© ARTVILLE, LLC.

Reproducible Research in Signal Processing

[What, why, and how]

[Patrick Vandewalle, Jelena Kovačević, and Martin Vetterli]

Have you ever tried to reproduce the results presented in a research paper? For many of our current publications, this would unfortunately be a challenging task. For a computational algorithm, details such as the exact data set, initialization or termination procedures, and precise parameter values are often omitted in the publication for various reasons, such as a lack of space, a lack of self-discipline, or an apparent lack of interest to the readers, to name a few. This makes it difficult, if not impossible, for someone else to obtain the same results. In our experience, it is often even worse as even we are not always able to reproduce our own experiments, making it difficult to answer questions from colleagues about details. Following are some examples of e-mails we have received:

- "I just read your paper X. It is very completely described, however I am confused by Y. Could you provide the implementation code to me for reference if possible?"

- "Hi! I am also working on a project related to X. I have implemented your algorithm but cannot get the same results as described in your paper. Which values should I use for parameters Y and Z?"

Similar frustrations arose from time to time in our labs when changes had to be made to a figure for a revision of a paper or reuse in another work, and we are sure we are not alone.

To address the problem, we have started making our research reproducible. As described on Wikipedia, "reproducibility is one of the main principles of the scientific method and refers to the ability of a test or experiment to be accurately reproduced, or replicated, by someone else working independently" [1]. This is precisely the goal we are trying to reach with our reproducible research efforts. Instead of only describing the developed algorithms to "sufficient" precision in a paper, we give readers access to all the information (code, data, schemes, etc.) that was used to produce the presented results as first advocated by Knuth [2] and Claerbout [3]. In the words of Buckheit and Donoho [4]: "An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures."

We are convinced that making research reproducible is not only a matter of good practice, but also increases the impact of our publications. Recent studies have shown that publications are cited more often if they are available online in an open-access manner [5], or, if related research data is shared [6]. We believe that a similar argument holds for reproducible research papers, as they offer easier access to the research work. This is expected to increase the number of citations for such papers, as well as their use by colleagues in their own research, in commercial applications, and for teaching.

RESEARCH PRACTICES IN OTHER DOMAINS
Common research practices vary considerably from one research domain to another. Concerns about reproducibility are very old and can be found in many fields. They are addressed differently depending on the specific needs in each domain.

In theoretical sciences, such as mathematics, researchers have a long tradition of reproducing each other's results, proofs of theorems in this case. As early as about 500 years ago, mathematicians sent each other letters containing challenges to reproduce proofs of their results. Communication about research was then limited to personal communication between scientists. A famous example of a nonreproducible proof was Fermat's last theorem. Fermat himself noted in 1637 that he had a marvelous proof for this theorem but did not have the space in the margin to write it out. He was reading a translation of Diophantos' *Arithmetica*, where a special case was given. He then famously wrote in the margin (originally in Latin)

It is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general, any

> **WE ARE CONVINCED THAT MAKING RESEARCH REPRODUCIBLE IS NOT ONLY A MATTER OF GOOD PRACTICE BUT ALSO INCREASES THE IMPACT OF OUR PUBLICATIONS.**

power higher than the second into two like powers. I have discovered a truly marvelous proof of this, which this margin is too narrow to contain.

It has taken mathematicians more than 350 years to reproduce a proof of this theorem. Once such a proof is verified to be correct (reproduced), it is added to the common-knowledge database. Proofs of new theorems can then build upon the statements of the previous ones.

In experimental sciences, such as physics, chemistry, biology, and life sciences, it is a very common practice to give detailed descriptions of experimental setups and boundary conditions in publications. In 1637, Descartes established the framework for the scientific method in his work "Discourse on Method," of which reproducibility is an essential component. Many journals, such as *Nature* and *Science*, require authors to make data sets and descriptions of materials and methods available online in recognized databases. Reviewers are asked to determine whether all of the supplementary material has been archived. Researchers in biology and life sciences have developed a tradition of reproducing and checking fellow researchers' experiments. A hypothesis is stated by a team of researchers, and an experiment is performed to test this hypothesis. Only after a different team of researchers has independently been able to check the validity of the hypothesis by reproducing the experiment, it is added to the common-knowledge database. In these fields of research, knowledge is built up on reproduced experiments, and the hypotheses and experiments are to an experimental researcher what the theorems and proofs are to the theoretical one. Checking the validity of the results by independent researchers helps detect fraud, a recent example of which is that of the stem-cell research scandal [7], [8]. In response to this and other cases of scientific misconduct, in 2002, the *Journal of Cell Biology* introduced systematic tests to detect manipulation of figures in publications submitted to the journal [9]. Other publishers followed, and adopted similar measures.

Computational sciences, such as computer science, statistics, many fields of engineering, and signal processing, started making their mark in the second half of last century. They form a hybrid between the theoretical and experimental sciences described above: theoretical developments, such as new theories and theorems, are combined with practical experiments and the development of new algorithms. While this can result in the combination of the best of both sides (rigorous proofs for the theorems and experiments that are properly set up and described), it sometimes results in a weak combination of the two. We often encounter less rigorous proofs than a mathematician would find acceptable and describe our experimental setups far less precisely than an experimental scientist would. While such statements do not generalize to all signal processing work, the broad spectrum

of practices suggests this may be a good time to discuss and decide what the minimum required level of reproducibility should be.

