



národní
úložiště
šedé
literatury

Další využití šedé literatury - propojení výsledků výzkumu a plných textů publikací

Farace, Dominic
2011

Dostupný z <http://www.nusl.cz/ntk/nusl-82067>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 30.11.2022

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

REPURPOSING GREY LITERATURE - LINKING RESEARCH DATA TO FULL-TEXT PUBLICATIONS: SOME PRELIMINARY RESULTS

DOMINIC FARACE

Dominic.farace@textrelease.com

GreyNet

Abstract

This presentation deals with the research design and preliminary results of EPP, an Enhanced Publications Project carried out among a pool of over 280 authors in the International Conference Series on Grey Literature. In a way, this project seeks to circumvent the data vs. documents camp in the grey literature community by way of a middle ground provided through enhanced publications. Enhanced publications allow for a fuller understanding of the process in which data and information are used and applied in the generation of knowledge. The enhanced publication of grey literature precludes the idea of a random selection of data and information, and instead focuses on the human intervention in data-rich environments. Enhanced publications inherently contribute to the review process of grey literature as well as the replication of research and improved visibility of research results in the scholarly communication chain.

Keywords

Grey Literature, research, enhanced publications, preliminary results

Introduction - A research in progress

The international grey literature community is both diverse in scale and foci of research. The HEP (High Energy Physics) community depicted at CERN in Geneva and the Karst community monitored at Florida State University are two such examples. In fact, GreyNet's own community of researchers in the field library and information sciences is yet another example – one in which my presentation is focused.

The main title of this presentation 'Repurposing Grey Literature' includes both a retrospective element as well as the element of enhancing already existing records. GreyNet has over the years been involved in providing retrospective access to its collections of conference preprints, which first started with bibliographic records in the SIGLE database, later their full-text records in the OpenSIGLE repository, and now the newly launched OpenGrey Repository. The element however of enhancing these metadata records with related research

data as well as post-publication data such as commentaries brings the use of the term 'repurposing grey literature' to a fuller understanding.

The definition of an enhanced publication is borrowed from the DRIVER-II project, "a publication that is enhanced with three categories of information: 1) research data, 2) extra materials, and 3) post-publication data". In EPP, four partnering organizations were brought together based on their expertise and tasks they would execute during the course of the project. GreyNet together with INIST-CNRS would devise a questionnaire and carry-out a survey among its author base in the acquisition of research data linked to conference preprints. GreyNet would facilitate data entry in the DANS Easy Repository with link backs to corresponding metadata records in the OpenGrey Repository; and would further cooperate with Pratt Institute to establish basic criteria upon which commentaries by LIS students would be compiled and added to existing metadata records.

A Phased Approach

EPP, Enhanced Publications Project, is being carried out in six phases, some of which overlap depending on the partnering organizations responsible for their implementation. The phases include:

Project Proposal and Formation of the Team

Design of the Questionnaire and Author Survey

Criteria for the commentaries and selection of eligible conference preprints

Acquisition and Submission of research data

Data upload and cross-linking

Draft of enhanced publication guidelines and the design of a future workflow

As indicated in the subtitle of the paper, this presentation is based on preliminary results limited to the survey and questionnaire. The final results of the six phases in EPP are scheduled for presentation during the Thirteenth International Conference on Grey Literature this December in Washington DC. However, I consider the opportunity to present these partial results to this national forum as extremely beneficial to the project – in anticipation of your comments and recommendations.

Survey and Questionnaire

The population of the survey was selected from among the 286 authors and co-authors in the GL-Conference Series. It was decided that only first authors would receive the questionnaire, which narrowed the potential population of the survey to 162 of which only 95 were actually sent the online questionnaire. The reason the other 67 first authors were not included in the final survey population was due to a number factors such as no current email address, retired, deceased, etc.

Number (co)authors in the GL-Series	of	Number authors in the GL-Series	of	first	Number Survey Recipients	of	EPP	Number Survey Respondents	of	EPP
286		162			95			37		

The survey was carried out using freeware ‘Survey Monkey’ and the questionnaire contained 10 items, 3 of which were open-ended. The results used in this presentation are based on 37 respondents, roughly a 40% response rate.

While maintaining the anonymity of the individual respondent it was possible to determine the geographic region from which they came as seen in the figure above. Authors from 14 countries responded to the questionnaire.

Some Preliminary Results
 In an effort to better inform and guide the respondents through the questionnaire, an introductory text was included in a personalized email inviting them to participate in the survey. Also, informative subheadings were inserted at intervals between the items on the questionnaire as shown in quotation marks.

“Data exchange is becoming the norm in open access communities”

Question 1: Does one or more of your conference papers in the GL-Series base its findings on empirical or statistical data?

While it is generally known that not all of the conference papers in the GL-Series are based on statistical data, it still has yet to be determined what percentage of the collection is. The fact however that 65% of the respondents’ state that their work relies on empirical research data provides a clear indication of the relevance of this for the project.

Question 2: If so, would these data and/or datasets still be available in part or whole for archiving purposes?

The fact that some 60% of the respondents maintain and store their research data is no doubt indicative of the value they place on their work.

“A data policy should be in place within research communities and organizations”

Question 3: Are you aware of any existing data archives or data initiatives in your country related to grey literature or other scientific publications?

