A Manifesto on Resource Re-Use in Interactive Information Retrieval

Gäde, Maria; Koolen, Marijn; Hall, Mark; Bogers, Toine; Petras, Vivien

Published in:
Proceedings of the 2021 Conference on Human Information Interaction and Retrieval

DOI (link to publication from Publisher):
10.1145/3406522.3446056

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

A Manifesto on Resource Re-Use in Interactive Information Retrieval

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

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

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

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

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

ABSTRACT

This perspective paper on resource re-use intends to draw the attention of the interactive information retrieval (IIR) community to the challenges of research documentation and archiving for future use. Resources are understood as encompassing research designs, research data and research infrastructures. It proposes eight principles for improving the re-use of resources in the IIR community and presents concrete steps on how to achieve them. A five-level system for data archiving and documentation envisions increasingly open and stable documentation and access infrastructures.

KEYWORDS

Research data, research design, research software, re-use, data documentation, data sharing practices

ACM Reference Format:
Maria Gäde, Marijn Koolen, Mark Hall, Toine Bogers, and Vivien Petras. 2021. A Manifesto on Resource Re-Use in Interactive Information Retrieval. In *Proceedings of the 2021 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '21), March 14–19, 2021, Canberra, ACT, Australia.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3406522.3446056

1 INTRODUCTION

What makes a manifesto a manifesto? While consulting research manifestos across different disciplines [11, 13, 16, 26, 33, 36, 37, 43, 48, 62, 64], we found that a manifesto (1) identifies and defines a problem or area that has not gathered much traction in the research community, (2) motivates and argues why the problem needs to be tackled, (3) develops principles or recommendations for the community to address the challenge, and (4) ends with concrete calls to action.

This manifesto on resource re-use wants to draw the attention of the interactive information retrieval (IIR) community to the challenges of research documentation and archiving for future use. It advocates the uptake of sharing and re-use in the IIR research community in order to support validation, standardization and ultimately comparability or even reproducibility of IIR research. While IIR prides itself on the heterogeneity of research approaches and the inclusiveness of the community, it is a sign of a maturing research area that methodologies and research designs increasingly adhere to certain standards and protocols. Some standardization of empirical research can already be observed in the IIR community, such as the evaluation of an experiment and its appropriate documentation.

While our vision of an ideal future does not include total conformity of research questions or methodologies, this manifesto does stress the importance of well-documented and—in the best of all cases—re-usable and shareable resources in IIR research in order to progress the discipline. Only if we successfully trace, document and reference our research designs, data and—optimally—infrastructure, will we be able to build stably on top of the empirical and experiment-based research foundation that has already been established for IIR and move the state of the art forward without the risk of repeating research efforts needlessly.

This manifesto is the result of a development process that started several years ago, when several IIR evaluation campaigns were organized at CLEF, including the Interactive Task of the Cultural Heritage at CLEF lab (CHiCi) [45, 58] and the interactive Social Book Search track (iSBS) [19, 20, 24], where organizers and participants grappled with preserving research designs and research data across the different years the track was running. The iSBS track's goal was to aggregate a large pool of experimental interaction data from a book search portal in order to support data analysis with multilingual and multicultural user groups. Data gathering was performed by different research groups using the same platform that was provided by the organizers several years in a row. The challenges that were encountered in just three years of attempting to maintain continuity in the research design and making the research data available, reinforced the idea that research data management, documentation and archiving are important issues that the IIR community has yet to tackle effectively [5]. These ideas were elaborated in two workshops at CHIIR, the *Barriers to Interactive Information Retrieval Resource Re-use* workshop series (BIIRRR) [4, 6], where the discussion focused on the challenges of re-using IIR resources—the aspects or components of an IIR experiment that should be documented and re-used. The BIIRRR workshop also

CHIIR '21, March 14–19, 2021, Canberra, Australia.

ACM ISBN 978-1-4503-8055-3/21/03. https://doi.org/10.1145/3406522.3446056

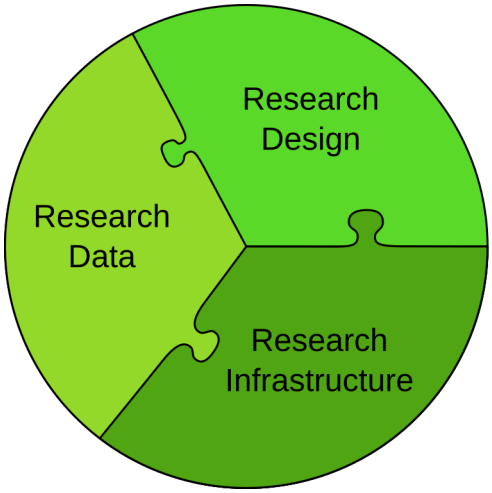

Figure 1: The three resource types that make up research: research design, research infrastructure, and research data

tested the publication of so-called *experience papers*, which allowed authors to discuss and document the research designs of previously published studies in a more detailed way—with room for describing failed experiments—which were not possible in the original outcomes-focused research publications.

This perspective paper—the Manifesto on Resource Re-use in IIR—is the culmination of these community discussions and thought processes. We first define the concepts re-use and resource (Section 2) before we motivate why re-use and documentation are important steps in moving IIR research forward (Section 3). We then propose eight principles for improving the re-use of resources in the IIR community and present concrete steps on how to achieve them (Section 4). Finally, we outline a plan on how to achieve the principles (Section 5).

Our manifesto is intended to draw the IIR community's attention to aspects in our research projects that are sometimes overlooked in favor of ever more creative and innovative design work. We hope to spark a discussion within the wider community on the principles which, at the very least, should increase awareness of research documentation aspects in our community and, ideally, have a lasting and positive effect on own research practices.