A BRIEF HISTORY OF REPRODUCIBLE RESEARCH
It is difficult to track reproducible research back to its origins, because concerns about the reproducibility of research results are probably as old as scientific research itself. The concept of reproducible research in computational sciences as we present it here is very similar in nature to the ideas on literate programming advocated by Knuth in 1984 [2]. Instead of talking about adding code to a publication, he presented literate programming as a better way to document code. In his own words: "Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do." The results are very similar to what we have in mind here: research results as a bundle, consisting of a publication describing the methods and results as well as the code and data used to produce these results.

To our knowledge, the term reproducible research in this context was first used by Claerbout and his colleagues in the early 1990s [3]. He introduced these practices in his Stanford Exploration Project (SEP) lab working on geophysics research. The approach is currently being further developed by Fomel and Hennenfent in the open-source project Madagascar [10]. The open-source software construction tool SCons (think of it as a new cross-platform version of Makefiles) is now used to generate a publication together with all its results using a single command.

The reproducible research ideas were then applied to wavelet research by Buckheit and Donoho at Stanford using MATLAB [4]. They developed the WaveLab toolbox to reproduce the results from a set of their publications.

From here, the reproducible research philosophy spread to other domains and drew attention in various fields such as econometrics [11], [12], neurophysiological data analysis [13], and epidemiology [14]. The discussion on the topic in the signal processing community started with the publication of an opinion article in *IEEE Signal Processing Magazine* by Barni and Perez-Gonzalez [15]. Subsequently, the authors of this article, together with Barni and Perez-Gonzalez, organized a special session on the topic at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2007. Vandewalle and Vetterli presented an overview of their experience with reproducible research [16]. The Madagascar project was described by Fomel and Hennenfent [10], while Kovačević, gave her point of view as a former editor-in-chief of *IEEE Transactions on Image Processing* [17]. Vandewalle et al. [18] discussed standardized data sets and public competitions for algorithm comparisons. Finally, two case studies of reproducible research were presented by Marziliano [19] and Barni et al. [20].

A special issue of *Computing in Science and Engineering* is also devoted to reproducible research [21], which further indicates this topic's relevance.

As a result of these discussions, the IEEE Signal Processing Society took steps to encourage reproducible research. On IEEE *Xplore*, it is now possible to add a hyperlink to the code and data to reproduce the results (or other additional material). The editor-in-chief of *IEEE Transactions on Signal Processing* now encourages authors to use this possibility in the e-mail sent to the authors of accepted papers.

> **IT IS DIFFICULT TO TRACK REPRODUCIBLE RESEARCH BACK TO ITS ORIGINS, BECAUSE CONCERNS ABOUT THE REPRODUCIBILITY OF RESEARCH RESULTS ARE PROBABLY AS OLD AS SCIENTIFIC RESEARCH ITSELF.**

WHEN IS RESEARCH REPRODUCIBLE, AND WHEN IS IT NOT?
We use the following definition of reproducible research.

DEFINITION: REPRODUCIBLE RESEARCH
A research work is called reproducible if all information relevant to the work, including, but not limited to, text, data and code, is made available, such that an independent researcher can reproduce the results.

The suggested format for publishing reproducible research is a Web page containing the following items:

1) title of the publication

2) authors (with links to the authors' Web sites)

3) abstract

4) full reference of the paper, with current publication status, and a portable document format (PDF) (depending on the journal's policy, this might be the PDF of the article itself, or the last preprint version)

5) all the code to reproduce the results, images and tables (code should be well documented, and a readme file should be added explaining how to execute it)

6) all the data (images, measurements, etc.) to reproduce the results, images, and tables (an additional readme file explains what the data represents)

7) list of configurations on which the code was tested (software version, platform)

8) a possibility to give comments and remarks (and to report bugs).

Depending on the research domain, the following (optional) information could be made available:

9) images with captions (so that people know what Figure X is about)

10) references (with abstracts and links to the sites where the full text papers can be downloaded).

We distinguish six degrees of reproducibility:

■ 5: The results can be easily reproduced by an independent researcher with at most 15 min of user effort, requiring only standard, freely available tools (C compiler, etc.).

■ 4: The results can be easily reproduced by an independent researcher with at most 15 minutes of user effort, requiring some proprietary source packages (MATLAB, etc.).

■ 3: The results can be reproduced by an independent researcher, requiring considerable effort.

■ 2: The results could be reproduced by an independent researcher, requiring extreme effort.

■ 1: The results cannot seem to be reproduced by an independent researcher.

■ 0: The results cannot be reproduced by an independent researcher.

REPRODUCIBLE RESEARCH STUDY

To assess the current reproducibility practices in signal processing, we decided to repeat the experiment from [17] on a larger scale. We performed a study on all of the 134 papers published in *IEEE Transactions on Image Processing* in 2004. We asked two or three reviewers per paper to check the reproducibility of a paper using a short list of questions. We split the questions into three main parts: the reproducibility of the 1) algorithm, 2) code, and 3) data.

1) Reproducibility of the algorithm

    a) Is the algorithm described in sufficient detail?

    b) Are exact parameter values given?

    c) Is there a block diagram?

    d) Is there a pseudocode?

    e) Are there proofs for all the theorems?

    f) Is the algorithm compared to other algorithms?

2) Reproducibility of the code

    a) Are implementation details (programming language, platform, compiler flags, etc.) given?

    b) Is the code available online?

3) Reproducibility of the data

    a) Is there an explanation of what the data represents?

    b) Is the size of the data set acceptable?

    c) Is the data set available online?

