications with up to 12 co-authors involving ICRAF to show the relationship between authors per publication and authors per publication with funding acknowledgement (FA), respectively.

Figure 1: Authors per publication and publication with funding acknowledgement (FA) (2008-2017)



As shown in **Error! Not a valid bookmark self-reference.** there is a clear linear relationship with the number of authors per publication and the length of the FA text. Publications with more than 12 co-authors do not show this strong correlation.

Figure 2: Relationship between the length and presence of funding acknowledgement (FA) and the number of authors per publication



Yet, the provision of funding acknowledgements is not consistently applied because of a lack of standardization of including funding information as well as cultural and political biases affecting how funding sources are acknowledged (Rigby, 2011). While the importance of funding acknowledgement analysis is highlighted to help attribute the contribution of funding organizations towards research activities, the need to be more consistent and structured in providing this information is reiterated (Costas & van

Leeuwen, 2012; Costas & Yegros-Yegros, 2013). The trends and changes in the occurrence of acknowledgements of particular bodies or certain journals and disciplines or even the nationality of the authors can vary greatly (Grassano, Rotolo, Hutton, Lang, & Hopkins, 2017; Kokol & Vošner, 2018). This shows that much of the funding-related information seems underutilized. For instance, there is an insignificant usage of field 536 of less than 1% of all MARC records in the OCLC WorldCat catalog.[1]

However, additional bibliographic notes such as funding information represent an untapped resource. These notes can be enriched and have the potential for future harvesting in linked open data systems as shown in a recent analysis of notes in MARC 508 and 511 note fields for creation/production credits and participants or performers, respectively, as well as the use of relator notes (Weitz, Toves, Vizine-goetz, Naught, & Bremer, 2016).

Research methods

Due to its institutional setup, the World Agroforestry Centre tracks research-related information in different systems. While publications are contained in an institutional repository[2] along with bibliographic records on research datasets, these datasets are maintained in a separate system.[3] Pre-award information (e.g., concept notes, proposals) and grant- and award-related documentation (e.g., grant agreements, technical donor reports) are contained in their own internal systems. In order to overcome the lack of a complete, interlinked research information environment (Jeffery & Asserson, 2009), the Centre is enhancing the bibliographic records of its repository by indexing funding-related information in its institutional research repository. Linking grant and award records to publication data programmatically would help to better understand the research lifecycle and track publications over time, for instance, computing the lag time between the grant award and corresponding publications (Ihli, 2017; Pappas & Williams, 2011). The Centre's institutional repository, based on Koha, an open source integrated library system using the MARC 21 format for bibliographic data (Library of Congress, 2017), serves as its research information management system (RIM). The Centre is in the process of evolving its RIM into an IR-as-CRIS by incorporating more research-related information that is traditionally associated with a CRIS, such as person or project data (de Castro, Shearer, & Summann, 2014). The Centre's RIM supports its overall research information infrastructure through publications and scholarship expertise, discoverability, access, and support as well as stewardship of the institutional record (Bryant et al., 2017).

Research shows that the MARC 21 format for bibliographic data is suitable for integration with CERIF-based research management systems through its ability to facilitate data exchange and consistently handle authority and publication data (Ivanović, Milosavljević, Milosavljević, & Surla, 2010; Ivanović, Surla, & Konjović, 2011; Ivanović, Surla, & Racković, 2011). Even though the MARC 21 format for bibliographic data has its limitations in terms of reuse in different applications for different purposes, which can be addressed through modifications based on Functional Requirements for Bibliographic Records (FRBR) (Aalberg & Žumer, 2013), the MARC 21 format is still rich in metadata and provides space for more detailed description of entities of RIMs while supporting interoperability through its description in XML (MARCXML)[4].

