

Overview of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2019)

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

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

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

Samuel Dodson
University of British Columbia
Canada
dodsons@mail.ubc.ca

Mark Hall
Martin-Luther-Universität
Halle-Wittenberg
Germany
mark.hall@informatik.uni-halle.de

Nils Pharo
Oslo Metropolitan University
Norway
nils.pharo@oslomet.no

Luanne Freund
University of British Columbia
Canada
luanne.freund@ubc.ca

Marijn Koolen
Royal Netherlands Academy of Arts
and Sciences
Netherlands
marijn.koolen@di.huc.knaw.nl

Mette Skov
Aalborg University
Denmark
skov@hum.aau.dk

ABSTRACT

This paper presents an overview of the BIIRRR 2019 workshop at CHIIR 2019, which had the explicit aim of understanding and promoting re-use of resources for interactive IR experimentation.

CCS CONCEPTS

• **Information systems** → **Users and interactive retrieval**; • **General and reference** → *Empirical studies*; *Evaluation*;

KEYWORDS

IIR, re-use, secondary use, research design

1 DESCRIPTION

What would be the equivalent of a re-usable TREC test collection for the interactive information retrieval (IIR) community? The goal of the BIIRRR 2019 workshop was to answer this question by continuing existing community-driven efforts to develop approaches for the collection, organization, maintenance, and sharing of resources for IIR experimentation. These efforts developed out of discussions at the CHIIR 2017 workshop on Supporting Complex Search Tasks (SCST 2017) [1] and were expanded upon during the BIIRRR workshop at CHIIR 2018 [3, 4].

The information retrieval (IR) community has a strong tradition of making research data of system-based experimentation available for re-use, as exemplified by the development of test collections, shared tasks and relevance assessments in large-scale initiatives such as TREC¹, CLEF², NTCIR³, and FIRE⁴. These efforts have had

¹<https://trec.nist.gov/>

²<http://www.clef-initiative.eu/>

³<http://research.nii.ac.jp/ntcir/>

⁴<http://fire.irsi.res.in>

Workshop on Barriers to Interactive IR Resources Re-use at the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2019), 14 March 2019, Glasgow, UK 2019. Copyright for the individual papers remains with the authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

significant benefits for the IR community, in particular enabling the re-use of aspects of the test collections or shared tasks in other IR research. Comparable efforts have been undertaken to bring this paradigm to IIR research, such as the TREC Interactive [22] and Session [6] tracks, the INEX Interactive track [21], and the Interactive Social Book Search track [24]. The high degree of variation between IIR studies, however, has meant that none of these have achieved similar degrees of standardization and re-use. It seems that the traditional shared task structure and test collections (i.e. resources to search on) are not successful in inducing re-use in IIR. Instead, an equivalent type and level of re-use is more likely to be achieved through increased sharing of research design (aspects), enabling better comparability, more transparent reporting, and greater methodological standards and rigour.

IIR research exhibits a large variety of research designs and methods [20]. This methodological diversity and richness needs to be investigated to identify potentials and ways of representing these research materials suitable for re-use. Therefore we need to develop an understanding of how and when researchers currently re-use or would like to find and re-use materials. This is, in fact, an IIR research problem itself!

The interactive workshop format is used to gather information and input from the IIR community regarding the kind of methods used, their experiences with both re-using materials and making materials available for re-use, and ideas on how to move towards increased sharing and re-use. The organizers already represent a wide range of IIR research perspectives, but to achieve the goal of encouraging re-use a high degree of community input and commitment is needed in order to identify feasible re-use aspects as well as limitations.

At the BIIRRR 2018 workshop, a number of high-level focus areas were identified [4]. To make further progress on these issues, the focus of the 2019 edition centered on the following five topics:

- **Terminology** The terms used in IIR studies, their definitions, and origins.

