

# Survey on Re\* Research

Oleksii Serhiienko

April 2018

Complementary Module

MSE program, Information and Communication Technology

Zurich University of Applied Sciences

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Motivation . . . . .	2
1.2	Problem statement . . . . .	4
1.3	Methodology . . . . .	5
<b>2</b>	<b>Re* research characteristics</b>	<b>7</b>
2.1	Repeatability . . . . .	7
2.1.1	Current state of art . . . . .	8
2.1.2	Recommendations . . . . .	8
2.2	Reproducibility . . . . .	9
2.2.1	Current state of art . . . . .	11
2.2.2	Recommendations . . . . .	11
2.3	Recomputability . . . . .	12
2.3.1	Current state of art . . . . .	13
2.3.2	Recommendations . . . . .	13
2.4	Reusability . . . . .	14
2.4.1	Current state of art . . . . .	15
2.4.2	Recommendations . . . . .	16
2.5	Replicability . . . . .	17
2.5.1	Current state of art . . . . .	18
2.5.2	Recommendations . . . . .	19
<b>3</b>	<b>Summary</b>	<b>20</b>
3.1	Good practices . . . . .	21

<i>CONTENTS</i>	1
3.2 Important workflows . . . . .	22
3.3 Software solutions . . . . .	23
3.4 Conclusion . . . . .	24
<b>Bibliography</b>	<b>27</b>
.1 APPENDIX. References for all works . . . . .	32

# Introduction

Computer science world and research which involves computation-based experiments have complex processes and interactions of embedded and real-time systems, compilers, networking, and operating systems. Unfortunately, with growing system complexity, it is getting more difficult to get the same results and conclusions which are provided by original scientific work which is leading to decreasing quality of research. That is why research world came with Re\* research concept. *Re\* research* refers to processes with outputs and documentation which ensure *repeatable, reproducible, recomputable, reusable* or *replicable* results. In the scientific community, there is increasing concern about these characteristics, as evidenced by recent reports from ACM, Dagstuhl, and other players. The goal of this work is to consolidate and structure the knowledge around Re\* research as well as the reflection and reception in the scientific community, including in conference and journal and submission requirements. All these best practices will lead to general paper and research quality improving.

## 1.1 Motivation

Publications are one of the most critical and essential foundations of academic life. At the same time Jan Vitek[1] tells that computational disciplines have much more advantages in comparison with other fields of science since in most cases conferences plays the role of a filter for selecting the best scientific works. Unfortunately, due to the short publication time and the one-stage process of reviewing papers, the conferences had a profound impact on how the procedures in research are now being carried out. To remain competitive due to time constraints and unregulated standards of reproducibility of scientific works, most researchers had to de-

crease the level of research quality and not spend the majority of time to prepare, provide and publish code and documentation. Also, due to the increasing complexity of scientific experiments with the rapid growth of technical progress, as well as the fact that the main criteria for most conferences in the evaluation of scientific papers are the level of novelty - most Re\*research concepts are increasingly ignored during publication writing.

In our days for computational-based experiments, all the processes that are performed while writing scientific work can be characterized as follows:

The researcher identifies the problem that has not yet been solved or solved non-effectively. In the future, the software is written, which conducts evaluation and tests of the task/problem. When programming is complete, initiated the process of preparation of the initial data that will be processed by the software. When the result data is ready, it is analyzed by researcher and proofs hypothesis or creates brand new ideas based on the data. This workflow is shown in the picture: 1.1

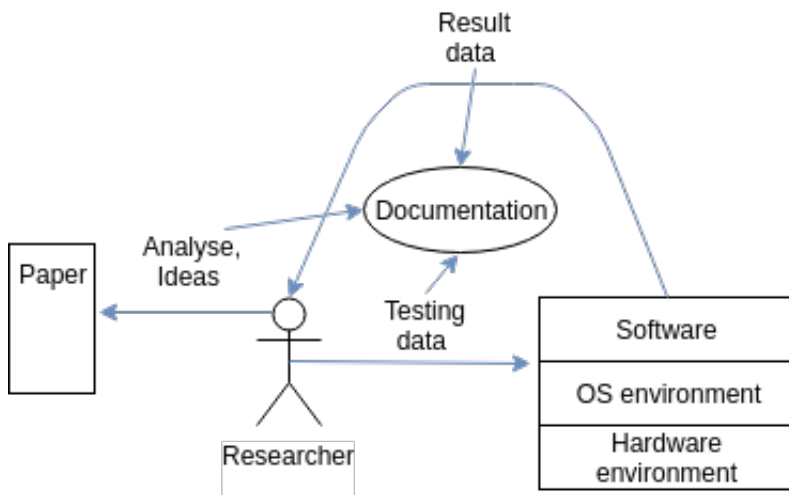


Figure 1.1: Research workflow[2]

By Jan Vitek[1] the essence of the scientific process in computer systems consists of:

- Identifying a problem and positing a hypothesis or model.
- Engineering a concrete implementation.
- Designing and conducting an experimental evaluation.

## 1.2 Problem statement

Experiments which includes computations, commonly represent complex ideas, and it is necessary to carry out a lot of tests on various of hardware and software platforms with several implementations. Typically, this is beyond the limitations caused by the amount of time and paper size. As a result, documents and scientific works that can be useful to the academic community are published without sufficient documentation, mention of restrictions and adequate instructions to reproduce the experiment and results.

Reasons above led to the fact that the academic community came to the idea of Re\* research. Re\* research refers to processes with outputs and documentation which ensure repeatable, reproducible, recomputable, reusable or replicable results. Instead, by Collberg[3] the symptoms of the current state of practice include:

- Unrepeatable results
- Unreproduced results
- Measuring the wrong thing
- Meaninglessly measuring the right thing.

Jan Vitek[1] defines more factors apart from pressure which are leading to bad quality from Re\* research point of view:

- Lack of benchmarks
- Lack of experimental methodology
- Lack of understanding of statistical methods

Tomas Kalibera in his work[4] "Repeatability, reproducibility in computer systems research" defines 7 "Deadly sins" in research which are making end results much useless :

- **"Unclear experimental goals."** First of all, it is necessary to formulate the purpose of the experiment and what kind of results and conclusions are expected in the end. Without clearly posed questions, the experiment has not much sense.

- **"Implicit assumptions or experimental methodology."** The documentation should have full information about the experimental setup and methodology that were used in conducting the experiment.
- **"Proprietary benchmarks and data sets."** For many researchers, as well as for industry, the availability of benchmarks and data sets used in the experiment could be a key point.
- **"Weak statistical analysis."** Even at high-level conferences like PLDI'11 [5], most of the works ( 39 of 42) did not mention about the division in a time of executed experiments.
- **"Measuring the wrong thing/right thing meaninglessly."** Often a situation occurs when complex and non-trivial experiments involve incorrect connections between various factors.
- **"Lack of a proper baseline."** Many researchers while writing a paper use own implemented baseline with differing levels of optimization. Therefore, it is essential at the beginning of the experiment to establish a reliable and credible baseline.
- **"Unrepresentative workloads."** One of the biggest dangers of writing a regular paper is the wrong load on the computer system. Processes that performs in parallel with the conducted experiment can have a fundamentally affect on the result and could not be reproduced in the future.