MARC 5XX fields provide space for bibliographic notes. One field that is relevant in our context is the specialized MARC field 536 for *funding information note*, which contains contract, grant, and project numbers when the material results from a funded project as well as information concerning the sponsor or funding agency (Library of Congress, 2017). It is one of several MARC codes and fields for identifying information needed to describe technical reports. Based on the CERIF classification for publications, a *report* is defined as "a separately published record of research findings, research still in progress, or other

[1] MARC Usage in WorldCat – 536: Funding Information. Retrieved from http://experimental.worldcat.org/marcusage/536.html
[2] World Agroforestry Centre institutional repository (Koha). Retrieved from http://outputs.worldagroforestry.org/
[3] World Agroforestry Centre - ICRAF Dataverse. Retrieved from https://dataverse.harvard.edu/dataverse/icraf
[4] Library of Congress, MARCXML: MARC 21 XML Schema. Retrieved from https://www.loc.gov/standards/marcxml/

technical findings, usually bearing a report number and sometimes a grant number assigned by the funding agency"[5] (Jörg et al., 2010). According to the ANSI/NISO standard Z39.18-2005 (National Information Standards Organization, 2010), scientific and technical reports as a type of grey literature should contain information identifying both the performing organization of a research done under a contract or grant as well as the sponsoring organization of such research activity as part of the title section. Many of the Centre's donor agencies do provide clear instructions and guidance as to how their technical reports should be structured and labeled.

In general, grey literature is characterized as a resource that is not released through regular publishing channels (not controlled by "commercial publishers"), and that most often is not subject to formal "bibliographic control". Not only can it be difficult to identify and obtain grey literature, and it often contains country-specific information (Schöpfel & Farace, 2010; Rothstein & Hopewell, 2009). According to the so-called Prague definition grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body (Schöpfel, 2010). Actually, research information repositories contain collect, index, maintain, and share grey literature products, which could be used as a significant evaluation set for both bibliometric analyses as well as for investigating research trends and developments (Biagi, Saccone, Trufelli, & Puccinelli, 2012).

However, these technical reports do not receive the same priority for MARC cataloging as do other types of government documents (Byrne, 1998, p. 242). Both literature databases Web of Science (WoS) of Clarivate Analytics (formerly maintained by Thomson Reuters) and Elsevier's Scopus began to systematically index funding acknowledgements (FA) only in 2008. As an example, we used the following Query #1 to retrieve publications of authors affiliated with the World Agroforestry Centre from the Web of Science database:

Query #1: OG=(World Agrofor*) or OG=(icraf) or OG=(Int* C* Res* Agrofor*) or AD=(World Agrofor*) or AD=(icraf) and PY=(2008-2017)

Based on (Kokol & Vošner, 2018; Tang, Hu, & Liu, 2017), we used the following query to search in the *funding text* (FT) field (Query #1) as a filter for funding-related information. We limited our search to publications from authors affiliated with the World Agroforestry Centre published in the years 2008 through 2017:

Query #2: OG=(World Agrofor*) or OG=(icraf) or OG=(Int* C* Res* Agrofor*) or AD=(World Agrofor*) or AD=(icraf) and FT=(A* OR B* OR C* OR D* OR E* OR F* OR G* OR H* OR I* OR J* OR K* OR L* OR M* OR N* OR O* OR P* OR Q* OR R* OR S* OR T* OR U* OR V* OR W* OR X* OR Y* OR Z* OR 0* OR 1* OR 2* OR 3* OR 4* OR 5* OR 6* OR 7* OR 8* OR 9*) and PY=(2008-2017)

Query #3: OG=(World Agrofor*) or OG=(icraf) or OG=(Int* C* Res* Agrofor*) or AD=(World Agrofor*) or AD=(icraf) and FO=(A* OR B* OR C* OR D* OR E* OR F* OR G* OR H* OR I* OR J* OR K* OR L* OR M* OR N* OR O* OR P* OR Q* OR R* OR S* OR T* OR U* OR V* OR W* OR X* OR Y* OR Z* OR 0* OR 1* OR 2* OR 3* OR 4* OR 5* OR 6* OR 7* OR 8* OR 9*) and PY=(2008-2017)

For robustness check, we conducted another query in the *funding organization* (FO) field (Query #2), which produced the same result of 1434 records.