Additionally, we asked the reviewers for the uniform resource locators (URLs) where they found the article, code, and/or data, and their confidence in their review. Each of the questions [except for 3(b)] had to be scored using the values 0, 0.5, 1, and N/A (not applicable). For question 3(b), we considered the size of the data set acceptable if the number of items (typically the number of images) was above four (quite an arbitrary number).

The results of this study are summarized in Table 1 and are well aligned with the smaller-scale experiment from [17] (the authors will answer individual requests for more details). In most of the cases (84%), the algorithm and the data are described with sufficient details, and similarly for parameter values (71%). In about two out of three papers, the algorithm is compared to other algorithms. Only about one third of the papers give a block diagram or pseudocode, which is rather surprising, given that this is an easy way of giving an overview of the algorithm (57% of the papers give at least one of the two). The small part of proofs for theorems (27%) is largely due to the large number of papers where there are simply no theorems to prove (and where at least one of the reviewers marked that there was no proof, instead of marking N/A). The size of the data set on which results are shown in the paper is above four in only about half of the cases. Finally, code (9%) and data (33%) are only available online in a minority of the cases, with data being available more often thanks to the frequent use of standard image data sets, such as Lena. Remark, however, that several versions of many of those popular test images exist, which might therefore still introduce uncertainty about reproducibility. An issue with URLs, if they are mentioned in a paper, is their generally limited lifetime. Several reviewers reported URLs in a paper that had become invalid. One way of addressing this is the digital object identifier (DOI), a unique number that identifies the digital object (and not its location, like a URL does). The DOI can be seen as the equivalent of an international standard book number (ISBN) for digital documents.

Overall, we conclude that algorithms are generally well described and are in many cases compared to other algorithms. Implementation details and online code and data are only given in a very small number of cases. The links to code and data that we gathered in this study are available online (see http://www.reproducibleresearch.net). While the algorithm description is essential to scientific publishing, and the online availability of code and data may be considered by some as subordinate, it is undoubtedly useful when other researchers want to test an algorithm themselves or compare it to their work. Barni et al. [20] also noted that it is often only when one is really trying to implement a paper that missing details such as initialization or stopping criteria are noticed. In our study, the reviewers only gave their impressions about reproducibility of the papers; they were not asked to implement them entirely themselves. Our conclusion from the above is that while the situation is not desperate, the state of reproducibility of research in the signal processing community can be improved.

[TABLE 1] RESULTS OF REPRODUCIBILITY STUDY ON *IEEE TRANSACTIONS ON IMAGE PROCESSING* PAPERS PUBLISHED IN 2004. AVERAGE SCORES OVER THE 134 PAPERS ARE PRESENTED.

ALGORITHM						CODE		DATA		
DETAILS	PARAMETER VALUES	BLOCK DIAGRAM	PSEUDO-CODE	PROOFS	COMPARISON	IMPLEM. DETAILS	CODE AVAIL.	EXPLANATION OF DATA	SIZE DATA SET	DATA AVAIL.
0.84	0.71	0.37	0.33	0.27	0.64	0.12	0.09	0.83	0.47	0.33

FAILURES OF REPRODUCIBILITY

Reproducibility fails for various reasons. For example, in our study, some papers gave simulation results on standard images that are widely available online, with Lena being the most well-known one. As different versions of this image circulate on the Web, it is often difficult to know on which version the simulations were performed. For example, we found examples of a 512×512 color image with 24 b/pixel, a 256×256 color image compressed with Joint Photographic Experts Group (JPEG), and a 256×256 grayscale image. It is unclear which type of scaling, compression, or color-to-grayscale conversion was used to convert one image into the other. As this evidently has a nonnegligible effect on compression performance, for example, it compromises the reproducibility of the results (except if the specific file is made available).

Extreme examples of reproducibility failures are Fermat's proof of his last theorem, or the stem-cell research scandal mentioned earlier. Less extreme examples can be found in many papers: the entire reproducibility of a paper is rare. We take one of our own papers as an example [22]. The paper describes a new method for registration of partially aliased images and super-resolution reconstruction. The algorithm is correctly described in the paper, and a further developed version is described in a reproducible way in a later paper [23]. However, it is impossible to reproduce the results presented in the first paper. The exact shift values used in the simulation, as well as the stopping criterion of the reconstruction algorithm, are not given. Simulations are performed on a cropped grayscale version of the lighthouse image, and the experimental set of images is not available online. Even we as authors are not able to reproduce the results, as we have different versions of the code on our disk, and it is unclear with which version the images in the final version were created. Thus, when it comes to the difficulties of making research reproducible, we speak from experience.

HOW TO DO REPRODUCIBLE RESEARCH?

COMPONENTS

As described above, signal processing can be considered as a hybrid between theoretical and experimental sciences. It is therefore not surprising that we can distinguish both a theoretical and an experimental component to most of our research, as well as a part on the edge between those two: the development of new algorithms. Each of these parts has different requirements for making them reproducible, which we now discuss.

THEORY

Theoretical work is rather easy to make reproducible. As soon as a theorem is proven and the different steps are carefully described, any other researcher can reproduce the proof. Often, to develop an intuitive feel for the results, a theorem is complemented by numerical simulations.