### 1.3 Methodology

To write this survey paper, scientific papers, articles, and posts were analyzed. Most of the time, data sources were found and taken from both DBLP[6] and Google Scholar[7]. In total, 75 scientific papers were analyzed, and for each of the characteristics of the Re\* research, 3-5 papers were selected. The analysis of the scientific material was carried out to compile a general picture and the relationship between the characteristics given above, each of which is central in a variety of scientific publications. All the selected scientific works were abstracted on the problem they cover, and year of publishing. The figure1.2 shows a graph with the number of publications where certain characteristic was a central problem. For making this analysis, the

full list of papers was used (see Appendix) and simple Python code which together with initial data can be found on GitHub[2].

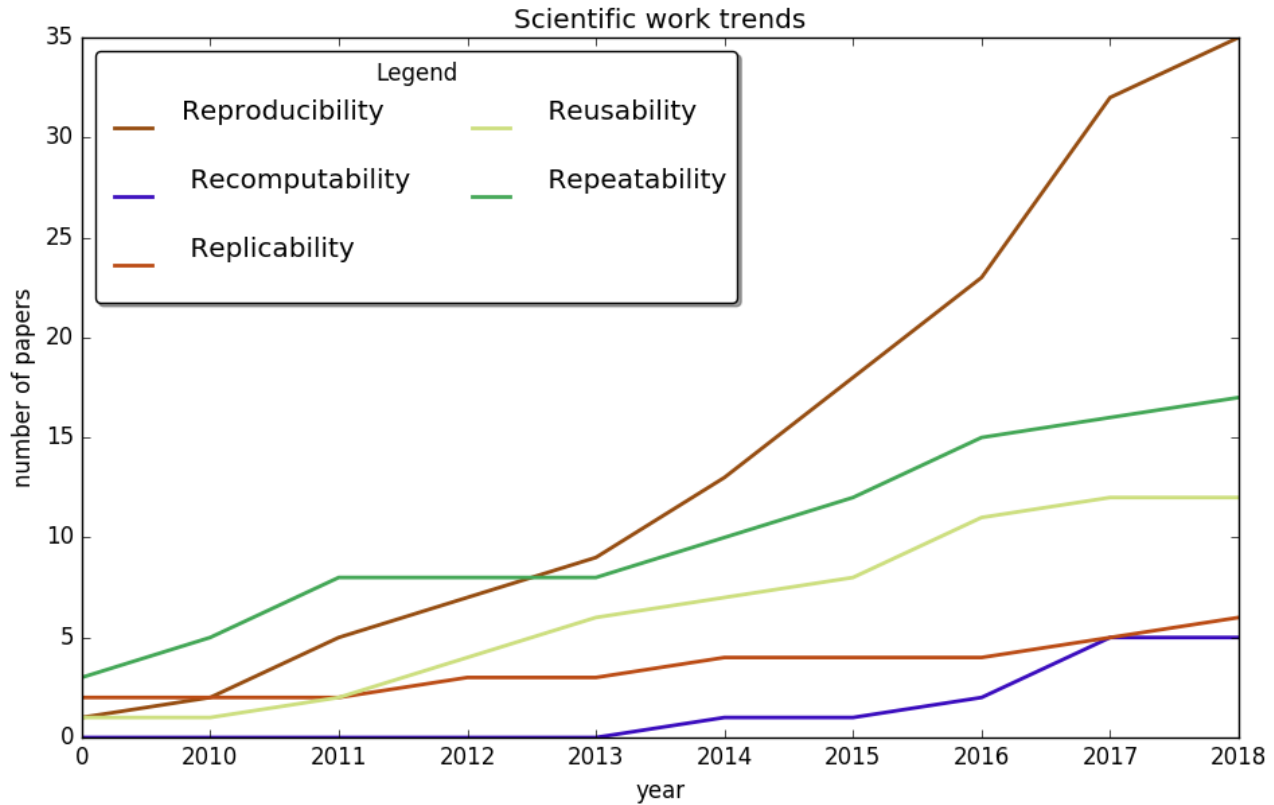


Figure 1.2: Paper analysis[2]

The primary requirements were the quality of the paper, documents with later dates had a privilege and, of course, papers were chosen which determined the research characteristics of Re \*research as a central topic. From the analyzed articles, the most prominent efforts were made to extract definitions, features, current solutions and generalize them to give a clear idea of these topics separately and to create a "general view" of the study. The "general view" implied a clear division and commonalities between each part of the Re\* study. As a result, the goal was also to make a set of simple requirements that will help make scientific work reproducible/repeatable/recomputable/reusable/replicable.



## **Re\* research characteristics**

### **2.1 Repeatability**

The repetition by Tomas Kalibera[4] definition is "the possibility of repeating the same experiment, simulation or scientific method in the same system and way and obtaining the same result." This requirement applies more to researchers to ensure stable results, but is also necessary for the scientific community in particular. Having the opportunity to repeat the result leads to new ideas. When the results are repetitive, it makes it easier for systems research. This characteristic is beneficial for scientific reviewers since those works/results that are repetitive and provide a code or web interface in the public domain can be tested for reliability. And also to evaluate the proposed method, the change in input data can be used. Especially this applies to those works for which no special hardware or software infrastructure is required.

Another author Collberg[3] defines the repeated study as " if it is possible to re-run the researchers experiment using the same method in the same environment and obtain the same results." In his opinion, it is necessary to keep results repeatable that reviewers, colleagues and other scientists could evaluate the outcome and methodology by full and accurate evidence. With this approach, a lot of repetitive work will be avoided; the researchers rely on the result and scientific progress will advance.

In this work, repeatability can be defined as an ability of scientific work to be repeated the various number of times and receive the entirely the same result of each experiment. The researcher needs to provide full information about his experimental set-up, software, codebase and initial datasets with instructions how to run them. This workflow is shown in the picture [2.1](#).

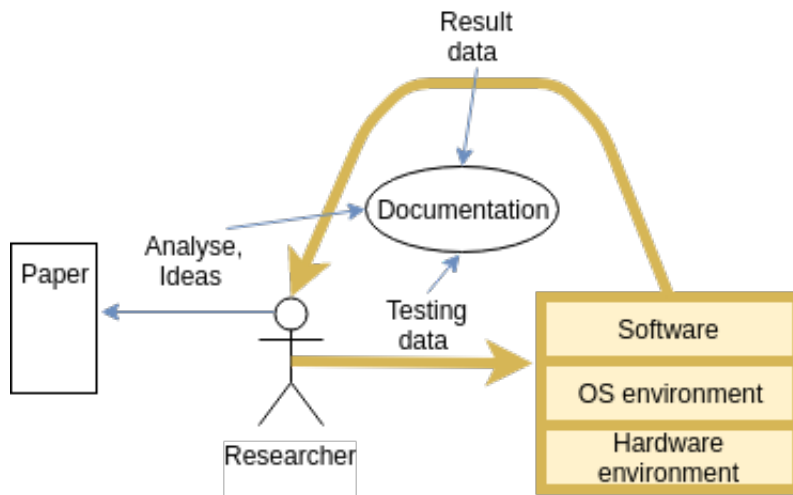


Figure 2.1: Repeatability[2]

### 2.1.1 Current state of art

The National Science Foundation Grant Proposal Guide[8] says, “Investigators and grantees are encouraged to share software and inventions created under the grant or otherwise make them or their products widely available and usable.” Typically, papers from industry have low repeatability because most of the code is under a private license, at the same time academic works, because they use open-source code, have a higher degree of repeatability.