The records were manually cleaned up to exclude authors who are not affiliated with the World Agroforestry Centre, but whose organizations are based at or hosted by the Centre, thus sharing some of the same affiliation in the respective fields (name disambiguation). The results shown in Table 1, confirm that literature databases such as Web of Science provide an increasingly complete set of funding-related information.

[5] See also the *CERIF 1.5 Vocabulary*. Retrieved from https://www.eurocris.org/Uploads/Web%20pages/CERIF-1.5/CERIF1.5_Semantics.xhtml

Table 1: Funding acknowledgement (FA) index of World Agroforestry Centre publications

Year	Total records	Records with FA	FA presence rate (%)
2017	221	197	90
2016	236	202	86
2015	204	160	78
2014	228	176	77
2013	110	74	67
2012	95	77	81
2011	102	63	62
2010	62	46	74
2009	60	29	48
2008	76	20	26
Total	1394	1044	75

Note: Data source: Clarivate Analytics *Web of Science* (March 2018)

Using the example in (Weitz et al., 2016), we assessed how to enrich the funding-related information of MARC records. Adapting the procedural steps of (Grassano et al., 2017) in extracting funding-related information in (peer-reviewed) journal articles, we identified first the text of the publications with full funding acknowledgement. As shown in Table 2, almost 90% of the Centre's publications are of the document type (journal) article; the same figure applies to the number of articles with funding acknowledgement.

Table 2: Funding acknowledgement index of World Agroforestry Centre (journal) articles

Year	Total articles	Articles with FA	FA presence rate (%)
2017	199	180	90
2016	221	193	87
2015	181	152	84
2014	187	149	80
2013	95	67	71
2012	87	73	84
2011	80	58	73
2010	57	43	75
2009	49	27	55
2008	71	18	25
Total	1226	959	78

Note: Data source: Clarivate Analytics *Web of Science* (March 2018)

In a second step, we reviewed the publications for completeness as well as for additional funding-related text sections because some of the corresponding information can also be reported in a separate dedicated section, such as "Funding", "Financial Information" or "Financial Support".

The MARC 21 field 536 for the funding information note, the following funding-relevant subfields are available:

$a Text of note
$b Contract number
$c Grant number
$d Undifferentiated number
$e Program element number
$f Project number
$g Task number
$h Work unit number

Using a recent Centre publication as an example (Wigboldus et al., 2017), a funding-related entry in a MARC bibliographic record would look like this; note that the last entry "$d" for "Undifferentiated number" contains the internal grant ID number of the Centre for this project:

536 $aGreen Rubber project supported by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) on behalf of the Federal Government of Germany (BMZ)$f13.1432.7-001.00$dGTZG-1091

[INSERT SAMPLE RECORD OF MARC 536 FIELD ELEMENTS]

Web of Science has indexed a total of 22 journal articles with a funding acknowledgement with regard to the above *Green Rubber* project; one (1) article is indexed, but lacks the corresponding funding acknowledgement (Mertens, Bruun, Schmidt-Vogt, He, & de Neergaard, 2015). However, as shown in Table 3, the funding-related information in the corresponding journal articles is not consistent—for instance, none of the articles actually uses the full project name, "GREEN RUBBER: Alleviating poverty and enhancing environmental integrity through restoring ecosystem services in a tropical plantation crop in the Upper Mekong Region":

Table 3: Details of a funding acknowledgement of a project in publications (n = 22)

Item	Number	Presence rate (%)
Abbreviated project name	12	57
Project number	20	95
Full name of funding government agency (BMZ)	6	29
Full name of funding development agency (GIZ)	1	5
Sequence of abbreviated funding agency names #1: BMZ/GIZ	9	43
Sequence of abbreviated funding agency names #2: GIZ/BMZ	8	38

Note: Data source: Clarivate Analytics *Web of Science* (March 2018)

Based on a query of the technical donor reports of this project showed that a number of articles (D. Zhai et al., 2018; D.-L. Zhai, Xu, Dai, & Schmidt-Vogt, 2017), contain the relevant funding acknowledgement, have not been indexed in Web of Science. Yet, another article (Zhao et al., 2015), which is also not indexed in Web of Science, does not contain the relevant information at all. A non-indexed World Agroforestry Centre Working Paper, which can be considered grey literature, does not have the corresponding funding acknowledgement either (Hammond, Yi, McLellan, & Zhao, 2015).