2 DEFINING RESOURCE AND RE-USE

As we will expand upon further in Section 4.2, we believe that clearly defining the terminology used in a research study is essential to its successful future documentation and re-use. We therefore start our manifesto by defining the terminology used in the rest of the paper.

2.1 Resources

Research is a complex process that is supported by three main resource types that should be documented and re-used. As shown in Figure 1, we view these resources as interlocking puzzle pieces representing the different elements of the research process: (1) research design, (2) research infrastructure, and (3) research data.

Research design, the first type of resource, is defined here as the methods and techniques used to collect and analyse the empirical data [10]. This includes the design of the research environment (e.g., location of study, study participants, tasks), any data collection protocols or other instruments used (including the questions and scales) as well as the data analysis methods and measures [15, 44].

We do not consider the physical research environment to be part of the research design, but part of the research infrastructure instead. In many research fields and in IIR in particular, a technical infrastructure is also required and designed to handle the experimental procedure and tools and support the overall execution of the research design. In IIR, the *research infrastructure* usually provides access to an IR system (including software, interfaces and collections) as well as the application of the data collection techniques such as user management, pre- and post-test questionnaires, assignment and randomization of user tasks, interaction logging, and other components [22, 51, 57].

Finally, the type of resource that has arguably received the most attention in recent years is *research data*. An increasing focus on open access and open science has been responsible for promoting research data publication and the further development of research data repositories [40]. Research data is typically defined quite broadly, for instance by Pryor [49, p. 3], who defines it as "the output of any systematic investigation involving a process of observation, experiment or the testing of a hypothesis, which when assembled in context and interpreted expertly will produce new knowledge", a heterogeneity that is echoed by Borgman [7]. We define research data as any data that has been collected, observed, generated or created during or as a results of the research process.

For the purposes of the IIR community and this manifesto, we view research as a holistic process and consider the term 'resource' to cover any component belonging to one of these three resource types: research design, research infrastructure and research data.

2.2 Re-use

Several recent articles discuss and disagree on what exactly constitutes re-use. Pasquetto et al. [42] define *re-use* as use of data by someone other than the creator of the data. When the creator uses the data beyond the original purpose for which it was created, it is still embedded in the creator's context, which they define as *use*. In the context of reproducibility there are different types of re-use, e.g., "reanalzying published data, repeating the study, reprocessing the 'raw' data, or replicating the findings under different conditions" [41]. Broader re-use includes activities such as "returning to one's own data for later comparisons, acquiring datasets from public or private sources to compare to newly collected data, surveying available datasets as background research for a new project, or conducting reanalyses of one or more datasets to address new research questions" [41]. Pasquetto et al. [41] therefore see re-use as a process, not as a single action.

In defining different levels of re-use associated with different curation intensities, the U.S. National Science Board [59, p. 20] propose a continuum of local to global re-use, which ranges from *research data collections* that are focused on the specific research project for which they were made with minimal curation, to *resource collections* that serve a community and with more curation, all

the way to *reference collections* that serve a wider community and follow robust and comprehensive standards. The characteristics of the resource collection determine its potential for re-use by others, with re-use in a wider community requiring increasing levels of curation and documentation.

There is a set of related concepts regarding re-usable research, which Schöch [52] considers distinguishable by three variables: research question, research data and research method. In Schöch [52]'s conceptual analysis, *re-use of data* happens when it is used for a different research question using a different method. Using the same data with the same method for a different research question would be *reinterpretation*. Using the same data with the same question and method would be *replication*, while using a different dataset with the same question and method would be *reproduction*. Re-use is here defined not by the data producers or re-users, but by the context within which the use takes places.

Van de Sandt et al. [60] compared 20 definitions of re-use found in the literature and provided an analysis of the discourse around use and re-use to identify and compare their characteristics. They find that there is no clear way to distinguish between use and re-use, as there are no attributes or characteristics that clearly delineate the two. This is particularly true in disciplines where research data is not necessarily generated for a single project, but for community use, such as social science panels on societal questions. Van de Sandt et al. define (re-)use as "the use of any research resource regardless of when it is used, the purpose, the characteristics of the data and its user".

For the purposes of the IIR community and this manifesto, we define re-use in its broadest sense as use of research data, research designs or infrastructure for more than an individual purpose.

3 THE IMPORTANCE OF RE-USE

Research cultures that promote re-use and support it through sustainable data sharing habits, policies and infrastructures allow its researchers to spend more of their time working on knowledge discovery and innovation instead of repetition of development work. It also allows for the replication and reproduction of earlier studies, thereby vetting those results and solidifying the joint foundation of scientific knowledge that future studies can build on [8]. Young researchers that aim to enter a research field and interdisciplinary projects building upon previous research will particularly benefit from data sharing. The formulation of the FAIR standards in 2016 with the explicit goal of ensuring the sustainable use of research data exemplifies this mindset [63]. FAIR means that research data should be (1) findable, (2) accessible, (3) interoperable and (4) re-usable—with these principles applying equally to researchers (data producers) and repositories (data providers). Our manifesto is closely aligned with the FAIR principles [63], but focuses on the specific issues encountered in IIR, while at the same time expanding the scope beyond research data to encompass research designs and infrastructure.

While research data access, re-use, and reproducibility have been promoted for years, there is still much we do not know about the intentions, incentives, practices and barriers to re-use [41, 60]. One of the major barriers to optimal re-use of research data is often the lack of contextualization and implicit knowledge transfer [41]. Sharing research data means more than just sharing the original input and output data to an experiment. For a complete understanding and reliable interpretation of the data, the context of data production also needs to be documented. This includes information about the environment in which the experiments were designed and took place as well as implicit knowledge that was gained and drawn upon before, during, and after the study.