EXAMPLE 1: SPECTRUM OF ULTRAWIDE BANDWIDTH SIGNALS

Ridolfi et al. developed a unifying formula to compute the power spectrum of the family of pulse modulated signals used in ultrawide bandwidth transmissions [24], [25]. In their publications, the authors give a detailed mathematical proof of their formula, using the point process formalism. Next to this proof, they also provide MATLAB code to compute such power spectra for various types of pulse modulations. The code is distributed under the GNU general public license (GPL) [26] and allows anyone to simulate an arbitrary ultrawide bandwidth transmission by specifying the parameters of the pulse modulation, the multiuser signature, and the errors affecting the transmission, such as the jitter or the losses. It is therefore possible to extensively test the formula and, with the help of a script, exactly reproduce the results provided by the authors. The code is well documented, and there is a detailed instruction manual that explains how to run the code. □

> **IDEALLY, AN INDEPENDENT RESEARCHER SHOULD BE ABLE TO REPRODUCE ALL THE RESULTS WITH A SIMPLE MOUSE CLICK.**

ALGORITHMS

At the core of many research projects in computational science is the development of new algorithms. With such work, it is important to give a careful description of the algorithm, including initialization procedures and specific parameter values [20]. A block diagram or a pseudocode description of the method is often helpful to increase clarity. Next to this detailed description, it is important to also add the code for the algorithm and, if possible, the data used to produce the presented results. Finally, the environment in which the algorithm was tested has to be specified. This can be a software tool, such as MATLAB, Octave, or Mathematica, or a compiler. If a compiler is used, it is important to note the additional compilation parameters and compiler flags that were used. While in theory, the code is supposed to be platform or hardware independent, it is useful to specify the platform on which the tests were performed. The tested version(s) of the software environment should also be precisely specified.

EXAMPLE 2: TONE MAPPING

Meylan et al. have developed a new algorithm for tone mapping of color filter array images, based on a model for retinal processing [27]. Next to the PDF of the paper, they also made the code available online that can be used to reproduce the results, as well as the figures and some additional results that did not fit in the paper. The code is written in MATLAB and is distributed under the GPL. The code was tested in MATLAB 7.0.1 on a Linux computer. □

EXPERIMENTS

Practical experiments are, by their nature, probably the most difficult type of work to reproduce. They often contain complex

setups, depend on various (environmental) parameters or random behavior, etc. An example is the SensorScope project described next.

Often, it is not even feasible to repeat the exact experiments, as results can in the best case only be statistically equivalent to the original results. Therefore, in this case, we opt for reusability rather than reproducibility of the data. This means posting full measured data sets online, along with sufficient documentation on the nature of the data set and instructions for use. Next to the measurement data, a detailed description of the experimental setup and measurement devices should still be given, for those interested in understanding the exact setup or in performing similar experiments.

EXAMPLE 3: SENSORSCOPE

The goal of the SensorScope project (see http://sensorscope. epfl.ch) is to gather a large amount of environmental measurements, which allows environmental scientists to study energy exchanges at the boundary between the earth and the atmosphere. This is achieved with a large-scale wireless sensor network. Each node in the network consists of a set of sensors measuring temperature, humidity, soil moisture, wind speed and direction, solar radiation, infrared temperature, and precipitation. The sensing platforms and their wireless connection interface have been specifically designed for this project, and all design specifications are available online. Deployments of such sensor networks have taken place in different locations: on the Ecole Polytechnique Fédérale de Lausanne (EPFL) campus as well as on the Plaine Morte glacier and several other locations throughout Switzerland.

As it is difficult (if not impossible) to reproduce such an experiment, the acquired data is also available online for other scientists. A database and a Web interface were developed to allow easy access to the data. The interface also contains clear documentation about the measurement conditions, such as precise sensor location, sensor specifications, and maintenance information. Such reusable data sets not only improve reproducibility of the work for our colleagues, but also provide a tremendous service to researchers in different domains (such as environmental sciences). It allows them to work with data they would otherwise never get. In return, we get valuable feedback about relevant measurements. □

CASE STUDY: SUPER-RESOLUTION

The first and third authors started the reproducible research efforts at EPFL during the first author's Ph.D. dissertation work. The first two conference papers written in this context starting in 2003 were not reproducible. It would score only one on our reproducibility scale. We did not save the precise parameter values for which certain results were obtained, original images used for experiments were lost, etc.

In 2004, upon a journal submission, we decided to make our work reproducible. It took about one week to check all the results and make sure they were reproducible, clean up and document the MATLAB code, and put everything online. Later, in undergraduate student projects, we developed a graphical user interface (GUI) in MATLAB to facilitate algorithm comparison without having to look at the code (Figure 1). This code is currently being downloaded at a rate of more than 100 times per month. Moreover, in the revisions of the paper, it was extremely helpful to have the code to reproduce each figure still at hand, as modifications to the figures could be easily made.

After this first publication, we repeated the exercise for our next two journal publications ([28] and [29]). This time, the actual work of making everything reproducible and putting the information online was much more efficient and was done in about one day. We already had reproducibility in mind from the start and thus carefully saved the code for each of the figures from the first time we produced them.