- **Methodology** General and specific methodologies employed in IIR studies, their origins, and their re-use.
- **Research design** The overall research design structures and patterns employed and their potential for re-use.
- **Re-use** Existing resources re-used as part of IIR studies, how to find them, and issues with re-using them.
- **Documentation** What and how to document aspects of IIR studies to maximize the potential for re-use.

BIARRR 2019 addressed these topics by combining short paper presentations with breakout groups. This workshop summary paper first describes earlier and related efforts preceding the BIARRR 2019 workshop (section 2), before delving into the focus areas (section 3). Section 4 describes the workshop program, the short papers presented and contains a brief outline of the breakout sessions. Section 5 discusses the outcomes and continuing activities planned after the workshop.

2 EARLIER AND RELATED EFFORTS

There have been several successful gatherings directed toward addressing the need for considering how to collect, organize, maintain, and share research resources for conducting IIR experiments. IIR campaigns on this topic include the TREC Interactive Track (1997–2002) [22], the INEX Interactive Track (2004–2010) [21, 26, 29], the Cultural Heritage in CLEF (CHiC) Interactive Task (2013) [24, 30], and the interactive Social Book Search (iSBS) task (2014–2016) [12, 13, 15] which provided great insight into the challenges and opportunities for long-term, re-usable IIR research materials.

While these demonstrate the ongoing interest in standardising the evaluation of IIR studies [20], they also show that establishing and maintaining a collaborative platform for the re-use of IIR research instruments is still an open issue. This is due to the complexity of IIR studies, which require a combination of system- and user-centered evaluation approaches [19] as well as some flexibility and individuality. In addition to the tasks and document collections that are needed and provided in most shared tasks, participants, search contexts, tasks, processes, systems, datasets and evaluation measures all need to be modeled to enable re-use for IIR studies. Past IIR studies (published at IiX and CHIIR) were analyzed with respect to these design components for this workshop [23].

There have been efforts to collect and make available some of these IIR research components. The Repository of Assigned Search Tasks (RepAST)⁵ collects, analyzes, and shares search tasks taken from publications of IIR studies. RepAST contains bibliographic data and abstracts from approximately 750 published papers, as well as a list of author-identified search task types (e.g., complex, simple, subject, known-item, factual), and the full text of any assigned search tasks reported in the papers [10]. As such, RepAST serves as a library of tasks. Members of the IIR community are encouraged to compare and contrast task descriptions as well as reuse the tasks in the collection. While RepAST is valuable, it has been underutilized to date, likely due to a lack of awareness.

Issues related to re-use in IIR have also been discussed at various workshops, including the Supporting Complex Search Tasks (SCST) workshops in 2015 and 2017 [1, 11], which were organized based on the experiences of running the iSBS shared task. In particular, the

discussions at the popular SCST 2017 workshop (co-located with CHIIR 2017) identified a strong desire within the IIR community to address the issues around re-use. This led to the BIARRR 2018 workshop at CHIIR 2018, which focused exclusively on the re-use issue and resulted in the publication of a summary paper [4] as well as a grant proposal, and spawned several informal follow-up meetings, which planned the BIARRR 2019 workshop. It also served as a starting point for a concrete, community-driven effort focused on the challenges and opportunities for designing and implementing a platform for the collection, organization, maintenance, and sharing of resources in IIR experimentation.

Further afield, related efforts and platforms can be identified, particularly in the social sciences⁶. There, the discussion on re-use of quantitative versus qualitative research data appears to grapple with similar challenges as the comparison between system-based IR evaluation and IIR experimentation [2, 9].

