Third Workshop on Building towards Information Interaction and Retrieval Resources Re-use (BIIRRR 2022)

Toine Bogers
Aalborg University Copenhagen
Denmark
toine@hum.aau.dk

Maria Gäde
Humboldt-Universität zu Berlin
Germany
maria.gaede@ibi.hu-berlin.de

Mark Hall
The Open University
Milton Keynes, United Kingdom
mark.hall@open.ac.uk

Marijn Koolen
Huygens ING, KNAW
Netherlands
marijn.koolen@di.huc.knaw.nl

Vivien Petras
Humboldt-Universität zu Berlin
Germany
vivien.petras@ibi.hu-berlin.de

Paul Thomas
Microsoft
Australia
pathom@microsoft.com

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in visualization**; • **Information systems** → **Users and interactive retrieval**.

KEYWORDS

Interactive IR, IIR, research data, research design, research software, re-use, data documentation, data sharing practices

1 FORMAT

Full-day workshop

2 BACKGROUND AND RELEVANCE TO CHIIR

The systems-focused Information Retrieval (IR) community has a strong tradition of making research resources available, as exemplified by the test collections and shared tasks developed in large-scale initiatives such as TREC, CLEF, NTCIR, and FIRE. These efforts have significant benefits for the IR community, in particular enabling the re-use of (parts of) the test collections and the shared tasks in other IR research. Efforts have been undertaken to transfer this paradigm to less systems-focused research, including the TREC Interactive and Session tracks [12], the INEX Interactive track [14, 16], the Interactive Social Book Search track [6, 7, 11], and the RePAST archive [5]. However, the high degree of variation in user-focused research has meant that none of these achieved similar degrees of standardisation and re-use.

At the 2018 and 2019 BIIRRR workshops [3, 4], we identified a number of focus areas essential to resource re-use in interactive information retrieval. We distilled these into a manifesto [8] published at CHIIR 2021, which highlighted the potential and challenges of research documentation and archiving for future use. We proposed eight principles for improving the re-use of resources that are applicable to the wider information interaction and retrieval (II&R) community and presented steps for moving towards increasing documentation and re-use. We now propose a third edition of the BIIRRR workshop with a focus on further discussing the challenges for re-use and developing recommendations for documenting the wide range of research designs that CHIIR encompasses.

One of the central aspects of the CHIIR community is that it brings together a wide range of research methodologies, ranging from desk-based to qualitative and quantitative studies, often incorporating both user and system aspects. In order to fully reap the benefits of this breadth and the cross-fertilisation possibilities it offers, the amount of re-use between studies needs to be increased. Awarding the CHIIR 2021 best paper award to our manifesto shows that there is a desire within the community to move forward on these issues.

It is important to note that speaking about re-use in the context of CHIIR, we are not focused on reproducibility as a necessary end goal, although we do not exclude it. Any part of a study (research design, research data, research infrastructure) is valuable for re-use. For example, a study looking at leisure search might re-use an interface from another study, but use a different set of tasks for their participants. The element to re-use can even be as small as a single question from a qualitative study.

One major barrier to re-use, that our initial survey revealed [4], was missing documentation or documentation lacking in sufficient detail to enable (the decision what to) re-use. This is where the heterogeneity of approaches creates additional challenges. Documentation for qualitative methods may require a different format than for quantitative methods, but both can be described in a structured and standardized way. Such guidelines for reporting qualitative [17] and quantitative research [9, 15] already exist in the medical community and are being developed in the psychology community [1]. This workshop provides a forum where researchers from the full spectrum of CHIIR interests can come together to develop ideas on how we can improve research documentation, to ensure that

the documentation standards are relevant and applicable to both the specifics and breadth of the CHIIR community.

In this context, it is particularly important that any guidelines or standards are flexible, adaptable, and can evolve as the research in the community evolves. The openness of the II&R community must not be restricted and extensive discussions are needed on how to balance standardisation, space for innovation and for aspects that can only be documented to a limited extent. The workshop will provide the opportunity to discuss the challenges for documenting research resources and developing documentation guidelines for the II&R community.

3 KEY OUTCOMES & CONTINUING ACTIVITIES

After focusing on identifying the barriers towards resource re-use in the first two editions, the goal of the 3rd edition of the BIIRRR workshop is to provide an interactive forum to address questions around documentation and re-use of II&R materials as a step further towards the development of community standards. We are also explicitly widening the scope to include any kind of research that sits under the banner of Information Interaction and Retrieval, to ensure that the workshop's outcomes are as widely usable within the CHIIR community as possible.

In the manifesto we stressed the importance of documentation standards, which we consider the most viable first step towards reaching a community consensus on how to move forward with the other elements mentioned in the manifesto. One of the main outcomes of BIIRRR 2022 will therefore be a set of requirements on documenting *research designs* that will be developed in the breakout groups, drawing on the presentations of *experience* and *position* papers and an ongoing analysis of II&R studies. This outcome is similar in spirit to the standardized format for exchanging user study instruments that came out of the previous edition of BIIRRR [10]. That work demonstrated a way to standardize, while also allowing for flexible extensions of the standard.