For contextual information, the research designs of data producing studies also need to be documented, archived, and shared to enable their re-use. While complete and transparent documentation of research designs is important for all types of studies, it is perhaps even more important in the context of qualitative research, which is part of the IIR method portfolio. Here, re-using the research design is often the only possibility to reproduce or validate previous studies, since it is often more problematic to make qualitative research data available for re-use due to ethical or legal issues, in particular privacy protection—such as interview data for example [32, 34].

The development of the research infrastructure represents a major investment of time and effort, much of which is duplicated or could be reduced through adaptation and re-use. The difficulty is that—except for very generic components, such as the underlying information retrieval system—re-use generally requires rewriting significant parts of the code base [22], resulting in a tendency to rewrite from scratch. Even for more generic aspects of the IIR research infrastructure, such as software to manage the participants flow through an experiment, the attempts at developing a re-usable solution [25, 54, 57] have not achieved any uptake. This mirrors the state of research software re-use in general [2, 14, 21].

Some disciplines, such as the social sciences or psychology, have already progressed far in documenting and providing access to research data and research designs [50, 55]. Resource re-use is no less important for the IIR field: IIR studies are typically complex due to the interaction between users and the search system(s) and the combination of user- and system-centered evaluation [30]. This necessitates sophisticated research designs with rich combinations of data collection and analysis methods and a research infrastructure with many interdependent IR system components. In the past few decades, there have been several multi-year IIR campaigns that have attempted to re-use the experimental setup from year to year within the campaign, such as the TREC Interactive Track (1997–2002) [39], the INEX Interactive Track (2004–2010) [38, 47, 56], the Cultural Heritage in CLEF (CHiC) Interactive Task (2013) [46, 58], and the interactive Social Book Search (iSBS) task (2014-2016) [19, 20, 24].

While these campaigns have provided a great deal of insight into the challenges and opportunities for longer-term re-use of IIR resources [5, 22], there has been relatively little re-use of resources related to research design or research infrastructure from campaign to campaign. Despite an ongoing interest in standardising IIR experimentation and evaluation within evaluation campaigns, this shows that establishing and maintaining a collaborative process and platform to support the re-use of IIR resources remains an unsolved issue. This is further compounded by a need for flexibility in being able to tailor the research design to the specific search scenario of each study or campaign.

Nevertheless, there have been efforts to collect, archive, and make available certain types of IIR study resources. For instance,

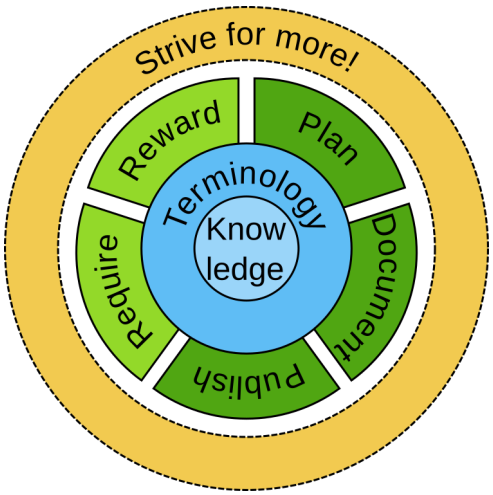

Figure 2: Principles to Support Re-Use in IIR

the Repository of Assigned Search Tasks (RepAST)[1] collects, categorises, and provides access to a large collection of search tasks mined from publications of IIR studies [18]. Apart from the RepAST initiative, however, no continuity or satisfying re-use practices appears to have been established throughout the years [5]. Our goal with this manifesto is to provide both guiding principles as well as concrete steps to stimulate and strengthen the re-use of IIR resources.

4 PRINCIPLES TO SUPPORT RE-USE IN IIR

We now present eight principles for improving resource re-use in the IIR community, which are visualized in Figure 2. At the core lie the principles on *Knowledge* (#1) and *Terminology* (#2) that underpin everything else. The remaining principles are split into those aimed at the individual researcher—*Plan* (#3), *Document* (#4), and *Publish* (#5)—and those aimed at the IIR community as a whole—*Require* (#6) and *Reward* (#7). Around all of these sits the *Strive for more!* principle (#8) that encourages all of us, both individually and as a community, to move re-use forward. For each principle, we describe who we argue is primarily responsible for it as well as which of the three resource types the principle addresses; Table 1 contains an overview of this information. All principles apply to all resource types (see Section 2.1), but may be more relevant for some than for others (see Section 5.1).

4.1 Principle 1: Make tacit knowledge explicit

All implicit knowledge, choices, and study aspects should be noted, codified, and made explicit.

Transforming existing protocols and research methods, with all the distinct elements that they consist of, into a study requires a large number of choices, that are often based upon implicit knowledge and experience. These implicit aspects of the study are often hard to codify, but they need to be documented and made explicit to enable any re-use. This extends into the study itself, where checklists and notes should be used to capture exceptions and unusual aspects or events [9].

Inspired by the Manifesto for Research Services [11], this principle helps not only with reproducibility, but also reduces the risk of biases and other issues that later call into question the study's results. Additionally, by describing the tacit knowledge of a study and comparing it with others, it becomes possible to gradually develop and merge that tacit knowledge into the research design, formalisms and procedures, which can then be more easily codified [27] and become common ground.

Concrete actions.

- Prepare a checklist of the practical matters to take into account before, during, and after the experiment.
- Document your exact study goals, research questions, premises, and (if applicable) hypotheses.
- Take notes, screenshots (and photos or videos where appropriate) during your experiments.
- Take note of actions and decisions that are not part of standardized research designs and not on the checklist.
- Take note of aspects that are difficult to reproduce and describe them.

4.2 Principle 2: Define your terminology

All concepts and terms should be precisely defined, ideally referring to existing, published definitions.