### 2.1.2 Recommendations

- One of the most widely used practices to ensure scientific work is repeatable is the packaging of all source data, code, operating system, and third-party software into a **virtual image**. This practice will allow the project to be run by any researcher and solve problems with the versioning. The more it leads to own issues and limitations that need to be resolved. Like how to solve the problem with the launching of a virtual image that may have had an outdated operating system or applications that have identified the security vulnerabilities and at the moment they are not patched, and still, results might be different because of hardware setup, etc.
- **Fund repeatability engineering.** Special agencies should encourage researchers to request additional funds for “repeatability engineering,” including hiring programming staff

to document and maintain code, do release management, and assist other research groups wanting to repeat published experiments. In the same way funding agencies conduct financial audits to ensure costs claimed by grantees are allowed, they should also conduct random audits to ensure research artifacts are shared by what was promised in the grant application

- **Require sharing contract.** Publishers of conference proceedings and journals should require every article include a sharing contract specifying the level of repeatability to which its authors will commit. While the first point will have the effect of shifting some funding from pure research to engineering and oversight, both are important because they ensure research results continue to benefit the academic community— and the public funding it—past the project end date.
- **Sharing contracts.** The sharing contract should be provided by the authors when a paper is submitted for publication (allowing reviewers to consider the expected level of repeatability of the work), as well as in the published version (allowing readers to locate research artifacts). The contract commits the author to make certain available resources that were used in the research leading up to the paper and committing the reader/reviewer to take these resources into account when evaluating the contributions made by the paper

## 2.2 Reproducibility

Many research departments recently have been increasingly worried by the reproducibility of scientific papers, especially in computation-based experiments, where computers and computations are the central part of the experiment. To ensure performance is necessary to publish all the computational tools, documentation, datasets and benchmarks that were used by the researchers for this experiment. For any researcher who would like to reproduce results and experiments open source code, documentation, datasets should be available. This movement in the scientific world has a de facto title "Reproducible research" [9].

In the work of Tomas Kalibera[4], reproducibility is defined as: "Independent confirmation of a scientific hypothesis through reproduction by an independent researcher/lab is at the core

of the scientific method." Reproduction of scientific work should be done by independent researchers after publishing the paper with available datasets, code, experimental setup, etc. Required data can also be requested from the author. It is also possible to say that in the case of error or deception of work reproducibility is better to use compared to repeatability since it provides a higher level of abstraction.

Comparing repeatability and reproducibility, another author Colberg[3] notes that reproducibility does not require an identical experimental setup with the same hardware. Instead, it is an independent confirmation of the scientific hypothesis/results that are performed after the publication by other research units conducting the same experiments with the same or similar data periodically changing them that come to the similar or the same conclusions. It can be concluded that: Repeatability and reproducibility are characteristics of the scientific workflow that is necessary to prevent the proliferation of erroneous results.

In this work, reproducibility can be defined as an ability of scientific work to be reproduced by any other research group not depending on hardware requirements. Compared to repeatability, results should not be identical but rather conclusions and data trends should be confirmed. A researcher needs to provide full information about codebase and initial datasets with instructions how to run them. Although an alternative solution that also confirms the conclusions can be considered reproducible. This workflow is shown in the picture 2.2.

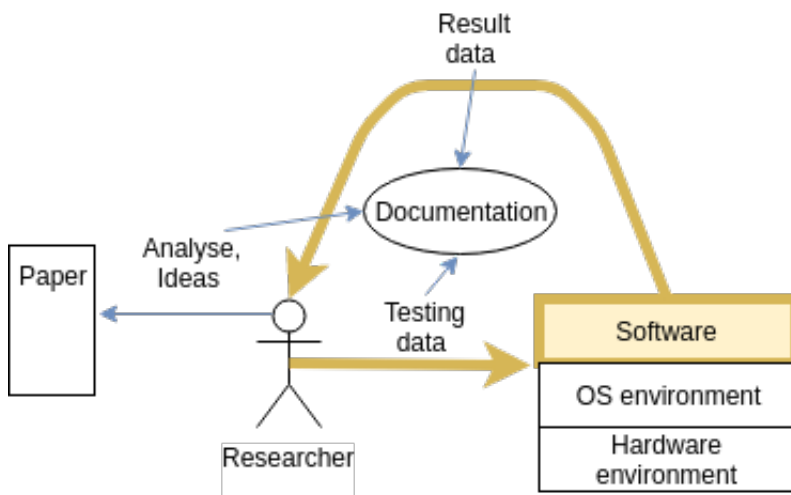


Figure 2.2: Reproducibility[2]

### 2.2.1 Current state of art

Since the problem of the reproducibility of scientific works is not new, many scientific journals like PLOS and Nature Methods with some governmental institutions such as NIH require authors to provide all the data obtained during the experiments. Data include documented steps of experiments and intermediate results, in order so that interested parties have access to this information, and there is an opportunity to promote similar research. Often, studies that reproduce already existing hypotheses are also published in journals whether they confirm whether or not the results of the original study. In the world of computer science, unfortunately, this does not happen, although there have been numerous attempts at the Reproducing research planet[10], and the Evaluation of Collaboratory[11]. The reproducible research planet is a website for scientists who must store research data including data sets, code, documentation, experimental setup, etc. There have been attempts to make this website part of a standard research process in several articles and journals. In turn, the authors thereby increased their chance to be cited and reproducibility in general. The Evaluate Collaboratory is in addition to organizing workshops on experimental evaluation of software and computer systems, initiated a petition to program committee chairs of conferences and workshops that called for acceptance of reproduction studies as first-class publications.

### 2.2.2 Recommendations

- **Develop open source benchmarks.** Required benchmarks, well documented, open source and carefully evaluated benchmarks should be fully accepted as first-class contributions for publications at leading conferences and journals and should be supported.
- **Codify best practice documentation, methodologies and reporting standards.** It is necessary to harmonize the minimum standards for documenting experiments and reporting. This can be done through community efforts and target working groups. It is also required to understand the factors that affect the measurable properties of computer systems, and better understand their statistical properties.
- **Require repeatability of published results.** Repeatability should be part of the publica-

tion review process. Based on the document and additional material on the experiments (documentation, configuration, source code, input sets, scripts), reviewers should make sure that the experiments are repeatable.