If the above percentages are indicative, then half of the respondents appear unaware of data initiatives related to grey literature in their country. Unfortunately, we do not yet have cross tabulations by country and we can only refer back to the first figure in which the respondents were grouped by geographic region. The incentive to further explore this finding takes on the implication that EPP could contribute to increased awareness of this within the grey literature community.

Question 4: If so, please provide the name(s) and corresponding URL(s) here?

The responses to this open-ended question were categorized into one of three clusters depending on how specific, general, or non-applicable the responses were. It can also be noted that the number of respondents to this question was significantly less than to the other questions. However, if we look at the 70% of those who provided specific names and/or URLs, then we are able to compile a short list of archives housing research data on grey literature: PLEIADI (IT), NUSL (CZ), NOAA (US), NASA (US), NSIDC (US), ICPSR (US), METIS (NL), NSF (US), NIH (US), ADP (SI), ORNL (US), TRAIL/CRL (US), SYNABA (PL), IATUL (INT), IOP (INT), Morphbank/FSU (US), ISS (IT), DANS (NL), SIDR (FR). These above cited data archives/portals still have to be further examined for their relevance to our project.

“Data counts as science output and should be recognized in references and citations”

Question 5: Would you be willing to submit data, datasets, or subsets to DANS (Data Archiving and Networked Services) that would in turn be linked to their existing metadata records in the OpenGrey Repository?

Almost half of the respondents (48%) appear willing to submit their data to the DANS Archive, while 46% express uncertainty. The results from this question proves to be one of the more challenging aspects of the project for it will be necessary to address issues that authors have regarding their hesitance to submit data. Perhaps, by underscoring the advantages authors would gain by increased referencing and citing of their work and by enabling them through ready guidelines, this group of respondents would be more willing to participate in the project.

Question 6: If so, would you prefer that GreyNet entered your retrospective data and/or datasets in DANS, or would you prefer to do this directly?

The ‘no preference’ response to this question (41%) can be interpreted not only as encouraging for retrospective data but also for future data and datasets that would be added to

the DANS Repository and subsequently linked to the OpenGrey full-text records. While the scope of this project is geared to retrospective input, it is understood that empirical and statistical data related to future GreyNet collections of conference preprints will be directly entered in the DANS Repository by the authors.

“Research data should be preserved and accessible in order to enhance scholarly communication”

Question 7: Do you agree that both the data producer as well as the data user stand to benefit by submitting data, datasets, or subsets for this Enhanced Publications Project?

Over 90% of the respondents agree that both the researcher/author as well as the potential data user would stand to gain from the enhancement of conference preprints with related data, datasets, and subsets. It is long understood that researchers and authors are at the same time ardent information users. It has also been demonstrated that specific communities of researchers are more likely to use the sources and resources produced within their own community. Now, we see that wider audiences (net users) will have open access to this research.

Question 8: Do you think that guidelines for data entry should be available for future conference papers and other types of grey literature?

The most significant percentage of respondents (94%) is of the opinion that guidelines should be made available for data entry, and such guidelines will be addressed in the final phase of this Enhanced Publications Project.

“Data is disciplinary or subject based and this accounts for differences in formats used to acquire it - Examples may range from spreadsheets to SPSS, etc.”

Question 9: What kind of data and data formats have you used/are using in your research?

The responses to this open-ended question were categorized into one of three clusters depending on how specific, general, or non-applicable the responses were. If for example a respondent replied that they were ‘no longer engaged in research’ or were ‘not engaged in research at the moment’ these responses were categorized as non-applicable. If the respondents replied to this question with non-specific or general terms such as tables, charts, times series, etc., these responses were categorized as general. And, those respondents who actually named particular software or statistical packages were then labeled specific responses. Among those named were SPSS, Meko, Excel, Minitab, MS Access, etc.

Question 10: Please enter your name, email address, and any other comments or recommendations for this Enhanced Publications Project?

The respondents to this open-ended question were classified into three groups based on their prior affiliation in the GL-Conference Series. Preliminary results show that a little over 40% were involved over the past year, about 20% were involved 2 to 4 years ago, and a little less than 40% were involved five or more years ago. While we might have anticipated that the majority of respondents would have been more recent to the GL-Conference Series, it was quite interesting to find that twice as many authors who contributed 5 or more years ago to the conference series chose to respond to this survey than authors who were distanced 2-4 years from the conference series.

In Close

While the summary and conclusion of this project await completion, a number of remarks by authors and researchers may lend to open discussion and debate among those gather here today. To this end, I have paraphrased a few remarks taken from the completed EPP questionnaires and state them as follows: “I’m a firm believer that not all data is worth archiving”; “Will your system support the preservation and migration to new platforms?”; “For many, it would need to be a local activity linked to our own sites”; I share data with my colleagues and research teams, but I’m not sure if I would be willing to share them with anybody else at the moment?”.

References

<http://greynet.org/opengreyrepository.html>

<https://easy.dans.knaw.nl/ui/home>

<http://www.datacite.org/>

The driving and evolving role of grey literature in High-Energy Physics / Anne Gentil-Beccot
<http://www.reference-global.com/doi/abs/10.1515/9783598441493.2.155>

Grey Literature in Karst Research: The Evolution of the Karst Information Portal, KIP / Chavez
<http://www.reference-global.com/doi/abs/10.1515/9783598441493.2.181>

<http://www.driver-repository.eu>

Enhanced Publications: State-of-the-Art (PDF)
Enhanced Publications: Object Models and Functionalities (PDF)

http://www.textrelease.com/images/Whois_in_Grey_Literature_2011.pdf