3 BIARRR 2019 FOCUS AREAS

In order to incorporate the community’s experience, the workshop invited *experience papers* that detail methodological and re-use aspects of previously published or in-press IIR studies. Rather than focus on research questions and results, *experience papers* were to focus on the following aspects of IIR studies, which are generally under-reported in scientific publications:

- **Terminology** What terminology did you use to describe the different components of the study? Why did you choose this terminology? How did you develop this terminology?
- **Methodology** What overarching and specific methodologies did you employ in the study? How did you decide which methodologies to employ? Examples of overarching methodologies include qualitative/quantitative/mixed methods, theory/practice/design, distant/close reading, big data/small data. Specific methods include, for instance, log studies, eye tracking, A/B testing, and simulated work tasks.
- **Research designs** What research design(s) did you use? Which (aspects) of these have the potential to be re-used? To capture the variation within the IIR field, we use the broad definition of research designs from Cheek [7]: “*the way in which a research idea is transformed into a research project or plan that can then be carried out in practice by a researcher or research team*”.
- **Re-use** What previously created materials did you re-use? This can cover all aspects, such as research designs, software, interfaces, data, scales, and specific survey questions. How did you decide what to re-use? How did you discover the materials that you re-used? Which problems did you encounter searching for them?
- **Documentation** What aspects of your study could be re-used and how have you documented and represented them to enable re-use? What aspects were fully documented in your publication? What aspects do you feel should be documented outside the main publication?

⁵<https://ils.unc.edu/searchtasks/search.php>

⁶See, for example, [8] and <https://qdr.syr.edu/>

4 WORKSHOP ACTIVITIES

The BIIRRR 2019 was a highly interactive, full-day workshop, which combined accepted presentations, discussion lead-ins, and break-out discussions.

4.1 Workshop Program

We started the day with a full round of introductions of all participants, asking them to identify their interest in the workshop. The morning program⁷ contained presentations of six accepted *experience papers*, organized to elicit more interaction and discussion. After each paper, a short question & answering round followed to collect critical aspects for re-use that were mentioned by the presenters. After the paper presentations, participants discussed the experiences and interplay of the presented work and positions.

The afternoon was organized around break-out sessions dedicated to discussion of the workshop themes and the identified concrete re-use issues from the paper presentations. The workshop closed with a concrete mission statement and a clear plan for future work.

4.2 Workshop Papers

Six *experience papers* were presented at the workshop. They ranged from describing a concrete IIR study [17], the development of a re-usable dataset [31], the development and challenges of re-usable software for IIR experiments [14], the re-use aspects of two interactive evaluation tracks [24, 25] and an analysis of IIR experiments published at previous IiX and CHIIR conferences [23]. We briefly summarize each of the experience papers below and relate them to the re-use themes covered in the workshop.

The Multi-Stage Experience: the Simulated Work Task Approach to Studying Information Seeking Task Stages. Huurdeman et al. [17] demonstrate the complexity of IIR experiments by pointing to the several aspects and decisions need to be considered during the preparation and implementation of an IIR study. In order to reconstruct and possibly re-use previous work a detailed description and documentation is needed that is often missing due to space limitations in conference papers. In general it was pointed out that checklists and protocols as provided by Borlund [5] are essential and helpful tools that should be reported and provided for re-use. However, for their own material the authors identified several challenges with respect to the maintenance of system components as well as legal and privacy issues when it comes to the storage and transmission of research data.

Data Sets for Spoken Conversational Search. Trippas and Thomas [31] compare their experiences creating two data sets for spoken conversational search by focusing on differences in terminology, methodology (e.g., transcription protocols), and research designs (e.g. in the form of tasks provided to the participants). They found that despite their involvement in the creation of both data sets, they still had to be careful with regard to comparing them and identifying where re-use would be possible. They therefore caution about reuse in interactive studies “despite careful design and description, and despite close similarity in protocol”.

To Re-use is to Re-write: Experiences with Re-using IIR Experiment Software. Hall [14] describes his experiences with developing and re-using software for both of experimental workflow and search user interfaces in IIR experiments. ESS, the software for implementing experimental workflows, saw a considerable amount of re-use over a five-year period by the main developer. However, re-use by other academics has been limited due to missing documentation and the complexity of using it. Additionally after five years of intermittent development and expansion the maintainability of the software has significantly decreased. The second software component described by Hall, PyIRE, was a workbench for developing search user interfaces to be used in conjunction with ESS. The huge variety in required UI components between the different IIR experiments led to a large amount of re-writing and very limited re-use. As a result, Hall argues for the importance of good documentation for re-use purposes, and suggests that future systems of this kind should focus on either ease-of-use or flexibility, but not both.