Even though many concepts and terms are common in the community, there are often multiple definitions with subtle or not so subtle differences. The specific definitions used in a study should be made explicit, including references to publications where they exist. Where existing definitions are not used, a clear explanation should be provided for why new definitions are required. The terminology should also be used precisely and consistently throughout the study.

This is particularly important as IIR is a highly interdisciplinary field, with researchers coming from various academic backgrounds, who might not all be familiar with the same concepts and terminology, or who would assume different definitions based on their disciplinary background and training. Reproducibility, for instance, is defined by Claerbout and Karrenbach [12] as the same researchers using the same experimental setup, whereas the ACM definition [1] refers to this as repeatability. Instead, ACM defines reproducibility as a different research group using a different setup to confirm the reported outcomes and findings. The potential misunderstandings that could arise from such subtle differences could be avoided by providing clear and specific definitions.

Concrete actions.

- Identify which concepts and terminology you will use while creating your research design.
- Look for alternative definitions or related concepts.
- Discuss differences in definitions and uses of concepts and terms that you find across publications.
- Discuss to what extent they apply to your work.

[1] https://ils.unc.edu/searchtasks/search.php, last visited January 19, 2021.

Principle	Primary responsibility		Resource types addressed		
	Researcher	Community	Designs	Data	Infrastructure
#1 Knowledge	✓		✓		
#2 Terminology	✓		✓		
#3 Plan	✓			✓	
#4 Document	✓		✓	✓	✓
#5 Publish	✓		✓	✓	✓
#6 Require		✓	✓	✓	✓
#7 Reward		✓	✓	✓	✓
#8 Strive	✓	✓	✓	✓	✓

Table 1: Overview of the primary responsibility for each of the eight principles (left) as well as the different resource types addressed by them (right).

- Argue for the definitions you choose or why you introduce your own.
- Cite papers that define concepts that you use including those with alternative definitions that you do not use.

4.3 Principle 3: Plan the data life cycle

The full data life cycle, covering collection, processing, analysis, publication, and archiving, should be planned before starting any data collection.

Data collection is often time- and resource-consuming. Therefore, IIR research should start with an initial Data Management Plan (DMP) that clearly describes what kind of research data will be collected or generated during the study, how it will be stored, managed, described and analysed, and how it will be shared and archived at the end of the study. This ensures that the collection and usage of data is transparent to the participants and research community.

In the fields of medical and health studies, presenting such as DMP in advance is generally a basic requirement [31, 35, 61] and these plans are usually made publicly available. Increasingly, funding organizations require a DMP to be submitted together with a project proposal in most data-producing research areas [53]. There are several tools, which will support the creation of a DMP[2], which will also provide guidelines on how to document and archive data appropriately.

Concrete actions.

- Start with a data management plan and document what data will be collected and for which purpose.
- Make sure that your data collection practice follows current data protection and privacy regulations such as the General Data Protection Regulation (GDPR) in the European Union.
- Publish your data management plan as part of your research project and/or publication of results.

[2]See for instance https://dmponline.dcc.ac.uk/, last visited January 19, 2021.

4.4 Principle 4: Document your research resources

All aspects of the research resources should be documented in detail.

In principle, everything required to reproduce an experiment, including the raw research data, analyses and systems used should be documented. As an initial minimum requirement, the research design must be documented to enable a basic level of resource reuse. This includes the experimental setup (e.g., location of study, study participants, tasks), any data collection protocols or other instruments used (including the questions and scales) as well as the data analysis and measures [44].

Documenting the research design must include implicit knowledge and terminology definitions as covered by principles #1 and #2. Additionally, contextual information, such as the study goals, research questions, or associated publications should also be documented. Where available, standardized and machine-readable documentation formats should be used, e.g., the User Study Exchange Format [23].

In a more advanced re-use context, the research data and research infrastructure (including software) should also be documented in similar ways.

Concrete actions.

- Review existing practices and guidelines (in the social sciences it is common to archive codebooks for research designs, for example).
- Define your data model or use a community standard for research resource documentation.

4.5 Principle 5: Archive & publish your research resources

All documentation (see principle #4) should be archived in open access, long-term storage to be available for publication reviews and should then be published in a referenceable location.

Publications are generally limited in the amount of space available to authors and details on the research designs, data and infrastructure are frequently the first victim of this limit [29]. There are several platforms where additional research data and documents can be formally published, e.g., the Open Science Framework[3], Zenodo[4], or figshare[5]. Many funding organisations already require that publications must be made available via public open access repositories—some of them can also be used to archive the research designs and other research resources.

A dedicated repository for research designs in IIR does not (yet) exist, but for some resources related to the research designs, specific repositories exist, for example RePaST for tasks [18]. Tasks used in studies should therefore also be provided to those repositories to provide more access points for the community. This may involve some duplication of work, but if machine-readable formats are used, this should reduce the effort required.

Concrete actions.

- Identify an appropriate open access archive repository.
- Separately archive those aspects of the resources that can be freely shared and those that cannot for legal or other reasons.
- Include links to the archived resources in the publication submission.
- After publication of research outcomes, publish those resource aspects that can be freely shared and update the publication to refer to those.
- Publish parts of the resources to specialised repositories, where such exist.

4.6 Principle 6: Require basic sharing practices

Publication venues should require that resources be archived and available to reviewers.

The foundations for this principle are often already in place with publishers [17]. What is needed is that journals and conferences in the IIR field adopt them. This principle also goes hand-in-hand with principle #7, which is focused on rewarding good resource sharing behaviour. Together they should produce better, more co-operative outcomes [3]. Optionally, publication venues could also require that the archived resources are published after the main publication.