Finally, these reproducible research results were also very helpful when writing the Ph.D. dissertation [30]. Results and figures used in previous publications could be easily adapted (changing notations or input images) for use in the dissertation. Although hard to quantify, this represented a significant gain in time.
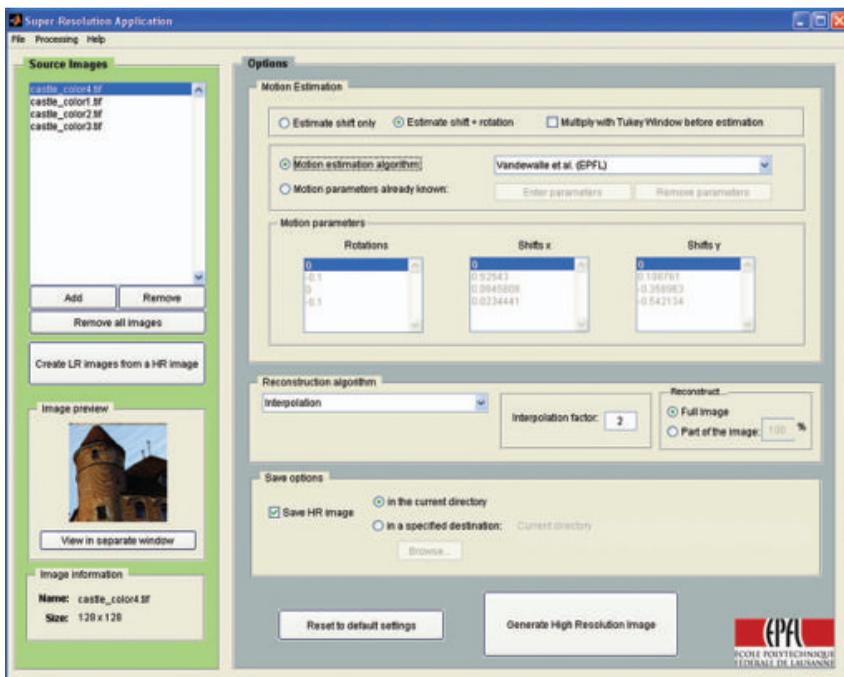
[FIG1] The GUI developed for comparison of super-resolution algorithms. It was developed on top of the reproducible research code from [23].

IMPLEMENTATION AND ISSUES

How to go from good intentions about reproducibility to actually publishing reproducible research requires a good setup as well as taking care of some important details (for example, about open access).

PUBLICATIONS REPOSITORY

A publications repository is an online database of publications where authors can upload their publications by filling out a form with information about the publication (authors, title, abstract, etc.), after which the publications are made accessible online. Such repositories are being created or are already in place at many institutions. Next to institutional repositories, a number of subject-based repositories exist, grouping publications around a certain topic. A well-known early example of this is arXiv (see http://www.arxiv.org). However, with most setups, it is nontrivial to add code or figures to the upload and have them displayed prominently with the publication itself. Note also that other material that is important for a publication but does not fit within the journal constraints (limited availability of space, video material, etc.) can be added online. More and more journals (for example, *IEEE Transactions on Signal Processing* or *BMC Bioinformatics* [31]) give this opportunity to their authors.

In collaboration with the EPrints team, we have therefore built a repository configuration for reproducible research. It is based on the widely used EPrints open-source software developed by Gutteridge and his team at University of Southampton (for more information see http://www.eprints.org). It requires a standard installation of EPrints and can then be configured using the files available online (http://rr.epfl.ch/information.html). The entire repository software is freely available online and can be installed and set up with minimum effort. Its main software requirements are Apache, MySQL, and Perl (with some extra modules). Next to the PDF and bibliographic information, the author is required to upload the code and data used in the paper and can also add other figures and relevant documents. For each uploaded file, the author can specify under which license he wants to distribute the file. Following our definition of reproducible research, the author can also specify the configurations tested. Then, as an additional functionality, readers can also give comments about a publication and evaluate the reproducibility of the work. Altogether, this offers great possibilities to make reproducible publications available online, adding a certain amount of interactivity, with a minimum installation effort.

Note that IEEE allows authors to self-archive their publications on their own Web pages or on an institutional repository, as long as an IEEE copyright notice is displayed on the same page for the readers [32]. The rights metadata for open archiving (ROMEO) Web page gives a useful overview of the open access policies for most popular journals. For more information on ROMEO, see http://www.sherpa.ac.uk/romeo.php.

Another issue is who is going to verify the additional information (code, data, etc.) made available online. Appointing a specific person at the publisher's side who verifies the code and data would be a good way of ensuring that results can be easily reproduced by an external person and would require a certain uniformity in ways of reproducing results. However, such a solution is very costly. Some publishers already ask their reviewers to check the availability of supplementary material, which could be generalized and could also include checking the code. Next to this, we have developed an additional feature in our repository setup that allows any user of the code to give an evaluation, which is then visible as information to other researchers reading the Web page. A user can quickly mark whether he has tested the code and if it works or does not work and whether he was able (or not) to reproduce the results. Next to this, it is also possible to add user comments.

> RECENT STUDIES HAVE SHOWN THAT PAPERS THAT ARE FREELY AVAILABLE ONLINE ARE CITED SIGNIFICANTLY MORE THAN PAPERS THAT ARE NOT.

LIFETIME OF WEB PAGES

The short lifetime of many Web pages is a very troubling issue when making research reproducible and keeping it so. When a researcher moves to another institution or retires, or when a lab Web site gets redesigned, the Web addresses at which reproducible research is available are often no longer valid. Sometimes the material is taken along and made available from a different location, but this is currently not a common practice.

As mentioned previously, a DOI could be a partial answer to that, as it makes it easier to retrieve the data when it has moved, although it does not bring the information back when it disappeared. We believe librarians can play an important role here, as they are generally experienced with the long-term preservation of works. One possible method is to use institutional repositories as described above and extend them to handle reproducible research compendia.