Noticing the inconsistency in the presence of funding-related information as well as its presentation, utilizing text mining, in particular, named entity recognition, would help to overcome these issues (Weiss, Indurkhya, & Zhang, 2015). On the one hand, existing funding acknowledgements can be analyzed towards named entities to restructure the funding note text in the respective bibliographic records. On the other hand, grey literature can be harvested to retrieve missing funding-related information, which can then be incorporated into the corresponding bibliographic records. Adapting the framework for managing, processing, and analyzing grey literature and metadata of (Motta, Puccinelli, & Saccone, 2016), we are building a document corpus of technical donor reports and the relevant grant agreements to complement the bibliographic records of known project-related publications. In a subsequent phase, using these documents as our training and test corpus we want to enhance this process to capture publications and enrich their description where the funding information is missing.

[INSERT PROCEDURAL STEPS OF TEXT MINING]

Because of the heterogeneous nature of names of funding organizations, we manually built a dictionary of the corresponding names and aliases by adapting (Loster, Zuo, Naumann, Maspfuhl, & Thomas, 2017). This list was mapped to and verified with authority records such in VIAF[6] and other relevant sources such DBpedia with regards to funding agency names. In addition, the named entity recognition (NER) tagger model is being developed using the Stanford NLP, which creates a conditional random field (CRF) model (Sutton & McCallum, 2011), for which a document training set as well as the dictionary with the known

[6] Virtual International Authority File (VIAF). Retrieved from https://viaf.org/

named entities is needed. We applied the text processing features of the KNIME Analytics platform[7] to create a workflow for processing the documents.

[INSERT SCREENSHOT OF KNIME TEXT PROCESSING WORKFLOW]

We used the funding-related information as retrieved through the text mining process for editing exiting records by adding the corresponding funding information notes to the 538 fields. Similar to the process of upgrading a historical climate and weather data collection catalog (Mayernik, Huddle, Hou, & Phillips, 2018), we selected a subset of items to develop and test a protocol for upgrading the metadata in the records of the RIM. In the first phase, we selected publications with existing, but unstructured funding acknowledgement. We tested the approach by focusing on a single funding agency, the International Fund for Agricultural Development (IFAD), which supports many research activities of the Centre.

[INSERT TABLE OF RESULTS TAGGING IFAD]

At the moment, we rely on a combination of automated and manual steps of updating the records with the funding-related information. The automated part of the process involved using the conversion tool of the metadata editing suite MarcEdit[8] and the import features of Koha (Sanchez, Fatout, Howser, & Vance, 2006; Todd, 2018).

[INSERT TABLE OF PROCEDURAL STEPS FOR UPGRADING MARC RECORDS]

We yet have to investigate how to fully automate this process similar to the methodology for automating the matching of certain metadata fields with the corresponding controlled vocabulary as described in (Myntti & Cothran, 2013). The other manual process involved augmenting the enriched bibliographic records to include full descriptive funding-related information and utilize the relator notes to designate the relationship between an agent and a resource in a bibliographic record where applicable, that is in our case, "*funder [fnd]: A person or organization that furnished financial support for the production of the work.*"[9].