Towards this end, the workshop has two main goals:

- **Collect practical experiences with documentation and re-use** Similar to the format we introduced at BIIRRR 2019, we wish to solicit more experience and position papers to collect more examples of practical barriers, challenges, and approaches to operationalizing our eight principles, this year with a particular focus on the challenges around documenting complex study setups. In an experience paper (and its subsequent presentation at the workshop), authors can further document one of their previously published II&R studies and analyze it in terms of the possibilities and challenges for the re-use of their research data, research designs and research infrastructure. Position papers discuss challenges and proposals in a more abstract manner. Both will push the discussion forward and solicit insights and pitfalls and challenges for improving documentation.
- **Develop recommendations for research design documentation** We argue in our manifesto that research designs are the first step in documenting research resources - research data and research infrastructure are even more complicated and will be worked on in later editions. A survey

conducted as part of the BIIRRR 2018 workshop [4] indicated that the lack of documentation was a major limiting factor, but also that how to appropriately document was something researchers struggled with. A 2019 review of the IIX and CHIIR experimental papers [13] showed common characteristics of research designs that could be used to develop recommendations on how to document in a more structured way. The second goal of this year's workshop is thus to start the process of discussing and developing recommendations for research design documentation. The explicit aim is to develop an open proposal that can be used across the full range of study types in the II&R community and that is open to further development.

To achieve these outcomes, wide dissemination to the community is a key activity, in particular to maintain momentum directly after the workshop. We propose to provide more dissemination activities on the workshop website directly after its completion. A short summary produced and discussed at the end of the workshop will be published within a few days of the workshop. An overview paper and the experience and position papers will also be published open access at CEUR-WS as was done with the BIIRRR 2019 workshop papers [2]. Long-term we are also aiming for an article in an appropriate journal to articulate our vision, report the results of an ongoing analysis of II&R re-use practices, and detail the knowledge gained from the workshop outcomes and the concrete, practical frameworks developed based on the BIIRRR workshops' results and community feedback.

Participants will also be invited to commit to participating in or leading some of the continuing activities, such as working on documentation guidelines for different II&R research designs. The aim of this is to encourage wider community participation in this process, but also to enable participants to organically form into interest groups with a view towards collaboration or even potential grant applications in their area of interest, a welcome outcome from the previous workshops.

The central aspect of this workshop is that it will enable follow-up events to have a solid foundation from which to continue, rather than having to re-invent the wheel each time.

4 WORKSHOP ORGANIZERS

Five of the organizers have jointly organized several successful CHIIR-related workshops: Supporting Complex Search Tasks (SCST) at ECIR in 2015 and CHIIR in 2017, and Barriers to Interactive Information Retrieval Resource Reuse (BIIRRR) at CHIIR in 2018 and 2019. These five authors have co-organised the interactive tracks at the Social Book Search Lab at CLEF in 2014−2016. For this workshop they are joined by Paul Thomas, to emphasise that the workshop's outcomes must be relevant to industry research as well.

Toine Bogers (http://www.toinebogers.com) is an associate professor at the Department of Communication & Psychology at Aalborg University Copenhagen in Copenhagen, Denmark.

Maria Gäde (https://www.ibi.hu-berlin.de/de/ueber-uns/personen/gaede) is a senior lecturer at the Berlin School of Library and Information Science at Humboldt Universität zu Berlin, Germany.

Mark Hall (https://www.room3b.eu/) is a lecturer in Computing & Communications at The Open University in the UK.

Marijn Koolen (https://marijnkoolen.com) is a researcher of information science and digital humanities at Huygens ING, KNAW in Amsterdam, the Netherlands.

Vivien Petras (https://www.ibi.hu-berlin.de/en/about-us/people/petras) is a professor at the Berlin School of Library and Information Science at Humboldt Universität zu Berlin, Germany.

Paul Thomas (https://www.microsoft.com/en-us/research/people/pathom/) is a senior applied scientist at Microsoft, based in Australia.

REFERENCES

[1] Katarina Blask, Lea Gerhards, and Maria Jalynskij. 2021. PsyCuraDat: Designing a User-Oriented Curation Standard for Behavioral Psychological Research Data. *Frontiers in Psychology* 11 (2021), 3892. https://doi.org/10.3389/fpsyg.2020.579397

[2] Toine Bogers, Samuel Dodson, Luanne Freund, Maria Gäde, Mark Hall, Marijn Koolen, Vivien Petras, Nils Pharo, and Mette Skov (Eds.). 2019. *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2019) co-located with the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2019)*. Number 2337 in CEUR Workshop Proceedings. Aachen. http://ceur-ws.org/Vol-2337

[3] Toine Bogers, Samuel Dodson, Luanne Freund, Maria Gäde, Mark Hall, Marijn Koolen, Vivien Petras, Nils Pharo, and Mette Skov. 2019. Workshop on barriers to interactive IR resources re-use (BIIRRR 2019). In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. ACM, 389–392.