## 2.3 Recomputability

In work "The Recomputation Manifesto"[12] Gent indicates that in the scientific development, replication is of great importance, but usually, in the computer science world, replication does not take much attention and researchers do not repeat the same experiment over and over again. And those experiments that are rarely implemented do this in full. The manifesto proposes to make the source code more accessible, although this is a small step. Usually, even when the source code can be launched this does not mean that the experiment can be repeated. In this paper, it is proposed that the discipline of computer science include the replication of experiments as a standard practice and also that tools and repositories are available..

Some of the authors[13] define recomputability the same as reproducibility as the ability to achieve the same results on the same data regardless of the computing environment or when the analysis is performed.

In his paper[14], Wehrle notes that research is currently associated with highly specialized software and complex data processing workflows. While technology accelerates scientific progress, this acceleration also risks the sustainability of progress. Since almost every piece of information is created, transmitted and stored digitally, and the information is not necessarily associated with single, easily identifiable digital objects, but can be embedded in complex environments and processes, which allows to check, replicate and reuse data and accompanying methods not guaranteed, especially in the long run.

Matthias Kricke[15] provides one of the most accurate definitions of replication, this "ability to recompile results from raw data at any time" is essential for these studies to ensure data stability and selectively include new data in the already supplied data product. The competence of these products allows the client to restore earlier versions of data, reports and analysis results to compare them with new ones.

In this work, recomputability can be defined as an ability of scientific work to be recomputed

any point in time with any data and by anyone who would like to do it. A researcher needs to provide full information about codebase and make it publicly available. Compare to reproducibility and repeatability the goal is not to get the same results or conclusions, but instead, other people can reuse codebase for any purposes. This workflow is shown in a picture 2.3.

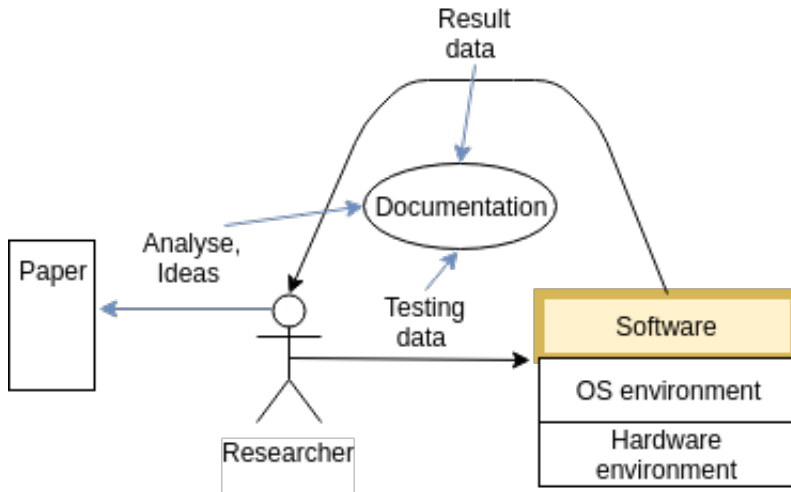


Figure 2.3: Recomputability[2]

### 2.3.1 Current state of art

Minimum information standards for recomputable experiments have been discussed and publically available for some years (e.g., MIAME[16], MIABi[17]), although much recent discussion on the practical implementations of recomputability frequently takes place on the Internet and in blogs rather than in peer-reviewed literature[18][19][20].

### 2.3.2 Recommendations

- Each publication presenting computational results should be aimed at creating a "virtual reference environment" accompanying the publication. This reference environment should be the minimum implementation of the software stack needed to reproduce some or all of the computational part of the results. For a desktop tool, this will be the operating system, libraries, tools, and data. By downloading this image and placing it on a virtual machine monitor, such as VMWare [21] or VirtualBox [22], readers and reviewers

can immediately reproduce and examine the results with minimal setup effort. Many researchers now also have access to institutional or national "cloud" environments, and a virtual reference environment can easily be applied to one of them.

- Using open and free software, such a reference medium is now technically simple for production and distribution; it does not require specialized knowledge in the field of software development or system administration beyond the skills set of a typical bioinformatics researcher. If the result is operating systems and tools with open source code or software, there are step-by-step tools for creating a snapshot of the working environment (for example, Ubuntu Builder [\[23\]](#) or Relinux).
- Using already existing tools like Popper[\[24\]](#), which goal is to implement executable papers in today's cloud-computing world by treating an article as an open source software (OSS) project. Popper is realized in the form of a convention for systematically implementing the different stages of the experimentation process following a DevOps approach.

## 2.4 Reusability

One of the complete works on the scientific works reusability belongs to Costantino Thanos[\[25\]](#) in his the term re-use of data means the ease of using data collected for one purpose, to study a new problem. This term implies re-use of separate data sets in the context of different contexts. Re-use of data becomes a unique feature of modern scientific practice, as it allows for the re-examination of evidence, reproduction, and verification of results, minimizing duplication of effort and relying on the work of others.

In another paper, Gabriel Fils[\[26\]](#) say that the minimal use case for sharing a computational experiment (in the form of a joint research object) involves repeating its initial implementation and verifying its results. However, to use its potential, it must support modified reuse. The objective of research should be created and stored not as a simple aggregate of digital content, but in an efficiently computable form: as a reusable object.

In this work, reusability can be defined as an ability of scientific work to be reused any point of time with any data and by anyone who would like to do it and ultimately everything should



be documented starting from the progress, small conclusions, ideas, methods, and benchmarks choice. A researcher needs to provide full information about codebase, datasets, and documentation. The main idea that other researcher can reuse the data to make brand new conclusions/papers. This workflow is shown in the picture 2.4.

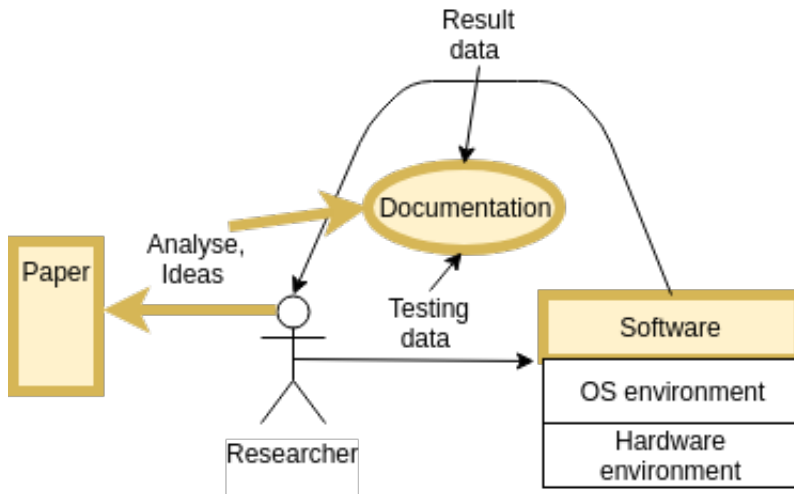


Figure 2.4: Reusability[2]