To ensure uptake, minimally the research design documentation must become a required part of the publication process. For complete transparency, also the research data and infrastructures should be made available first to reviewers and then—if at all possible—to the public. If a data archival section is already a requirement for a publication, the research design documentation can be integrated there, but if not, then there should be a required section in the paper that includes a referenceable link to the documentation.

Concrete actions.

- Require that at least documented research designs for IIR studies are available to reviewers at CHIIR.
- Petition journals relevant to the IIR field to add research design and data archiving into their publication workflow.

4.7 Principle 7: Reward good sharing practices

Researchers who follow good sharing practices should be visibly rewarded.

Complementing the basic requirements in principle #6, the IIR community should also develop a range of rewards to highlight good reusability practice. Changing the incentives for responsible sharing practices is a problem that requires a coordinated effort by all stakeholders to alter existing reward structures—authors, reviewers, chairs/editors, publishers, funders, institutions, and societies.

One particular reward mechanism that shows promise are *badges* to acknowledge reusability practices. Promoting or even requiring such badges would signal that the IIR community values these practices. An example of the successful use of such a badging system is the *Psychological Science* journal, where the adoption of open data badges has had a positive effect, increasing data sharing by a factor of more than ten [37]. Another such initiative is by the Association for Computing Machinery (ACM), which proposes a set of badges to denote functional, reusable, and available research artifacts [1].

Another approach for rewarding good practices would be to provide incentives towards providing more detail when reporting methodology, study procedures, i.e., the research design, and results. This is achieved by either allowing authors to add extra pages to a publication or provide additional publication venues or categories, dedicated to the reporting of research designs. This would remove the current incentive structure towards providing simplified, 'clean' narratives, which can result in the incomplete reporting of research designs or results.

Concrete actions.

- Allow for adding pages to a publication that are focused on reporting research designs.
- Institute a badging system with different levels at CHIIR, ISIC, and topically related journals.
- Institute an award for best-documented research paper at conferences like CHIIR.
- Add a publication category focused on reporting research designs, such as the experience papers at the 2019 BIIRRR workshop.

4.8 Principle 8: Strive for more!

Unlike the previous seven principles, which cover the basics required to embed re-use practices into the IIR community, this principle covers where the IIR community could go next. As such, the principle does not contain any concrete steps, only potential actions.

Potential actions.

- If you have shared your research design and data management plan, the next step is to share your research data and infrastructure. If you have shared your data and research

[3] https://osf.io
[4] https://zenodo.org
[5] https://figshare.com/

infrastructure, the next step is to add documentation. Documentation can have different levels of detail and explicitness and different levels of structure and adherence to conventions. In the next section we offer concrete suggestions for increasing levels of documentation and sharing.

- Beyond sharing the designs, infrastructure and data of your research, you could pre-register your IIR studies, including hypotheses that are to be tested. There is a call for more transparency, including pre-registration of studies and hypotheses, across an increasing number of fields [31, 35, 61]. Although a significant part of our research is exploratory, where we do not have a very specific hypothesis to test, there are many studies where we have very concrete and specific questions and hypotheses. Pre-registration of the design, setup and hypotheses of such a study helps others to understand the original motivation behind the design and setup of the study, so that its components can more easily be re-used and compared against.
- Another step beyond sharing resources is to try to reproduce an IIR study. This requires a deep understanding of the theory as well as the research design of the original study. With proper re-use of IIR research resources, this type of reproducibility study could be used for graduate work. Organizing dedicated reproducibility tracks at conferences for Master's or PhD students and their supervisors could provide great training opportunities as well as a deeper understanding of which original studies truly contribute to our understanding of information seeking behavior.

5 THE WAY FORWARD

5.1 Five Levels of IIR Resource Re-Use

Inspired by the Five Star Linked Data system [28], which describes a hierarchy of steps to increase the level of openness and linking in data, we envision a five-level system for data archiving and documentation to increase the re-use of IIR resources. While level five is the ultimate objective for ensuring open documentation and re-use, every level contributes to a more standardized, documented and re-usable resource space for IIR. The levels build upon each other: a level can only be reached if the levels that come before it have already been accomplished. The last three levels are advanced levels, which depend on open and long-term accessible repositories and data models for structured documentation. While there are examples for research design and research data repositories available, there is no dedicated repository for IIR resources.

- *Level 1: Documentation of the resources in the research publication.* The description of the research design, data and infrastructure in the publication describing the study is already a community standard. We have included this as a first level, because very often, the description of the methodology and data lacks sufficient detail or is necessarily abbreviated because of typical length restrictions in conference and journal papers. Length restrictions should not preclude authors from at least precisely describing their research designs.
- *Level 2: Extended documentation of the resources on a project website.* To overcome information gaps and therefore challenges in reproducibility and re-use because of brevity restrictions, the study should be documented extensively on a publicly reachable website, which the authors create and maintain. Even though there is a risk that such websites may disappear, we argue that this level represents an important step. First, even if the documentation does completely or partially disappear after some time, this is still preferable over not having had any documentation at all. Second, 'free' and long-term options such as GitHub Pages are widely available, ensuring at least medium-term availability.
- *Level 3: Structured documentation of research design in a repository.* Documenting the research design in an openly accessible repository overcomes the challenges of maintenance and continuity within a research group, when researchers are leaving. Documenting research designs in a structured way—using a community-accepted data model, which includes the necessary information to reproduce an experiment—does not only provide a guideline on how to document research designs, but also makes them comparable across studies.
- *Level 4: Structured and documented archival of research design + research data.* While documenting a research design in an open repository is relatively straight-forward, because it shouldn't incur any rights or permission issues, this next level requires not only the documentation of the research design, but also the documentation and the archiving of the generated research data in a repository. Preparing research data for publication on a repository does not only include the documentation and preparation of the data, but also involves rights clearance, which sometimes is more difficult than the documentation (see principle 3).
- *Level 5: Structured and documented open archival of research design + research data + research infrastructure.* Ultimate re-use opportunities are provided, if not only the research design and research data are made available, but also the research system, so that other groups can experiment with it. Since research software is rarely stable and well-documented [22], this is may be an even bigger challenge then documenting and archiving research designs and research data, as evident from the lack of re-use of IIR research software [25, 54, 57] and research software in general [2, 14, 21].