CODE AVAILABILITY

Licensing

A variety of licenses exist under which code can be made available online: proprietary-source and open-source licenses. With the first type, a user is typically not allowed to make modifications and redistribute the code. As such licenses are of less interest to us, we concentrate here on the second type only. Generally speaking, an open-source license allows a user to make modifications and/or redistribute the code as desired. Many variations exist: A listing of more than 50 existing open-source licenses can be found on the Open Source Web page (see http://www.opensource.org/licenses/category). One of the most popular licenses, the GNU General Public License [26], allows

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on April 22, 2009 at 12:47 from IEEE Xplore. Restrictions apply.

everyone to use, modify, and redistribute the code, as long as it is redistributed under the same (GPL) license. This prohibits other users to modify and/or redistribute the code under a different license, such as a proprietary-source license, while allowing the author to redistribute the code under another license later (open or proprietary source). Stodden proposes the open research license (ORL) for reproducible research [33]. It is actually a compilation of licenses, using the Creative Commons attribution license (without Share Alike component) to the media components, the Berkeley Software Distribution (BSD) license to the code, and the Science Commons Database Protocol to the data set. This license is particularly tailored to the nature of reproducible research as a compendium of various materials. Different from the GPL, Stodden does not include a Share Alike component in the ORL license, giving other researchers a maximum of liberty to reuse the work (while still acknowledging the original authors).

It appears reasonable to us that work funded by a government or public institution is also made public by putting the reproducible research results online under an open-source license. In fact, some of the government-funding agencies are requiring just that: For example, in the United States, the National Institutes of Health (NIH) has recently instituted a requirement that mandates open access of the papers funded by NIH at PubMed (see http://publicaccess.nih.gov/). However, for privately funded research or when the researchers want to commercialize their work, people are often reluctant to put code online. One could argue that privately funded research where the goal is not to give full disclosure of the way results were obtained (with parameter values, etc.) should not be published at all. Often, in a company, the work is only published after it has been protected by a patent. Once intellectual property has been protected, it should again be possible to fully disclose the work. Still, many companies funding university research or doing research themselves consider the know-how attached to a patent valuable and are not willing to have the resulting code distributed online. Note also that, even if the code and data are not made available online, it may be very useful to set reproducible research standards internally, such that colleagues inside the company can reproduce the results.

As another example, consider the case of a Ph.D. researcher who would like to start a company building on the results from his Ph.D. research. For such cases, it may not be a good idea to put the code online, but one can think of a setup where a user can upload his data set on a Web server and specify parameter values. The data are then processed on the server, and the processed results are returned to the user. This does not allow a detailed analysis and reuse of the source code, but it does allow an external researcher to obtain results on his own data sets, and compare them to the results obtained using other algorithms. At the same time, such a setup advertises the researcher's work.

Another argument for not putting code online could be the successful policy of the Internet Engineering Task Force (IETF). For a protocol to be accepted as a standard, it has to be described with sufficient accuracy in natural language such that (at least) two independent implementations can be made from it. In this way, the completeness of the standard description can be verified.

Platform Choice: Ease of Reproducibility Versus Author Efforts

Ideally, an external researcher should be able to download a reproducible research work and reproduce all the results with a simple mouse click. However, there exists a large variety of computer platforms, and making even some simple code work on most platforms requires a lot of work. Also, which tools can we consider to be available to everyone? Can we suppose that programs such as MATLAB, Mathematica, or even Windows are available, or should we constrain ourselves to the use of open-source programs such as R, Octave, or SciLab in a Linux environment?

> PRACTICAL EXPERIMENTS ARE, BY THEIR NATURE, PROBABLY THE MOST DIFFICULT TYPE OF WORK TO REPRODUCE.

The setups using Sweave, SCons, or the Makefiles developed by Claerbout et al. at SEP probably come closest to the ideal one-click scenario described above. Claerbout and his colleagues use Makefiles that require a single command to "burn," "build," "view," and "clean up" a figure [3], [34]. A disadvantage of Makefiles is that they only work on Linux/Unix. Fomel et al. are using this experience with Makefiles to develop Madagascar, a new system based on SCons and Python [10]. Madagascar is an open-source software package for reproducible numerical experiments, mainly used (so far) in geophysical data processing. It is a cross-platform package, running on Mac OSX, Linux/Unix, OS/2, and Windows. It allows the reproduction of research results using a few generic commands. In this way, an article can be seen as a combination of text, code, and data, that are combined in a single package and can be compiled together. A list of reproducible documents using Madagascar can be found at http://reproducibility.org/wiki/Reproducible_Documents.

A similar concept is implemented by Leisch in Sweave [35], [36]. Sweave is part of the open-source statistical computing package R and allows a user to interweave LaTeX text with R code. The R code is then executed on compilation of the file, and results and figures are produced and included in the resulting document.

The first and third authors have developed all of their super-resolution work in MATLAB and hence also made the MATLAB code directly available online. For each figure or table in the papers, there is a MATLAB script figure_x.m or table_y.m that reproduces the figure or table. While MATLAB is a commercial product, and therefore not accessible to everyone, we consider it as a widely available program, and did not do the additional work of rewriting the code in another language.

From the above discussions, we can clearly conclude that some tradeoffs have to be made. We need to balance the ease of use for a reader with the additional work to be done after development of the algorithms and initial code. It is also not clear which software packages and tools can be considered to be widely used and familiar to other researchers. The answers to these questions will typically depend on the specific field of research and the researcher's experience and affiliation.