### 2.4.1 Current state of art

The ability to reuse data can be efficiently implemented within Open Science since the ultimate goal of Open Science is to make research data public and reusable. The European Commission is moving decisively towards the implementation of the "Open Science" concept in Europe: in 2015, the European Commission called on all member states of the European Union to ensure that states publish public research results to improve science and strengthen their knowledge-based[27]. The recent "Amsterdam Call to Action on Open Sciences" advocates "full open access for all scientific publications" and endorses an environment in which data sharing and management are the default method for all publicly funded research. This document was prepared as an open scientific conference organized by the Dutch Presidency of the Council of the European Union (April 4-5, 2016) [28]. Another initiative of the European Commission, which is worth mentioning, is the publication of the FAIR Data Management Guide in Horizon 2020, that is, a set of guidelines for creating data accessible, accessible, compatible and reusable [29]. The same as SNSF in "FAIR data principles"[30] defines a range of qualities a published

dataset should have in order to be Findable, Accessible, Interoperable and Reusable. Code, data, scenarios and the results of a temporary experiment - provide the means for sharing knowledge about computational analyses.

### 2.4.2 Recommendations

There are plenty of practices which can be applied to make proper implementation of the Data Publication procedures and, thus, to overcome the impediments to data reuse. Thanos[25] defines most important recommendations:

- **(Meta) Data Modeling.** In order to facilitate data understandability, it is necessary to define and develop formal models that adequately describe:
  - data representation needs of a given scientific discipline;
  - data provenance information;
  - data contextual information;
  - data uncertainty;
  - data quality information.

All this information is collectively called metadata information. If scientists need to reuse data collected by others, then the data must be carefully documented. Metadata is descriptive information about the data that explains the measured attributes, their names, units, accuracy, accuracy, a location of the data and, ideally, much more. Recently, a mechanism was proposed, a data document capable of improving the understanding of data and, therefore, the reuse of data. A data document can be defined as a scientific publication of a metadata document with a search capability describing a specific on-line available dataset or a group of datasets published by standard academic methods[31].

- **Domain-Specific Ontologies Ontologies.** A key technology that allows a wide range of data transmission services[32]. The increasing availability of data has shifted focus from closed applications with relatively low data levels, to mechanisms and applications for searching, integrating and using the vast amounts of data that are now available. Ontologies provide a semantic basis that allows you to reuse research data[33][34].

- **Data Discovering.** Under the discovery of data, meaning the ability to quickly and accurately identify and find data that supports research requirements. The process of detecting data existing in the data collection/database is supported by a search and query function that uses the capabilities of registering and citing data; and metadata descriptions contained in categorization/classification schemes, data dictionaries, and metadata registries.
- **Data Exchangeability.** By data exchange, we mean the ability of two objects, that is, the author of data and the data user, to exchange meaningful data sets. Data exchange is a prerequisite for the reuse of data. During the data exchange process, especially when data moves between scientific disciplines, it is necessary to solve three types of "heterogeneity".
- **Linking Data to Publications.** In science with a predominance of data, scientific communication undergoes significant changes. Modern scientific communication should support the practice of providing a link to data in the same way that researchers usually provide a bibliographic reference to printed resources. The need for quoting data is beginning to be recognized as one of the critical practices underlying the recognition of data as primary research products and not as a by-product of research

## 2.5 Replicability

The term replicability is not defined in computer science, and most of the time it refers to reproducibility. But for example in psychological scientific research Fournier[35] explains that the study should give the same results if they are repeated precisely. The replicability can be increased by an in-depth review of other similar experiments and exclude variables that you might not have thought of.

At the same time, Sinha[36] gives a different definition that replicability uses the same methods, but with different themes and experimenters, can also be reproduced someone else's work or results to establish its validity. For example, if someone executes experiment A and reaches conclusion B, and if someone does the same experiment A and come to a C output that is differ-

ent from B, then it will not replicate because the results will not be reproduced.

in the end, Fehr[37] gives quite a precise definition of replicability of the attribute describes the possibility of repeating experiments based on calculations and obtaining the same results. Sometimes this is an equivalent term. For replication, required documentation is required on how to run the software to get replicated results. Replicability, in turn, is the primary requirement of reliable software, as well as its results, as it demonstrates the absolute stability of the procedure against statistical influences and observer bias. Also, screening can serve as a benchmark against which new methods can be compared, as indicated in.

In this work, replicability can be defined as an ability of scientific work to be replicated by anyone. A researcher needs to provide detailed information and documentation on experiment workflow, codebase with documentation. The main idea that other researcher can get the same conclusions. This workflow is shown in picture 2.5.

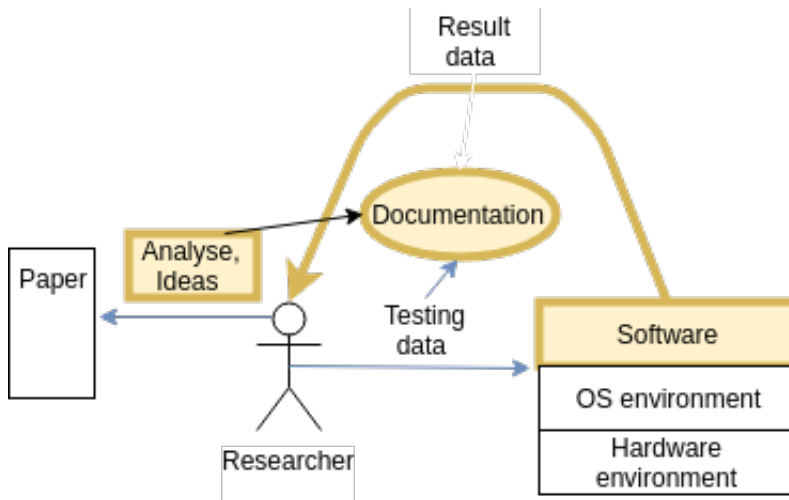


Figure 2.5: Replicability[2]

### 2.5.1 Current state of art

Concerns about reproducibility are confirmed by the observation that replication studies are rare and in some areas become less frequent with time. Various authors have established links between the rarity of replicas and the importance attached to innovation and original research, or to the perception that replications are not published in journals. The Nature study showed that a minority of researchers even attempted to publish replication, although, on a positive

note, 24% published successful replication, and 13% published a refusal to replicate the result. This is balanced approximately by the equivalent number (10%) of respondents who reported that they could not issue the error message[38]

### 2.5.2 Recommendations

- **Basic documentation.** A fundamental requirement for replicability is the abbreviated documentation, which includes instructions on how to generate an executable binary program in the case of a compiled language and a description of how to run the program to obtain replicated results. This documentation is crucial for the replication of the experiment because it determines the technical implementation and provides a possible repetition of the experiment. Often numerically calculated results are further processed to facilitate interpretation, for example, by visualization. The documentation on the evaluation of these results, descriptively or algorithmically, is necessary for replicating not only the results of calculations but also their evaluation.
- **Automation and testing.** The automation of the experiment makes it possible to quickly and reliably check the possibility of replicating experiments based on calculations. Usually, this means that one or more scenarios automatically prepare and run the experiment, and after processing the results. The reproducible behavior of all building blocks of the experiment, for which it is recommended to set up specific tests. Three categories of tests are usually considered: unittests, small sections of the source code exam; Integration tests, verification of the main components of the source code; and system tests, to evaluate the whole project. Tests usually include comparing a statistically significant sample with the analytical results, or a matching control problem[37].