Conclusion

The combination of enriching existing bibliographic records with funding-related information through named entity recognition (NER) in the corresponding funding acknowledgement and extracting this information from grey literature/technical donor reports will allow for improved tracking of research outputs across projects in the future. Especially in an environment in which research-related information is maintained in separate systems and not in a single CRIS, maintaining funding-related information in the bibliographic record alongside with the other bibliographic information of the publications helps to fill this void. The RIM will also be enhanced by way of increasing access points for end-users through authorized forms of corporate name headings for the respective donor or funding agency (O'Neill, Bennett, & Kammerer, 2014). And, standardizing the corporate names assures that the data is consistent so that the RIM has the potential to better incorporate its scholarly records—in their broader conceptualization (Lavoie et al., 2014) —in a linked data environment (Myntti & Cothran, 2013) to help improve tracking funded research outputs. Overall, these enriched records could be used for performance assessments, but also as initial indicators for planning more targeted impact evaluations of research activities in the future. The enriched information can also help in identifying research trends over time, which can support the decision making process of research planning, especially when targeting certain funding agencies. Overall, this enterprise-focused NER approach allows the Centre to be in compliance with current movements towards more transparency and accessibility such Open Access as well as the International Aid Transparency Initiative (IATI) as it can better demonstrate how and where donor funds have been invested in development cooperation and what kinds of outcomes it was able to achieve. At the same time this could also be useful for automatic assessment of funding agency activities, sentiment analysis towards particular agencies, or discovering networks of funding agencies and grantees from textual data.

[7] KNIME Analytics Platform. Retrieved from https://www.knime.com/knime-analytics-platform
[8] MarcEdit. Retrieved from:
[9] Library of Congress, MARC Code List for Relators. Retrieved from https://www.loc.gov/marc/relators/

References

Aalberg, T., & Žumer, M. (2013). The value of MARC data, or, challenges of frbrisation. *Journal of Documentation*, *69*(6), 851–872. https://doi.org/10.1108/JD-05-2012-0053

Biagi, L. de, Saccone, M., Trufelli, L., & Puccinelli, R. (2012). Research product repositories: Strategies for data and metadata quality control. *Grey Journal*, *8*(2), 83–94.

Bryant, R., Clements, A., Feltes, C., Groenewegen, D., Huggard, S., Mercer, H., … Wright, J. (2017). *Research information management: Defining RIM and the library's role* (Research Report). Dublin, OH: OCLC Online Computer Library Center. Retrieved from https://doi.org/10.25333/C3NK88

Byrne, D. J. (1998). *MARC manual: Understanding and using MARC records* (2nd ed.). Westport, CT: Libraries Unlimited.

Costas, R., & van Leeuwen, T. N. (2012). Approaching the "reward triangle": General analysis of the presence of funding acknowledgments and "peer interactive communication" in scientific publications. *Journal of the American Society for Information Science and Technology*, *63*(8), 1647–1661. https://doi.org/10.1002/asi.22692

Costas, R., & Yegros-Yegros, A. (2013). Possibilities of funding acknowledgement analysis for the bibliometric study of research funding organizations: Case study of the Austrian Science Fund (FWF). In *Proceedings* (Vol. 2, pp. 1401–1408). Vienna: AIT Austrian Institute of Technology. Retrieved from http://www.issi2013.org/Images/ISSI_Proceedings_Volume_II.pdf

de Castro, P., Shearer, K., & Summann, F. (2014). The gradual merging of repository and CRIS solutions to meet institutional research information management requirements. *Procedia Computer Science*, *33*, 39–46.

Freyman, C. A., Byrnes, J. J., & Alexander, J. (2016). Machine-learning-based classification of research grant award records. *Research Evaluation*, *25*(4), 442–450. https://doi.org/10.1093/resevaljrvw016

Grassano, N., Rotolo, D., Hutton, J., Lang, F., & Hopkins, M. M. (2017). Funding data from publication acknowledgments: Coverage, uses, and limitations. *Journal of the Association for Information Science and Technology*, *68*(4), 999–1017. https://doi.org/10.1002/asi.23737

Hammond, J., Yi, Z., McLellan, T., & Zhao, J. (2015). *Situational analysis report: Xishuangbanna autonomous Dai prefecture Yunnan, China* (Working Paper No. 194). Kunming, CN: World Agroforestry Centre. Retrieved from http://humidtropics.cgiar.org/wp-content/uploads/downloads/2015/04/Situational-Analysis-Report-Banna.pdf