Our five levels of archiving and documentation align well with the continuum of local to global (re-)use proposed by the US National Science Board [59] referenced also in Section 2.2. The more types of resources are shared and the more they are documented in a structured and standardized manner, the more we move IIR resources from local use to global re-use.

5.2 The Manifesto and the Future

Despite broad agreement over the necessity of being able to reproduce experimental research, the current academic infrastructure does not reward reproducibility studies the same way it rewards original work. This results in the overwhelming majority of research effort being directed at new research before the community has been able to properly verify the veracity and generalizability

of recent research. If original work is required for publication, one might argue that making research resources re-usable may not be a worthwhile objective for the community to pursue.

While we believe that documenting and re-using research resources will also support original research and improve good scientific practice, another important application area for resource documentation and sharing is education and community growth. Students, young researchers and others entering the community will equally benefit from having resources available to attempt to reproduce IIR studies for their introduction to the field.

The aim of this paper is to enable a community discussion on the role of sharing and re-use of resources in IIR research, what we should strive for and how to get there. This manifesto should be discussed by the entire community and modified where necessary to ensure broad consensus. Moreover, as our eighth principle indicates, our shared understanding of how to best facilitate resource re-use develops over time, so our principles should be adapted according to new insights. Therefore, the manifesto should be treated as a living document, with this paper being merely the first version. One idea is to maintain a community repository with up-to-date best practices and resources, including this and future versions of the manifesto.

REFERENCES

[1] ACM. 2020. Artifact Review and Badging. https://www.acm.org/publications/policies/artifact-review-badging.

[2] Alice Allen, Cecilia Aragon, Christoph Becker, Jeffrey Carver, Andrei Chis, Benoit Combemale, Mike Croucher, Kevin Crowston, Daniel Garijo, Ashish Gehani, Carole Goble, Robert Haines, Robert Hirschfeld, James Howison, Kathryn Huff, Caroline Jay, Daniel S. Katz, Claude Kirchner, Katie Kuksenok, Ralf Lämmel, Oscar Nierstrasz, Matt Turk, Rob van Nieuwpoort, Matthew Vaughn, and Jurgen J. Vinju. 2017. Engineering Academic Software (Dagstuhl Perspectives Workshop 16252). *Dagstuhl Manifestos* 6, 1 (2017), 1–20. https://doi.org/10.4230/DagMan.6.1.1

[3] J. Andreoni, W. Harbaugh, and L. Vesterlund. 2003. The Carrot or the Stick: Rewards, Punishments, and Cooperation. *American Economic Review* 93, 3 (2003), 893–902.

[4] Toine Bogers, Samuel Dodson, Luanne Freund, Maria Gäde, Mark Hall, Marijn Koolen, Vivien Petras, Nils Pharo, and Mette Skov. 2019. Overview of the CHIIR 2019 Workshop on Barriers to InteractiveIR Resources Re-use (BIIRRR 2019). In *CEUR Workshop Proceedings*, Vol. 2337. CEUR-WS.org, Aachen, Germany, 1–6. http://ceur-ws.org/Vol-2337/overview.pdf

[5] Toine Bogers, Maria Gäde, Mark M. Hall, Vivien Petras, and Mette Skov. 2017. Lessons Learned from the CHiC and SBS Interactive Tracks: A Wishlist for Interactive IR Evaluation. In *Proceedings of the Second Workshop on Supporting Complex Search Tasks co-located with the ACM SIGIR Conference on Human Information Interaction & Retrieval (CHIIR 2017), Oslo, Norway, March 11, 2017 (CEUR Workshop Proceedings, Vol. 1798)*, Marijn Koolen, Jaap Kamps, Toine Bogers, Nicholas J. Belkin, Diane Kelly, and Emine Yilmaz (Eds.). CEUR-WS.org, Aachen, Germany, 11–14. http://ceur-ws.org/Vol-1798/paper2.pdf

[6] Toine Bogers, Maria Gäde, Mark Michael, Luanne Freund, Marijn Koolen, Vivien Petras, and Mette Skov. 2018. Report on the workshop on barriers to interactive IR resources re-use (BIIRRR 2018). *ACM SIGIR Forum* 52, 1 (2018), 119–128.

[7] Christine L Borgman. 2012. The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology* 63, 6 (2012), 1059–1078.

[8] Christine L Borgman. 2015. *Big data, little data, no data: Scholarship in the networked world.* MIT press, Cambridge, MA.

[9] Pia Borlund. 2003. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information research* 8, 3 (2003), 8–3.

[10] Julianne Cheek. 2008. Research design. In *The SAGE encyclopedia of qualitative research methods*. SAGE Publications, Inc., Thousand Oaks, CA, 762–764.

[11] Henry Chesbrough and Jim Spohrer. 2006. A Research Manifesto for Services Science. *Commun. ACM* 49, 7 (July 2006), 35–40.

[12] Jon F. Claerbout and Martin Karrenbach. 1992. Electronic Documents give Reproducible Research a New Meaning. In *SEG Technical Program Expanded Abstracts 1992*. Society of Exploration Geophysicists, Tulsa, OK, 601–604.