## Summary

Overall most of the Re\* research components have similar ideas, approaches, and standard practices. In this part, all characteristics will be generalized, made clear differences and general recommendations which should satisfy all of the criteria.

First of all, a broad picture of all Re\* research characteristics together was made. The central definition is **recomputability** which refers to the code reusability. Recomputability require having the code and underlying documentation available online that everyone can have access and reuse computation approaches with any data. **Reproducibility** is inheriting the recomputability characteristic. The basic definition refers to the ability to have similar results with the same code and data sets so that the paper reviewer or researchers from other groups will be able to confirm results. The last characteristic which is aimed to give similar data results(data sets) is **repeatability**. It is required to have the full description of experimental setup that everyone can get the same results. Of course, it is not possible in real life to get the same results from repeating an experiment, and usually, this characteristic refers to the authors themselves so the experiment will be repeatedly done with the same setup and have in the end deviations adequately done. Another feature which inhering recomputability is **replicability**. The main idea of replicability has the same methods which were used to get the result data and codebase documented that no matter what kind of data was used by the initial author other research groups will reach similar conclusions of the particular problem. More complex characteristic is **reusability** which is by the end of the day means that all data can be reused by anyone else hand having thoroughly everything documented starting from the codebase and initial data and finishing with methods, benchmarks, and approaches. The main idea of reusability can reuse all of this components and make completely new ideas[3.1](#).

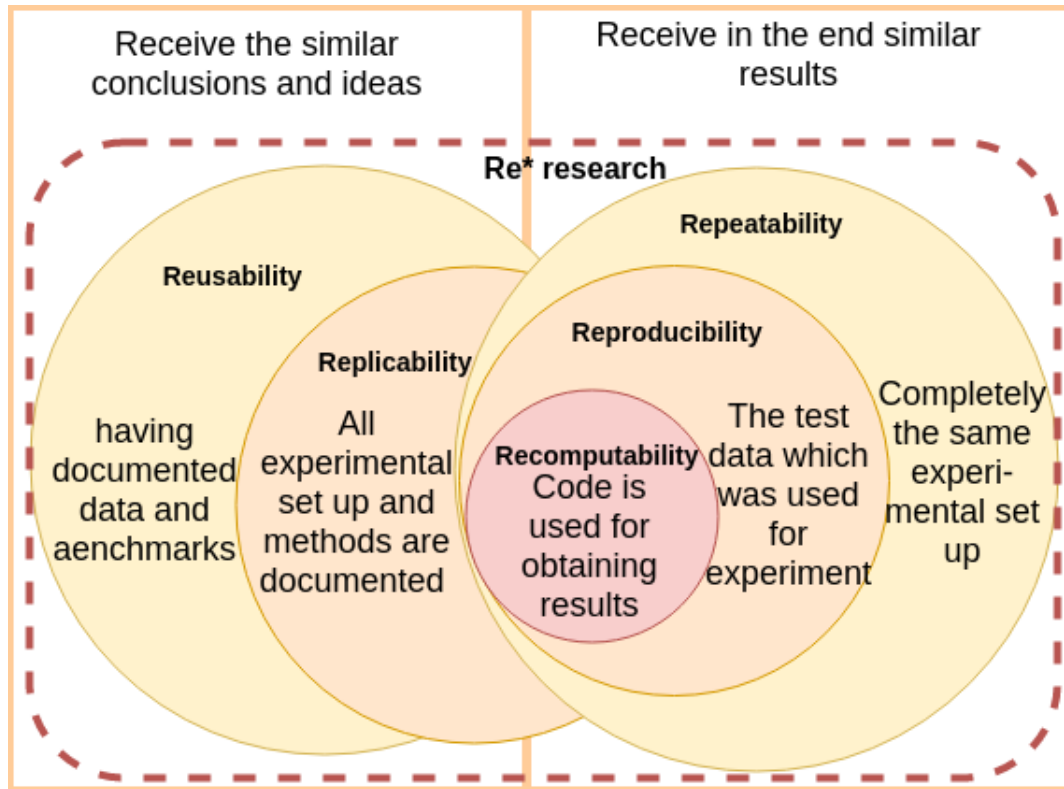


Figure 3.1: Relation diagram[2]

As can be seen, two characteristics for which includes all other aims, concepts and satisfied criteria are Repeatability and Reusability.

### 3.1 Good practices

- **Codebase.** computation-based experiments have the critical issue of having everything published and use as long as it possible only open source components. Everything should be documented and work out of the box. Using Docker image would be the good solution for this. Always leave links in the paper where the sources can be found.
- **Experimental set-up.** All hardware and software characteristics should be well described in the paper. For having similar software setup sharing the virtual machine image would solve most of the problems. It is more difficult for having the same hardware setup. To solve this problem all experiments can be done in a virtual machine, and the virtual machine image can be shared afterward.

- **Experiments.** Both datasets initial the same as result ones should be in open access. Also, all benchmarks and methods should be documented.
- **Documentation.** Complete workflow, processes, and problems should be fully documented and provided together with the paper.

## 3.2 Important workflows

- **Statistical methods.** Observable phenomena postulate hypotheses about and then measure. In system studies, the usual performance is the performance measured by the execution time. Some factors include many factors based on the architecture, operating system, compiler, and application. Some elements are controlled by the experimenter (i.e., architecture, compiler parameters or the time of the experiment). There are uncontrollable factors, some of which can be observed, and some of them can not. All independent elements should be randomized. Some parts of the experiment may be beyond the control of the experimenter. Real-time systems respond to real-world incentives, which are themselves random.
- **Documentation.** Proper archiving and documentation allow this even after a long time after the actual hardware or software infrastructure for repeating the experiment becomes unavailable. Indeed, this can also lead to negative results, such as the detection of errors in earlier experiments. The community must provide means of correcting published documents for these cases.
- **Benchmarks.** Experimental studies in systems are based on control applications. The test is a factor in the experiment, like everything else. When the experiment is started only in one of the stages, the value of interest is measured just for this test. If the standard is an application that the user wants to run, then this is normal (modulo). Otherwise, to make more general conclusions, the control factor should also be randomized: work with a variety of different criteria that are representative of real applications and statistically summarize the results.