INEX iTrack revisited: Exploring the potential for re-use. Similar to the previous paper, Pharo [25] presents the experiences of another IIR track: the INEX iTracks conducted over a seven-year period. He provides a comprehensive overview of how the different BIIRRR 2019 themes changed over the seven-year run of the INEX iTrack, such as terminology, research design, methodology, and resources. Interestingly only a few changes were with a relatively stable framework over the years were observed. Also most components are still available (on demand) and the provided corpora have been re-used in the past. However, no concrete responsibility for the storage and maintaining of IIR track data is given and due to changing resources and people this data often disappears or is only partly available through previous organizers or participants. A single access point or repository could help overcoming this problem.

Experiences with the 2013-2016 CLEF Interactive Information Retrieval Tracks. Petras et al. [24] present an overview of the four IIR tracks at the CLEF (Conference and Labs of the Evaluation Forum) conference, including the original planning, the set-up and lessons learned for each consecutive track from 2013-2016. The paper mentions that for later re-use, we need to document in much greater detail (1) the collection characteristics for keeping collections up-to-date, but stable at the same time, (2) the users’ cultural background for more context information, (3) the underlying information seeking model and (4) the data gathering tools. Another aspect that is rarely considered are guidelines on updating information systems, but still allow backward comparability (to a previous design and user data gathered on it). The paper suggests an external repository for system software components, research designs and other materials to enable researchers benefiting from created components and to avoid problems when researchers move institutions and re-use rights are not properly assigned.

Elements of IIR Studies: A Review of the 2006-2018 IiX and CHIIR Conferences. Petras et al. [23] present an analysis of 145 papers published at past IiX and CHIIR conferences with the goal of identifying aspects of IIR studies that would need to be documented for potential later re-use. It found 10 research design components and an additional 7 contextual aspects in a first coding round. The preliminary analysis of papers reveals that these details were not

⁷See the workshop website <http://toinebogers.com/biirrr2019/program/> for the detailed workshop program.

always reported in the text—this was also confirmed by the other experience papers published at the workshop.

4.3 Breakout Sessions

The break-out discussions were organized based on the interest in the different themes from the workshop participants and identified during the paper discussions. Comments for all sessions were collected through a shared document and organized at the end of the workshop.

Interoperability. The first breakout session focused on the issue of interoperability between systems and components for supporting IIR experiments. Participants in the group had in the past developed both generic and experiment-specific systems for conducting IIR experiments and / or had used such systems. A critical discussion emerged with respect to potential re-use of system (components). Partly the group agreed that the effort required for re-use exceeded the benefit of being able to re-use existing software. One particular issue was the need for continual maintenance of software, which is much harder when there are dependencies on re-used software, which is often no longer maintained.

However, after further discussion it became clear that specific, generic elements that have the potential for re-use: pre- and post-task survey instruments and generic components such as those for logging user–system interactions. The focus on these was in part driven by the maintenance issue, as these kinds of elements can be re-used across many studies, allowing for the development of a community around the elements. Such a community would ensure long-term availability and stability of the re-used elements, significantly reducing the risk associated with re-use.

Based on the discussion outcomes and the participants experiences with building IIR study systems, the decision was made to test how easily possible it is to transfer components between systems. After some discussion the pre- and post-task survey instruments were chosen as exemplary test components. The discussion then focused on the technical and administrative organisation of such a test.

Since the lack of awareness has been identified as one major barrier on IIR material re-use, the idea of setting up a simple information page that draws together all the available tools / frameworks / toolkits / components and similar for building IIR studies was discussed.