Ihli, M. (2017). Methods for synthesis of funding agency and publisher data. In *Proceedings of the 6th International Workshop on Mining Scientific Publications* (pp. 57–63). Toronto, ON, Canada. https://doi.org/10.1145/3127526.3127537

Ivanović, D., Milosavljević, G., Milosavljević, B., & Surla, D. (2010). A CERIF-compatible research management system based on the MARC 21 format. *Program*, *44*(3), 229–251. https://doi.org/10.1108/00330331011064249

Ivanović, D., Surla, D., & Konjović, Z. (2011). CERIF compatible data model based on MARC 21 format. *The Electronic Library*, *29*(1), 52–70. https://doi.org/10.1108/02640471111111433

Ivanović, D., Surla, D., & Racković, M. (2011). A CERIF data model extension for evaluation and quantitative expression of scientific research results. *Scientometrics*, *86*, 155–172. https://doi.org/10.1007/s11192-010-0228-2

Jeffery, K., & Asserson, A. (2009). Institutional Repositories and Current Research Information Systems. *New Review of Information Networking*, *14*(2), 71–83. https://doi.org/10.1080/13614570903359357

Jörg, B., Jeffery, K., Asserson, A., van Grootel, G., Rasmussen, H., Dvorak, J., … Simons, E. J. (Eds.). (2010). *CERIF 2008 - 1.2 Semantics*. euroCRIS. Retrieved from https://www.eurocris.org/Uploads/Web%20pages/CERIF2008/Release_1.2/CERIF2008_1.2_Semantics.pdf

Kokol, P., & Vošner, H. B. (2018). Discrepancies among Scopus, Web of Science, and PubMed coverage of funding information in medical journal articles. *Journal of the Medical Library Association*, *106*(1), 81–86. https://doi.org/10.5195/jmla.2018.181

Lavoie, B., Childress, E., Erway, R., Faniel, I., Malpas, C., Schaffner, J., & Van der Werf, T. (2014). *The evolving scholarly record*. Dublin, OH: OCLC Online Computer Library Center. Retrieved from http://www.oclc.org/research/publications/library/2014/oclcresearch-evolvingscholarly- record-2014.pdf

Library of Congress. (2017). *MARC 21 format for bibliographic data* (1999 Edition). Washington, D.C.: Network Development and MARC Standards Office, Library of Congress. Retrieved from https://www.loc.gov/marc/bibliographic/

Loster, M., Zuo, Z., Naumann, F., Maspfuhl, O., & Thomas, D. (2017). Improving company recognition from unstructured text by using dictionaries. In *Proceedings 20th International Conference on Extending Database Technology (EDBT)* (pp. 610–619).

Mayernik, M. S., Huddle, J., Hou, C.-Y., & Phillips, J. (2018). Modernizing library metadata for historical weather and climate data collections. *Journal of Library Metadata*, *0*(0), 1–21. https://doi.org/10.1080/19386389.2018.1440927

Mertens, C. F., Bruun, T. B., Schmidt-Vogt, D., He, J., & de Neergaard, A. (2015). Rural wood consumption patterns of local and immigrant households with differentiated access to resources in Xishuangbanna, Yunnan, China. *Energy Policy*, *80*, 112–121. https://doi.org/10.1016/j.enpol.2015.01.040

Motta, G., Puccinelli, R., & Saccone, M. (2016). Extracting value from grey literature: processes and technologies for aggregating and analyzing the hidden "big data" treasure of organizations. *The Grey Journal*, *12*(1), 15–25.

Myntti, J., & Cothran, N. (2013). Authority control in a digital repository: Preparing for linked data. *Journal of Library Metadata*, *13*, 95–113. https://doi.org/10.1080/19386389.2013.826061

National Information Standards Organization. (2010). *Scientific and technical reports - Preparation, presentation, and preservation* (ANSI/NISO Z39.18-2005 (R2010)). Baltimore, MD: National Information Standards Organization.