[4] Toine Bogers, Maria Gäde, Mark Hall, Luanne Freund, Marijn Koolen, Vivien Petras, and Mette Skov. 2018. Report on the Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2018). *SIGIR Forum* 52, 1 (Aug. 2018), 119–128.

[5] Luanne Freund and Barbara M. Wildemuth. 2014. Documenting and studying the use of assigned search tasks: RepAST. *Proceedings of the American Society for Information Science and Technology* 51, 1 (2014), 1–4.

[6] Maria Gäde, Mark Michael Hall, Hugo C. Huurdeman, Jaap Kamps, Marijn Koolen, Mette Skov, Toine Bogers, and David Walsh. 2016. Overview of the SBS 2016 Interactive Track. In *Working Notes of the CLEF 2016 Conference (CEUR Workshop Proceedings)*, Krisztian Balog, Linda Cappellato, Nicola Ferro, and Craig Macdonald (Eds.), Vol. 1609. CEUR-WS.org, 1024–1038.

[7] Maria Gäde, Mark Michael Hall, Hugo C. Huurdeman, Jaap Kamps, Marijn Koolen, Mette Skov, Elaine Toms, and David Walsh. 2015. Overview of the SBS 2015 Interactive Track. In *Working Notes of the CLEF 2015 Conference (CEUR Workshop Proceedings)*, Linda Cappellato, Nicola Ferro, Gareth J. F. Jones, and Eric SanJuan (Eds.), Vol. 1391. CEUR-WS.org.

[8] Maria Gäde, Marijn Koolen, Mark Hall, Toine Bogers, and Vivien Petras. 2021. A Manifesto on Resource Re-Use in Interactive Information Retrieval. In *CHIIR '21: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia). Association for Computing Machinery, New York, NY, USA, 141–149. https://doi.org/10.1145/3406522.3446056

[9] Jeremy Grimshaw. 2014. SURGE (The Survey Reporting Guideline). *Guidelines for Reporting Health Research: A User's Manual* (2014), 206–213.

[10] Mark Michael Hall and Toine Bogers. 2020. A Standardised Format for Exchanging User Study Instruments. In *CHIIR '20: Conference on Human Information Interaction and Retrieval, Vancouver, BC, Canada, March 14-18, 2020*, Heather L. O'Brien, Luanne Freund, Ioannis Arapakis, Orland Hoeber, and Irene Lopatovska (Eds.). ACM, 457–461.

[11] Mark Michael Hall, Hugo C. Huurdeman, Marijn Koolen, Mette Skov, and David Walsh. 2014. Overview of the INEX 2014 Interactive Social Book Search Track. In *Working Notes of the CLEF 2014 Conference (CEUR Workshop Proceedings)*, Linda Cappellato, Nicola Ferro, Martin Halvey, and Wessel Kraaij (Eds.), Vol. 1180. CEUR-WS.org, 480–493.

[12] Paul Over. 2001. The TREC Interactive Track: An Annotated Bibliography. *Information Processing & Management* 37, 3 (2001), 369–381.

[13] Vivien Petras, Toine Bogers, and Maria Gaede. 2019. Elements of IIR Studies: A Review of the 2006-2018 IIiX and CHIIR Conferences. In *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2019)*. CEUR-WS.org, Aachen, Germany, 37–41.

[14] Nils Pharo, Thomas Beckers, Ragnar Nordlie, and Norbert Fuhr. 2011. Overview of the INEX 2010 Interactive Track. In *INEX '10: Proceedings of the Ninth International Workshop of the Initiative for the Evaluation of XML Retrieval*, Shlomo Geva, Jaap Kamps, Ralf Schenkel, and Andrew Trotman (Eds.). Springer, Berlin, Heidelberg, 227–235.

[15] Akash Sharma, Nguyen Tran Minh Duc, Tai Luu Lam Thang, Nguyen Hai Nam, Sze Jia Ng, Kirellos Said Abbas, Nguyen Tien Huy, Ana Marušić, Christine L Paul, Janette Kwok, et al. 2021. A Consensus-based Checklist for Reporting of Survey Studies (CROSS). *Journal of General Internal Medicine* (2021), 1–9.

[16] Anastasios Tombros, Birger Larsen, and Saadia Malik. 2005. The Interactive Track at INEX 2004. In *INEX '04: Proceedings of the Third International Workshop of the Initiative for the Evaluation of XML Retrieval*, Norbert Fuhr, Mounia Lalmas, Saadia Malik, and Zoltán Szlávik (Eds.). Springer, Berlin, Heidelberg, 410–423.

[17] Allison Tong, Peter Sainsbury, and Jonathan Craig. 2007. Consolidated Criteria for Reporting Qualitative Research (COREQ): A 32-item Checklist for Interviews and Focus Groups. *International Journal for Quality in Health Care* 19, 6 (2007), 349–357.