[13] Rosaria Conte, Nigel Gilbert, Giulia Bonelli, Claudio Cioffi-Revilla, Guillaume Deffuant, Janos Kertesz, Vittorio Loreto, Suzy Moat, J-P Nadal, Anxo Sanchez, et al. 2012. Manifesto of computational social science. *The European Physical Journal Special Topics* 214, 1 (2012), 325–346.

[14] Joenio Costa, Paulo Meirelles, and Christina Chavez. 2018. On the Sustainability of Academic Software: The Case of Static Analysis Tools. In *Proceedings of the XXXII Brazilian Symposium on Software Engineering* (Sao Carlos, Brazil) *(SBES '18)*. Association for Computing Machinery, New York, NY, USA, 202–207. https://doi.org/10.1145/3266237.3266243

[15] John W. Creswell and J. David Creswell. 2017. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications, Thousand Oaks, CA.

[16] Giorgio De Michelis, Eric Dubois, Matthias Jarke, Florian Matthes, John Mylopoulos, Mike Papazoglou, Klaus Pohl, Joachim Schmidt, Carson Woo, and Eric Yu. 1997. *Cooperative information systems: a manifesto*. Accademic-Press, Cambridge, MA.

[17] Taylor & Francis. 2020. Open and FAIR data sharing policy. https://authorservices.taylorandfrancis.com/data-sharing-policies/open-and-fair/.

[18] 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.

[19] 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, Vol. 1609)*, Krisztian Balog, Linda Cappellato, Nicola Ferro, and Craig Macdonald (Eds.). CEUR-WS.org, Aachen, Germany, 1024–1038.

[20] 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, Vol. 1391)*, Linda Cappellato, Nicola Ferro, Gareth J. F. Jones, and Eric SanJuan (Eds.). CEUR-WS.org, Aachen, Germany, 14.

[21] Teresa Gomez-Diaz and Tomas Recio. 2019. On the evaluation of research software: the CDUR procedure. *F1000Research* 8 (2019). https://doi.org/10.12688/f1000research.19994.2

[22] Mark M Hall. 2019. To re-use is to re-write: experiences with re-using IIR experiment software. In *CEUR Workshop Proceedings*, Vol. 2337. CEUR-WS.org, Aachen, Germany, 19–23.

[23] Mark Michael Hall and Toine Bogers. 2020. A Standardised Format for Exchanging User Study Instruments. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. ACM, New York City, NY, 457–461.

[24] 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, Vol. 1180)*, Linda Cappellato, Nicola Ferro, Martin Halvey, and Wessel Kraaij (Eds.). CEUR-WS.org, Aachen, Germany, 480–493.

[25] Mark M Hall, Spyros Katsaris, and Elaine Toms. 2013. A Pluggable Interactive IR Evaluation Work-bench. In *European Workshop on Human-Computer Interaction and Information Retrieval*. Springer, Heidelberg, Germany, 35–38.

[26] Diana Hicks, Paul Wouters, Ludo Waltman, Sarah De Rijcke, and Ismael Rafols. 2015. Bibliometrics: the Leiden Manifesto for research metrics. *Nature* 520, 7548 (2015), 429–431.

[27] Stephen Hilgartner and Sherry I. Brandt-Rauf. 1994. Data Access, Ownership, and Control: Toward Empirical Studies of Access Practices. *Knowledge* 15, 4 (1994), 355–372. https://doi.org/10.1177/107554709401500401 arXiv:https://doi.org/10.1177/107554709401500401

[28] Timothy Holborn. 2014. What is 5 Star Linked Data? https://www.w3.org/community/webize/2014/01/17/what-is-5-star-linked-data/.

[29] Hugo C. Huurdeman, Jaap Kamps, and Max L. Wilson. 2019. The Multi-Stage Experience: the Simulated Work Task Approach to Studying Information Seeking Stages. In *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use co-located with the ACM SIGIR Conference on Human Information Interaction and Retrieval, BIIRRR@CHIIR 2019, Glasgow, UK, March 14, 2019 (CEUR Workshop Proceedings, Vol. 2337)*, Toine Bogers, Samuel Dodson, Maria Gäde, Luanne Freund, Mark M. Hall, Marijn Koolen, Vivien Petras, Nils Pharo, and Mette Skov (Eds.). CEUR-WS.org, Aachen, Germany, 7–13. http://ceur-ws.org/Vol-2337/paper1.pdf

[30] Diane Kelly. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends in Information Retrieval* 3, 1–2 (2009), 1–224.

[31] Florian G Kern and Kristian Skrede Gleditsch. 2017. Exploring pre-registration and pre-analysis plans for qualitative inference. *Preprint ahead of publication* (2017), 1–15.

[32] Youngseek Kim and Melissa Adler. 2015. Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *International journal of information management* 35, 4 (2015), 408–418.

[33] Effie Lai-Chong Law, Arnold POS Vermeeren, Marc Hassenzahl, and Mark Blythe. 2007. Towards a UX manifesto. In *Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK 21*. BCS, The Chartered Institute for IT, Swindon, UK, 1–2.

[34] Sara Mannheimer, Amy Pienta, Dessislava Kirilova, Colin Elman, and Amber Wutich. 2019. Qualitative data sharing: Data repositories and academic libraries as key partners in addressing challenges. *American Behavioral Scientist* 63, 5 (2019), 643–664.

[35] Edward Miguel, Colin Camerer, Katherine Casey, Joshua Cohen, Kevin M Esterling, Alan Gerber, Rachel Glennerster, Don P Green, Macartan Humphreys, Guido Imbens, et al. 2014. Promoting transparency in social science research. *Science* 343, 6166 (2014), 30–31.

[36] Geoffrey Miller. 2012. The Smartphone Psychology Manifesto. *Perspectives on Psychological Science* 7, 3 (2012), 221–237. https://doi.org/10.1177/1745691612441215 arXiv:https://doi.org/10.1177/1745691612441215 PMID: 26168460.