O'Neill, E. T., Bennett, R., & Kammerer, K. (2014). Using authorities to improve subject searches. *Cataloging & Classification Quarterly*, *52*, 6–19. https://doi.org/10.1080/01639374.2013.850018

Pappas, C., & Williams, I. (2011). Grey literature: Its emerging importance. *Journal of Hospital Librarianship*, *11*(3), 228–234. https://doi.org/10.1080/15323269.2011.587100

Rigby, J. (2011). Systematic grant and funding body acknowledgement data for publications: new dimensions and new controversies for research policy and evaluation. *Research Evaluation*, *20*(5), 365–375. https://doi.org/10.3152/095820211X13164389670392

Rothstein, H. R., & Hopewell, S. (2009). Grey literature. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 103–126). New York, NY: Russell Sage Foundation.

Sanchez, E., Fatout, L., Howser, A., & Vance, C. (2006). Cleanup of NetLibrary cataloging records: A methodical front-end process. *Technical Services Quarterly*, *23*(4), 51–71. https://doi.org/10.1300/J124v23n04_04

Schöpfel, J. (2010). Towards a Prague definition of grey literature. In D. Farace & J. Frantsen (Eds.), *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues* (pp. 11–26). Prague, Czech Republic: GreyNet - Grey Literature Network Service. Retrieved from https://archivesic.ccsd.cnrs.fr/sic_00581570

Schöpfel, J., & Farace, D. (2010). Grey literature. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of library and information sciences* (3rd ed., pp. 2029–2039). Boca Raton, FL: CRC Press.

Sutton, C., & McCallum, A. (2011). An introduction to conditional random fields. *Foundations and Trends in Machine Learning, 4,* 267-373. https://doi.org/10.1561/2200000013

Tang, L., Hu, G., & Liu, W. (2017). Funding acknowledgment analysis: Queries and caveats. *Journal of the Association for Information Science and Technology*, *68*(3), 790–794. https://doi.org/10.1002/asi.23713

Todd, C. R. (2018). Librarian as data migrator: a functional pathway from Millennium to Koha. *Digital Library Perspectives*, *34*(1), 60–69. https://doi.org/10.1108/DLP-09-2017-0035

Weiss, S. M., Indurkhya, N., & Zhang, T. (2015). *Fundamentals of predictive text mining* (2nd ed.). London: Springer.

Weitz, J., Toves, J., Vizine-goetz, D., Naught, N., & Bremer, R. (2016). Mining MARC's Hidden Treasures: Initial Investigations Into How Notes of the Past Might Shape Our Future. *Journal of Library Metadata*, *16*(3–4), 166–180. https://doi.org/10.1080/19386389.2016.1262653

Wigboldus, S., Hammond, J., Xu, J., Yi, Z.-F., He, J., Klerkx, L., & Leeuwis, C. (2017). Scaling green rubber cultivation in Southwest China—An integrative analysis of stakeholder perspectives. *Science of The Total Environment*, *580*, 1475–1482. https://doi.org/10.1016/j.scitotenv.2016.12.126

Zhai, D., Dong, J., Cadisch, G., Wang, M., Kou, W., Xu, J., … Abbas, S. (2018). Comparison of pixel- and object-based approaches in phenology-based rubber plantation mapping in fragmented landscapes. *Remote Sensing*, *10*. https://doi.org/10.3390/rs10010044

Zhai, D.-L., Xu, J.-C., Dai, Z.-C., & Schmidt-Vogt, D. (2017, June 25). Lost in transition: Forest transition and natural forest loss in tropical China. https://doi.org/10.1016/j.pld.2017.05.005

Zhao, M., Nalaka, G., Harrison, R. D., Xu, J., Khin, M. M., Ridwan, N. D., & Ekananda, P. (2015). Diversity and vertical distribution characteristics of vascular epiphytes in Bulong Nature Reserve. *Plant Diversity and Resources*, *37*(3), 327–338. https://doi.org/10.7677/ynzwzj201514110