Community involvement & advocacy. The second breakout session focused on the issue of community involvement and advocacy: what are the best ways of advocating for IIR resource re-use and reproducibility. The group members taking part in this breakout session agreed that setting up new and improving existing repositories for IIR components—tasks, instruments, document collections, system components—would be a great way of stimulating sharing and re-use. One of the original BIIRRR 2018 goals was to brainstorm about single ‘iRepository’ that encompasses all of these components in one place. However, after more discussion on this topic in the lead-up to BIIRRR 2019 and in this breakout session, participants agreed that it would probably be more beneficial to start by setting up and/or improving specialized repositories for specific types of components, such as the RePAST for search tasks. Specialized

repositories would require less effort to set up, they allow for more specialized searching and the burden of maintenance can be spread over multiple partners. This collection of repositories and resources could then be linked together explicitly by gathering them in a network and implicitly by using the DOI of the papers the different components featured in. In time, perhaps the iRepository could be then be re-imagined as a wrapper on top of these repositories that allows researchers to search through all the different specialized repositories at once, search for components by specifying a paper, or by offering a wizard-like functionality that could advise in putting together an IIR study by combining components from the different repositories.

Another idea for stimulating resource re-use in the IIR community could be to produce a manifesto on how to perform more responsible and reusable IIR research, similar in spirit to the Leiden Manifesto [16] on responsible use of research metrics and evaluation. A BIIRRR manifesto could similarly advocate for issues like transparency, documentation, accessibility, and best practices for resource sharing and re-use.

Another proposed initiative was to continue the series of experience paper presentations at a future BIIRRR workshop. The experience papers were well-received because of the transparency and the behind-the-scenes look they offered on recent research projects, in addition to the educational purpose they served for young researchers attending the workshop. Future experience papers could also come with the requirement to unearth and make available as many of the resources used in the study as possible upon acceptance to the workshop. Another possibility that was discussed could be to investigate the possibilities of a special issue dedicated to the experience of conducting specific IIR studies. We also discussed if and how documentation and sharing of resources could be included as part of the main CHIIR acceptance process. Finally, participants floated the idea of a dedicated “Most reproducible paper” award at CHIIR to reward researchers for the transparency, documentation and accessibility of their research.

IIR Study Elements. The third breakout session discussed reporting elements for IIR studies based on the presented review paper [23]. The group members taking part in the session agreed that the suggested list of elements is appropriate for describing IIR studies, and may even be flexible enough to summarize more general information behaviour studies.

The first part of the discussion focused on concrete elements in the proposed set. During the day, alternative and additional elements were suggested, such as *time* and *domain*. Terminological discussions in the session also concluded that the broader term *document corpus* should replace the *test collection* element. A discussion on whether to add additional knowledge about the researchers performing a study concluded that to record their name and affiliation was sufficient. It became apparent that the developed list needed to be revised and added to - possibly with the help of the wider IIR community.

The discussion then moved to methodological aspects of an element schema for documentation. For example, it was noted that it is necessary to standardize and have a controlled set of element terms and fields, but that this must not be at the cost of creativity in describing methods and other experimental components. A

discussion on whether to create a taxonomy for each of the the class elements was postponed to future work. Similarly, ideas to create a hierarchy for study types or a decision tree were deemed to challenging for the current state. A guide for documentation may nevertheless arise from the development of the element set. The possibility for the schema to be grouped in relationship to a theoretical (IIR) framework, such as the ones presented by Tague-Sutcliffe [28], Ingwersen and Järvelin [18] and Kelly [19] or APA's guidelines on how to report on studies and research designs was also discussed.⁸

The last part of the discussion focused on using the element documentation schema once it would be established. The group brainstormed how the schema could be instantiated with examples to demonstrate its usefulness. A systematic review of the CHIIR conference papers appeared to be a good starting point for validating the schema. Asking CHIIR authors at forthcoming conferences to document their studies with the provided schema would have the additional advantage of introducing the schema to the IIR community. The documentation schema could also be used as a teaching instrument, either to help students in summarizing IIR studies or even to help students consider the important elements of a study.