[37] Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. A Manifesto for Reproducible Science. *Nature Human Behaviour* 1, 1 (2017), 1–9.

[38] 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, Heidelberg, Germany, 13–23.

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

[40] Heinz Pampel, Paul Vierkant, Frank Scholze, Roland Bertelmann, Maxi Kindling, Jens Klump, Hans-Jürgen Goebelbecker, Jens Gundlach, Peter Schirmbacher, and Uwe Dierolf. 2013. Making research data repositories visible: the re3data. org registry. *PloS one* 8, 11 (2013), e78080.

[41] Irene V. Pasquetto, Christine L. Borgman, and Morgan F. Wofford. 2019. Uses and Reuses of Scientific Data: The Data Creators' Advantage. *Harvard Data Science Review* 1, 2 (15 11 2019). https://doi.org/10.1162/99608f92.fc14bf2d https://hdsr.mitpress.mit.edu/pub/jduhd7og.

[42] Irene V Pasquetto, Bernadette M Randles, and Christine L Borgman. 2017. On the reuse of scientific data. *Data Science Journal* 16 (2017), 8. https://doi.org/10.5334/dsj-2017-008

[43] Ray Pawson. 2013. *The science of evaluation: a realist manifesto*. SAGE Publications, Los Angeles, CA.

[44] 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.

[45] Vivien Petras, Toine Bogers, Elaine G. Toms, Mark M. Hall, Jacques Savoy, Piotr Malak, Adam Pawlowski, Nicola Ferro, and Ivano Masiero. 2013. Cultural Heritage in CLEF (CHiC 2013). In *CLEF '13: Proceedings of the 4th International Conference of the CLEF Initiative (Lecture Notes in Computer Science, Vol. 8138)*, Pamela Forner, Henning Müller, Roberto Paredes, Paolo Rosso, and Benno Stein (Eds.). Springer, Heidelberg, Germany, 192–211.

[46] 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.org, Aachen, Germany, 29–36.

[47] 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.

[48] Rosalind W Picard, Seymour Papert, Walter Bender, Bruce Blumberg, Cynthia Breazeal, David Cavallo, Tod Machover, Mitchel Resnick, Deb Roy, and Carol Strohecker. 2004. Affective learning—a manifesto. *BT technology journal* 22, 4 (2004), 253–269. https://doi.org/10.1023/B:BTTJ.0000047603.37042.33

[49] Graham Pryor. 2012. *Managing Research Data*. Facet Publishing, London, UK.

[50] Karsten Boye Rasmussen and Grant Blank. 2007. The data documentation initiative: a preservation standard for research. *Archival Science* 7, 1 (2007), 55–71.

[51] Gareth Renaud and Leif Azzopardi. 2012. SCAMP: A Tool for Conducting Interactive Information Retrieval Experiments. In *IIiX '12: Proceedings of the 4th Information Interaction in Context Symposium*. ACM, New York City, NY, 286–289.

[52] Christof Schöch. 2017. Wiederholende Forschung in den digitalen Geisteswissenschaften. In *DHd2017: Digital Nachhaltigkeit (DHd2017), Bern, Switzerland, 13-18 February 2017*. Zenodo, Genève, Swizterland, 207–212. https://doi.org/10.5281/zenodo.277113

[53] Nicholas Smale, Kathryn Unsworth, Gareth Denyer, and Daniel Barr. 2018. The History, Advocacy and Efficacy of Data Management Plans. *bioRxiv* (2018), 30. https://doi.org/10.1101/443499 arXiv:https://www.biorxiv.org/content/early/2018/10/17/443499.full.pdf

[54] Diana Soltani, Matthew Mitsui, and Chirag Shah. 2019. Coagmento v3. 0: Rapid prototyping of web search experiments. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. ACM, New York City, NY, 367–371.

[55] Susan E Swogger. 2013. PsycTESTS. *Journal of the Medical Library Association: JMLA* 101, 3 (2013), 234.

[56] 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.

[57] Elaine G Toms, Luanne Freund, and Cara Li. 2004. WiIRE: the Web interactive information retrieval experimentation system prototype. *Information Processing & Management* 40, 4 (2004), 655–675.

[58] 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, Vol. 1179)*, Pamela Forner, Roberto Navigli, Dan Tufis, and Nicola Ferro (Eds.). CEUR-WS.org, Aachen, Germany, 12.

[59] U.S. National Science Board. 2005. Long-lived digital data collections: Enabling research and education in the 21st century (No. US NSF-NSB-05-40). https://www.nsf.gov/pubs/2005/nsb0540/.

[60] Stephanie van de Sandt, Artemis Lavasa, Sünje Dallmeier-Tiessen, and Vivien Petras. 2019. The Definition of Reuse. *Data Science Journal* 18 (2019), 22.

[61] Anna Elisabeth van't Veer and Roger Giner-Sorolla. 2016. Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology* 67 (2016), 2–12.

[62] Katrin Weller and Katharina E. Kinder-Kurlanda. 2016. A Manifesto for Data Sharing in Social Media Research. In *Proceedings of the 8th ACM Conference on Web Science* (Hannover, Germany) *(WebSci '16)*. Association for Computing Machinery, New York, NY, USA, 166–172. https://doi.org/10.1145/2908131.2908172

[63] Mark Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gaby Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Olavo Bonino da Silva Santos, Philip Bourne, Jildau Bouwman, Anthony Brookes, Tim Clark, Merce Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris Evelo, Richard Finkers, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (03 2016). https://doi.org/10.1038/sdata.2016.18

[64] Paul Willis and Mats Trondman. 2000. Manifesto for Ethnography. *Ethnography* 1, 1 (2000), 5–16. https://doi.org/10.1177/14661380022230679