### 3.3 Software solutions

There are plenty of solutions made to solve Re\* research problem not only in computation-based works, this software solutions are evaluated and compared on table 3.1:

- **PopperCI**[24]: Continuous Integration Service (CI), located in UC Santa Cruz, which allows researchers to automate end-to-end execution and verification of experiments. PopperCI suggests that tests follow Popper, a convention for implementing operations and writing articles after the DevOps approach, which recently proposed. PopperCI launches trials on cloud infrastructures with state, private or public funding in a fully automated way.
- **ARTENOLIS**[39]: The automated environment for reproducibility and testing for licensed software. The software is a universal software application for the infrastructure that implements continuous integration for open source software with licensed dependencies. It uses the master-slave structure, tests the code on several operating systems and several versions of the dependencies of the licensed software. ARTENOLIS provides stability, integrity and cross-platform compatibility of the codebase in the COBRA Toolbox and related tools.
- **ReproZip**[40]: Computational reproducibility using Ease.ReproZip is the recommended packaging tool for reviewing the reproducibility of SIGMOD. ReproZip was designed to simplify the process of creating an actual computational experiment, reproduced on different platforms, even when the experiment combines without repeatability. The tool creates a stand-alone package for the experiment, automatically monitoring and determining all the dependencies it needs. The researcher can share the package with others.
- **COLIBRI**[41]: Studio, which supports researchers in creating Case-based Reasoning (CBR) systems through documented views with varying degrees of abstraction. These workflows - mandatory templates - can be passed on to the community to advance their future reference and reproducibility.
- **CARE**[42]: A comprehensive archiver for reproduced performance in Linux. CARE works

in userland does not require configuration, and performs one task: creating an archive containing the selected executable files and files. To reproduce the final results of this initial launch is enough to unpack the archive, which has all the necessary tools for re-execution in a limited environment.

- **SUSHI**[43]: A defined workflow for a fully documented, reproducible and reusable analysis of NGS data. SUSHI is a flexible data analysis system that eliminates bioinformatics from the administrative problems of their data analysis. SUSHI allows users to create reproducible workflows of data analysis from individual applications and manage input data, parameters, meta information with user-controlled semantics, and work scripts. As distinctive features, SUSHI provides an expert command line interface, as well as a user-friendly web interface for launching bioinformatics tools. SUSHI datasets are self-contained and self-documented in the file system.
- **Dugong**[44]: It is Docker image which based on Ubuntu Linux focused on the reproducibility and reproducibility for the analysis of bioinformatics. Image Docker is base on Ubuntu 16.04, which automates the installation of more than 3500 bioinformatics tools (along with their respective libraries and dependencies) in alternative computing environments. The software runs through the user-friendly graphical interface XFCE4, which allows the user to manage and install the Jupyter Notebook to help in the delivery and exchange of serial and reproducible protocols and results in the labs, improving in the development of open science projects.
- **S2P**[45]: A software tool for fast implementation of reproducible biomedical research projects. The software provides various functionalities for the process of identifying data based on 2D-gel proteins and MALDI-mass spectrometry with an automated reproduced manner.

### 3.4 Conclusion

Recently computation-based experiments start getting more and more complex regarding setup, processes, methods, and interactions. With growing complexity of tests, it has become

more difficult to reproduce results from the author. Not reproducible results are not convenient in a scientific world that is why a concept of Re\* research became so popular recently. Re\*research consist of repeatability, reprehensibility, recomputability, reusability, and replicability. Most of the authors who were trying to solve this problem were generalizing it as term reproducibility, but in fact, they were always talking about Re research.

In this work, we have defined all of this characteristics separately with their definitions, features, current state of problems and recommendations. Also, we made a relational diagram where all of them were placed. The fact that research is repeatable and reusable typically mean that it satisfied all Re\*research requirements. Most of the recommendations can be said in simple words document and upload everything. This problem is fundamental now and a lot of pieces software in different areas of research trying and solving most of the challenges.

	PoperCI	ARTENOLIS	ReproZip	COLIBRI	CARE	SUSHI	Dugong	S2P
Last up-date	Jul 20, 2017	Jan 23, 2018	Mar 31, 2018	Jan 21, 2015	Feb 6, 2018	Apr 14, 2018	Jan, 7, 2017	Dec 7, 2017
Area to use	Computation-based experiments	Computation-based experiments	Reproducible documents storage	Case-based reasoning research	Reproducible documents storage	Reproducible documents storage	Scientific computational analysis for biomedical research	Bio-medical research
Architecture approach and proposed solution	Runs experiments on cloud infrastructure in a fully automated way.	Uses a master-slave framework, tests code on multiple operating systems.	Tracks operating system calls and creates a package that contains all the binaries, files and dependencies.	Built on top of the jCOLIBRI framework and enables the composition of its CBR components.	Monitors the execution of the specified command to create an *archive* that contains all the material required to *re-execute* it in the same context.	Takes care of all aspects of documentation: Input data, parameters, software tools and versions are stored persistently.	Automates installation of more than 3500 bioinformatics tools.	Core component contains a GUI application based on AIBench framework.
Language or environment	Python	Jenkins	Python	jCOLIBRI	C++	Ruby	Docker	Java

Table 3.1: Comparison table of Re\*research existing solutions

## Bibliography

- [1] Jan Vitek and Tomas Kalibera. Repeatability, reproducibility, and rigor in systems research. In *Proceedings of the 11th International Conference on Embedded Software, EMSOFT 2011, part of the Seventh Embedded Systems Week, ESWeek 2011, Taipei, Taiwan, October 9-14, 2011*, pages 33–38, 2011.
- [2] Oleksii Serhiienko. EVA1-statistics. <https://github.com/lexxito/EVA1-statistics>, 03 2018.
- [3] Christian S. Collberg and Todd A. Proebsting. Repeatability in computer systems research. *Commun. ACM*, 59(3):62–69, 2016.
- [4] Jan Vitek. Repeatability, reproducibility and rigor in CS research. In *Proceedings of the Programming Languages Mentoring Workshop, PLMW@POPL 2015, Mumbai, India, January 14, 2015*, page 12:1, 2015.
- [5] Mary W. Hall and David A. Padua, editors. *Proceedings of the 32nd ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2011, San Jose, CA, USA, June 4-8, 2011*. ACM, 2011.
- [6] University of Trier. DBLP. Digital Bibliography Library Project. <http://dblp.uni-trier.de/>, 03 2018.
- [7] Google LLC. Google scholar. <https://scholar.google.com>, 03 2018.
- [8] NSF. The National Science Foundation Grant Proposal Guide. [http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg\\_index.jsp](http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_index.jsp), 12 2016.