5 OUTCOMES & CONTINUING ACTIVITIES

One of the main aims of the BIIRRR workshops has always been to deliver concrete outcomes and encourage continuing activities around IIR re-use. Based on the discussions in BIIRRR 2019 the following five areas were identified and necessary worktasks assigned:

- *CHIIR Manifesto* First of all, some of the workshop organizers plan to write a CHIIR manifesto, perhaps as a CHIIR 2020 perspectives paper, for responsible and reusable IIR research in the form of guidelines for how to conduct meaningful IIR research that supports reuse and reproducibility of IIR studies. This manifesto will not only serve as a commitment by its authors to re-use, but also as an educational tool for researchers entering the field.
- *Elements for re-use* Another crucial worktask that will be undertaken is the definition of IIR study elements as well as the development of a categorization framework for categorizing that can be related to other frameworks like [18, 19, 28] and guidelines.
- *BIIRRR'20* In order to continue the fruitful exchange we will also submit another BIIRRR workshop proposal for CHIIR 2020, but with a small name change to *Building towards IIR Resource Reuse* to rather focus on the positive (Building) than on the negative (Barriers). We want to keep the experience papers 1) as a way for researchers to both reflect on their own previous work, 2) as case studies to discuss challenges and potential solutions and 3) as educational tools for early-career researchers. Some suggestions are to include a request for experience paper authors to make as much of their resources available as possible.
- *Resource Portal* A simple portal will be developed that collects IIR resources and links to toolkits, frameworks. This

will serve both as a practical and educational resource for experienced and new IIR researchers.

- *Technical / Infrastructure* Finally, two technical initiatives have been planned to support the practical aspects of re-use. First, some of the organisers plan to work on improving the RepAST repository. Second, a group of developers plan to develop and test formats for exchanging pre- and post-study survey instruments between systems.

A number of long-term issues were identified that in future might be addressed. One is data citability and options such as having a data-oriented journal for the SIGIR community or a special section in SIGIR Forum. For CHIIR 2021 we consider proposing a tutorial around reuse, focusing on e.g. reusing tasks using RepAST, or building an IIR experiment with Coagmento[27]. We are also considering a multi-day workshop to discuss documentation guidelines and the development of distributed resource portals and to plan further community dissemination activities.

Finally, we are aiming for an article in an appropriate journal to articulate our vision, detail the knowledge gained from the BIIRRR 2018 survey, outcomes of the 2018 and 2019 workshops, results of the IIR study re-use analysis, and the concrete, practical frameworks developed based on the workshops' results.

REFERENCES

- [1] Nicholas Belkin, Toine Bogers, Jaap Kamps, Diane Kelly, Marijn Koolen, and Emine Yilmaz. 2017. Second Workshop on Supporting Complex Search Tasks. In *Proc CHIIR 2017*. ACM, New York, NY, 433–435.
- [2] Libby Bishop and Arja Kuula-Luumi. 2017. Revisiting Qualitative Data Reuse: A Decade On. *Sage Open* 7, 1 (2017), 2158244016685136.
- [3] Toine Bogers, Maria Gäde, Luanne Freund, Mark Hall, Marijn Koolen, Vivien Petras, and Mette Skov. 2018. Workshop on Barriers to Interactive IR Resources Re-use. In *Proc CHIIR 2018*. ACM, New York, NY, 382–385.
- [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] Pia Borlund. 2003. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Inf. Res.* 8 (2003).
- [6] Ben Carterette, Paul Clough, Mark Hall, Evangelos Kanoulas, and Mark Sanderson. 2016. Evaluating Retrieval over Sessions: The TREC Session Track 2011-2014. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '16)*. ACM, New York, NY, USA, 685–688. <https://doi.org/10.1145/2911451.2914675>
- [7] Julianne Cheek. 2008. Research design. In *The SAGE Encyclopedia of Qualitative Research Methods*, Lisa M. Given (Ed.). SAGE.
- [8] Louise Corti. 2011. The European Landscape of Qualitative Social Research Archives: Methodological and Practical Issues. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 12, 3 (2011). <https://doi.org/10.17169/fqs-12.3.1746>
- [9] Colin Elman, Diana Kapiszewski, and Lorena Vinuela. 2010. Qualitative Data Archiving: Rewards and Challenges. *PS: Political Science and Politics* 43, 1 (2010), 23–27. <http://www.jstor.org/stable/25699288>
- [10] 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.
- [11] Maria Gäde, Mark M. Hall, Hugo Huurdeman, Jaap Kamps, Marijn Koolen, Mette Skove, Elaine Toms, and David Walsh. 2015. Report on the First Workshop on Supporting Complex Search Tasks. *SIGIR Forum* 49, 1 (June 2015), 50–56.
- [12] 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.
- [13] 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.