In this tradeoff, it is also important to consider the long-term availability of a platform. Due to changes in data formats, compilers, and research platforms, research is often only reproducible in a limited time window. One could argue that life-long maintenance should be given to reproducible research compendia. We believe that this will happen in practice for the most useful results, as people will take care of converting or rewriting those results. Other documents will unfortunately become obsolete after a few years and will not be readable anymore.

DATA AVAILABILITY

Just like for the code, there are several important issues to think about before putting a data set online. The first, and by far most important, item is to make sure that the authors are allowed to make the data set available. If the authors have generated the data set themselves, they are typically the owners, and can publish the data as they wish (see [31] for an example). Instead of generating a custom data set each time, it may also be a good alternative to use an existing data set. Although it has to be checked carefully that the data set satisfies the requirements, this directly facilitates comparisons between different algorithms. For example, in stereo image disparity estimation, the Middlebury data set provides the possibility for easy comparison, and has become a de facto standard data set [37].

Publishing the data may not always be possible: if the data used in the experiments is copyrighted by a third party, if the data is acquired/produced by another research group not willing to make it available, or if the data is protected under confidentiality agreement as is the case with patient data. The authors should then check with the copyright holders which policies (for example, making data anonymous, or making only a sample available) are allowed.

Another very important issue is to give a detailed description of what the data represent, and how they were acquired. This may be very trivial or irrelevant for a set of images acquired with a digital camera for image classification, but might be of great importance when discussing for example a specific type of noise removal algorithm.

DOES REPRODUCIBLE RESEARCH INCREASE IMPACT?

Making research reproducible lowers the entrance barrier to a publication and therefore also increases its potential impact. Readers can get into the research work and use its results much more easily if code and data are also available.

As discussed previously, the first advantage of reproducible research we noticed is a gain in our own efficiency. It is much easier to pick a reproducible work up again at some later point, because everything is well organized and documented. We also received positive feedback from colleagues and students who downloaded the code and were happy about its availability. This allowed and simplified some collaborations and is a source of easily reusable demo material for students and visitors.

Recent studies have shown that papers that are freely available online are cited significantly more often than papers that are not. A first study on this topic was performed by Lawrence in 2001 [5]. More recently, the same exercise was repeated in
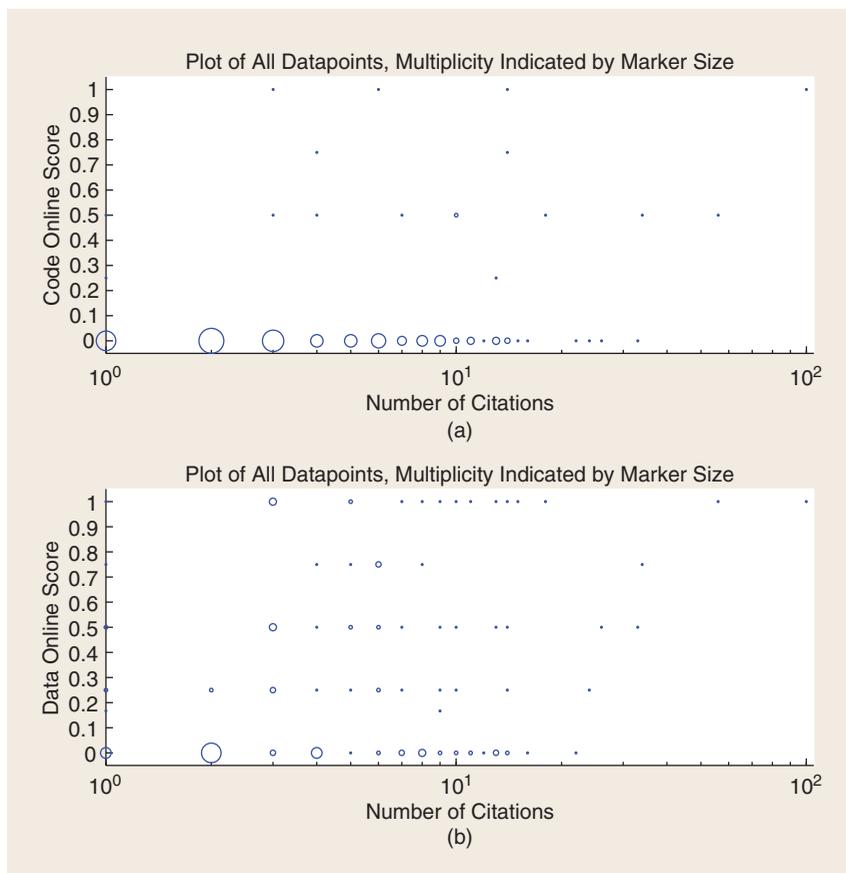
[FIG2] Online availability of (a) code and (b) data versus number of citations, as measured on 6 March 2008. The multiplicity of a data point is indicated by the size of the marker. The right part of the plots shows that highly cited papers typically have code and data available online.

other studies for publications in various fields [38], [39]. We are convinced that the same increased impact can be seen for publications that are reproducible, with an even larger effect. As they are easier to find and to use, such papers will typically also get cited more often. Piwowar et al. have performed a study in this direction, showing that publications for which the data sets are available online, are cited more often [6]. It is (currently) difficult to perform such an analysis for reproducible publications, because of the small scale (and varying degree of reproducibility) at which they are available.