- [9] Chris Drummond. Reproducible research: a minority opinion. *J. Exp. Theor. Artif. Intell.*, 30(1):1–11, 2018.
- [10] Reproducible Research Planet Blog. Reproducible research planet. <http://www.rrplanet.com/>, 03 2015.
- [11] Faculty of Informatics: USI. Evaluate collaboratory: Experimental evaluation of software and systems in computer science. <http://evaluate.inf.usi.ch/>, 03 2018.
- [12] Ian P. Gent. The recomputation manifesto. *CoRR*, abs/1304.3674, 2013.
- [13] Gaurav Kaushik, Sinisa Ivkovic, Janko Simonovic, Nebojsa Tijanic, Brandi Davis-Dusenbery, and Deniz Kural. Rabix: An open-source workflow executor supporting re-computability and interoperability of workflow descriptions. In *Biocomputing 2017: Proceedings of the Pacific Symposium, Kohala Coast, Hawaii, USA, January 3-7, 2017*, pages 154–165, 2017.
- [14] Dennis Wehrle, Thomas Liebetraut, Isgandar Valizada, and Klaus Rechert. Emulation-as-a-service - workflows and infrastructure to support recomputable science. In *Proceedings of the 7th IEEE/ACM International Conference on Utility and Cloud Computing, UCC 2014, London, United Kingdom, December 8-11, 2014*, pages 962–967, 2014.
- [15] Matthias Kricke, Martin Grimmer, and Michael Schmeißer. Preserving recomputability of results from big data transformation workflows - depending on external systems and human interactions. *Datenbank-Spektrum*, 17(3):245–253, 2017.
- [16] Brazma A, Hingamp P, and Quackenbush J. Minimum information about a microarray experiment MIAME—toward standards for microarray data. *Nat Genet*, 29:365–71, 2001.
- [17] Tan T, Tong JC, and Khan AM. Advancing standards for bioinformatics activities: persistence, reproducibility, disambiguation and minimum information about a bioinformatics investigation MIABi. *BMC Genomics*, 11:S27, 2010.
- [18] C. Titus Brown. Living in an ivory basement. <http://ivory.idyll.org>, 03 2018.

- [19] University of St. Andrews. Recomputation and best practices. <http://recomputation.org>, 03 2018.
- [20] Michael Barton. Bioinformatics. <http://www.bioinformaticszen.com/>, 2016.
- [21] Dell Technologies. Vmware. <http://www.vmware.com>, 03 2018.
- [22] Oracle Corporation. Virtualbox. <https://www.virtualbox.org>, 03 2018.
- [23] Canonical Ltd.
- [24] Ivo Jimenez, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau, Jay F. Lofstead, Carlos Maltzahn, Kathryn Mohror, and Robert Ricci. PopperCI: Automated reproducibility validation. In *2017 IEEE Conference on Computer Communications Workshops, INFOCOM Workshops, Atlanta, GA, USA, May 1-4, 2017*, pages 450–455, 2017.
- [25] Costantino Thanos. Research data reusability: Conceptual foundations, barriers and enabling technologies. *Publications*, 5(1):2, 2017.
- [26] Dai Hai Ton That, Gabriel Fils, Zhihao Yuan, and Tanu Malik. Sciunits: Reusable research objects. *CoRR*, abs/1707.05731, 2017.
- [27] European Commission. Commission Recommendation on Access to and Preservation of Scientific Information. [https://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/recommendation-access-and-preservation-scientific-information\\_en.pdf](https://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf), 2012.
- [28] The Netherlands EU Presidency. Amsterdam Call for Action on Open Science. <https://english.eu2016.nl/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science>, 2016.
- [29] European Commission. Directorate-General for Research Innovation. H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020. [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf), 2016.

- [30] Fond national suisse FNS. Explanation of the fair data principles. [http://www.snf.ch/SiteCollectionDocuments/FAIR\\_principles\\_translation\\_SNSF\\_logo.pdf](http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf), 03 2018.
- [31] Vishwas Chavan and Lyubomir Penev. The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, 12(15):S2, Dec 2011.
- [32] Meghyn Bienvenu, Balder ten Cate, Carsten Lutz, and Frank Wolter. Ontology-based data access: A study through disjunctive datalog, csp, and MMSNP. *ACM Trans. Database Syst.*, 39(4):33:1–33:44, 2014.
- [33] Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.*, 43(5-6):907–928, December 1995.
- [34] Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. *J. Data Semantics*, 10:133–173, 2008.
- [35] Fournier G. Replicability. psych central. <https://psychcentral.com/encyclopedia/replicability/>, 03 2018.
- [36] Sharad Sinha. Replicability research. <https://www.quora.com/What-is-replicability-in-research-How-does-it-differ-from-plagiarism>, 03 2018.
- [37] Jörg Fehr, Jan Heiland, Christian Himpe, and Jens Saak. Best practices for replicability, reproducibility and reusability of computer-based experiments exemplified by model reduction software. *CoRR*, abs/1607.01191, 2016.
- [38] Alireza Ahadi, Arto Hellas, Petri Ihantola, Ari Korhonen, and Andrew Petersen. Replication in computing education research: researcher attitudes and experiences. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research, Koli, Finland, November 24-27, 2016*, pages 2–11, 2016.
- [39] Laurent Heirendt, Sylvain Arreckx, Christophe Trefois, Yohan Yarosz, Maharshi Vyas, Venkata P. Satagopam, Reinhard Schneider, Ines Thiele, and Ronan M. T. Fleming.



- ARTENOLIS: automated reproducibility and testing environment for licensed software. *CoRR*, abs/1712.05236, 2017.
- [40] Fernando Chirigati, Rémi Rampin, Dennis E. Shasha, and Juliana Freire. Reprozip: Computational reproducibility with ease. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, pages 2085–2088, 2016.
- [41] Juan A. Recio-García, Belén Díaz-Agudo, and Pedro A. González-Calero. The COLIBRI open platform for the reproducibility of CBR applications. In *Case-Based Reasoning Research and Development - 21st International Conference, ICCBR 2013, Saratoga Springs, NY, USA, July 8-11, 2013. Proceedings*, pages 255–269, 2013.
- [42] Yves Janin, Cédric Vincent, and Rémi Duraffort. Care, the comprehensive archiver for reproducible execution. In *Proceedings of the 1st ACM SIGPLAN Workshop on Reproducible Research Methodologies and New Publication Models in Computer Engineering, TRUST 2014, Edinburgh, United Kingdom, June 9-11, 2014*, pages 1:1–1:7, 2014.
- [43] Masaomi Hatakeyama, Lennart Opitz, Giancarlo Russo, Weihong Qi, Ralph Schlapbach, and Hubert Rehrauer. SUSHI: an exquisite recipe for fully documented, reproducible and reusable NGS data analysis. *BMC Bioinformatics*, 17:228, 2016.
- [44] Fabiano B. Menegidio, Daniela L. Jabes, Regina Costa de Oliveira, and Luiz R. Nunes. Dugong: a docker image, based on Ubuntu Linux, focused on reproducibility and replicability for bioinformatics analyses. *Bioinformatics*, 34(3):514–515, 2018.
- [45] Hugo López-Fernández, Jose E. Araújo, Susana Jorge, Daniel Glez-Peña, Miguel Reboiro-Jato, Hugo Miguel Santos, Florentino Fdez-Riverola, and José L. Capelo. S2P: A software tool to quickly carry out reproducible biomedical research projects involving 2d-gel and MALDI-TOF MS protein data. *Computer Methods and Programs in Biomedicine*, 155:1–9, 2018.

**.1 APPENDIX. References for all works**