⁸American Psychological Association - <https://www.apastyle.org/>

- [14] Mark Michael Hall. 2019. To Re-use is to Re-write: Experiences with Re-using IIR Experiment Software. In *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2019)*. CEUR-WS.
- [15] 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.
- [16] Diana Hicks, Paul Wouters, Ludo Waltman, Sarah De Rijcke, and Ismael Rafols. 2015. Bibliometrics: the Leiden Manifesto for Research Metrics. *Nature News* 520, 7548 (2015), 429.
- [17] Hugo C. Huurdeman, Jaap Kamps, and Max L. Wilson. 2019. The Multi-Stage Experience: the Simulated Work Task Approach to Studying Information Seeking Task Stages. In *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2019)*. CEUR-WS.
- [18] Peter Ingwersen and Kalervo Järvelin. 2005. *The Turn: Integration of Information Seeking and Retrieval in Context*. Vol. 18. Springer Science & Business Media.
- [19] Diane Kelly. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends in Information Retrieval* 3, 1–2 (2009), 1–224.
- [20] Diane Kelly and Cassidy R Sugimoto. 2013. A Systematic Review of Interactive Information Retrieval Evaluation Studies, 1967–2006. *Journal of the American Society for Information Science and Technology* 64, 4 (2013), 745–770.
- [21] Ragnar Nordlie and Nils Pharo. 2012. Seven years of INEX interactive retrieval experiments—lessons and challenges. In *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer, 13–23.
- [22] Paul Over. 2001. The TREC Interactive Track: An Annotated Bibliography. *Information Processing & Management* 37, 3 (2001), 369–381.
- [23] Vivien Petras, Toine Bogers, and Maria Gaede. 2019. Elements of IIR Studies: A Review of the 2006–2018 IIX and CHIIR Conferences. In *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2019)*. CEUR-WS.
- [24] Vivien Petras, Marijn Koolen, Maria Gaede, and Toine Bogers. 2019. Experiences with the 2013–2016 CLEF Interactive Information Retrieval Tracks. In *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2019)*. CEUR-WS.
- [25] Nils Pharo. 2019. INEX iTrack revisited: Exploring the potential for re-use. In *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2019)*. CEUR-WS.
- [26] 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.
- [27] Chirag Shah. 2010. Coagmento—a collaborative information seeking, synthesis and sense-making framework. *Integrated demo at CSCW 2010* (2010).
- [28] Jean Tague-Sutcliffe. 1992. The pragmatics of information retrieval experimentation, revisited. *Information Processing & Management* 28, 4 (1992), 467–490.
- [29] 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 Szilávik (Eds.). Springer, Berlin, Heidelberg, 410–423.
- [30] Elaine G. Toms and Mark M. Hall. 2013. The CHiC Interactive Task (CHiCi) at CLEF 2013. In *Working Notes of the CLEF 2013 Conference (CEUR Workshop Proceedings)*, Pamela Forner, Roberto Navigli, Dan Tufis, and Nicola Ferro (Eds.), Vol. 1179. CEUR-WS.org.
- [31] Johanne Trippas and Paul Thomas. 2019. Data Sets for Spoken Conversational Search. In *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2019)*. CEUR-WS.