A quantitative analysis that can be performed, and which was shown to be related to the later citations of a publication [38], is to look at the download statistics of the publication and/or the code. We have been recording these numbers for the EPFL lab for the past few years, showing that for some of the papers, the code is being downloaded up to 200 times per month.

We also reexamined the results from our reproducibility study described above and related them to the number of citations for each of the papers, as available on Web of Science. In Figure 2, the score on questions 2(b) and 3(c) (about the online availability of code and data, respectively) is displayed versus the number of citations of the papers. Although a clear one-to-one relation cannot be derived, we can see that the papers with a high number of citations (the right part of the plots) typically also have a high score on both questions. For the papers with a low number of citations (which are also much larger in number, represented in the left part of the plots), this is most often not the case. Of course, there were exceptions: papers scoring high on these questions with a low number of citations. This is to be expected, as online availability of code and data (in whatever form) is no guarantee of a quality publication. Due to a lack of space, we only plot these two results here (as they are also the most illustrative); the full set of figures is available online (see http://rr.epfl.ch/). In summary, computational papers that do not have code and data available online have a low chance of being cited.

CONCLUSIONS

What should we do to raise the quality of signal processing publications to an even higher level? We believe it to be crucial to maintain the precision in describing our work in publications, ensured through a high-quality reviewing process. We also believe that if the experiments are performed on a large data set, the algorithm is compared to the state-of-the-art methods, the code and/or data are well documented and available online, we will all benefit and make it easier to build upon each other's work. It is a clear win-win situation for our community: we will have access to more and more algorithms and can spend time inventing new things rather than recreating existing ones.

> **READERS CAN GET INTO THE RESEARCH WORK AND USE ITS RESULTS MUCH MORE EASILY IF CODE AND DATA ARE ALSO AVAILABLE.**

To come to the best possible methods for making our research reproducible, we invite you to share your opinions, ideas and examples of good practices by visiting our discussion forum at http://www.reproducibleresearch.net/forum. We look forward to a lively discussion!

AUTHORS

Patrick Vandewalle (Patrick.Vandewalle@philips.com) received an M.S. degree in electrical engineering from Katholieke Universiteit Leuven in 2001 and a Ph.D. degree from Ecole Polytechnique Fédérale de Lausanne (EPFL) in 2006. From 2006 to 2007, he was a postdoctoral researcher at EPFL. He joined Philips Research in September 2007, where he is currently a senior scientist. His research interests are in signal and image processing, sampling, digital photography, and 3-D.

Jelena Kovačević (jelenak@cmu.edu) received a Ph.D. degree from Columbia University. She is a professor of Biomedical Engineering and Electrical and Computer Engineering and the director of the Center for Bioimage Informatics at Carnegie Mellon University. She received the Belgrade October Prize and the E.I. Jury Award at Columbia University. She is a coauthor on a Signal Processing Society award-winning paper and is a coauthor of the book *Wavelets and Subband Coding* (Prentice-Hall, 1995) and a forthcoming book, *The World of Fourier and Wavelets: Theory, Algorithms and Applications*. She was the editor-in-chief of *IEEE Transactions on Image Processing*. Her research interests include multiresolution techniques and biomedical applications. She is a Fellow of the IEEE.

Martin Vetterli (martin.vetterli@epfl.ch) received an engineering degree from Eidgenoessische Technische Hochschule Zuerich (ETHZ), his M.S. degree from Stanford University, and his doctorate from Ecole Polytechnique Fédérale de Lausanne (EPFL). He was on the faculty of Columbia University, New York, and the University of California at Berkeley before joining the Communication Systems Department of EPFL, where he also serves as vice-president of the university. He works on signal processing and communications, in particular, wavelet and sampling theory with applications, source compression, joint source-channel coding, self-organized communication systems, sensor

networks, and their application to environmental monitoring. He is the author or coauthor of about 120 journal papers, and is an ISI highly cited scientist in engineering. He won the 1984 EURASIP Best Paper Award and the IEEE Signal Processing Society Best Paper Award in 1991, 1996, and 2006. He is the coauthor of the books *Wavelets and Subband Coding* (Prentice-Hall, 1995), *Signal Processing for Communications* (EPFL Press, 2008), and a forthcoming book, *The World of Fourier and Wavelets: Theory, Algorithms and Applications*. He is a Fellow of the IEEE.

REFERENCES

[1] Wikipedia (2009). *Reproducibility—Wikipedia, The Free Encyclopedia.* [Online]. Available: http://en.wikipedia.org/w/index.php?title=Reproducibility&oldid=262130461

[2] D. E. Knuth. (1984 May). Literate programming. *Comput. J.* [Online]. 27(2), 97–111. Available: http://www.literateprogramming.com/knuthweb.pdf

[3] J. Claerbout, "Electronic documents give reproducible research a new meaning," in *Proc. 62nd Ann. Int. Meeting of the Soc. of Exploration Geophysics*, 1992, pp. 601–604. [Online]. Available: http://sepwww.stanford.edu/doku.php?id=sep:research:reproducible:seg92

[4] J. B. Buckheit and D. L. Donoho. (1995). "WaveLab and reproducible research," Dept. of Statistics, Stanford Univ., Tech. Rep. 474. [Online]. Available: http://www-stat.stanford.edu/~donoho/Reports/1995/wavelab.pdf

[5] S. Lawrence. (2001). Free online availability substantially increases a paper's impact. *Nature.* [Online]. 411(6837), 521. Available: http://www.nature.com/nature/journal/v411/n6837/full/411521a0.html

[6] H. A. Piwowar, R. S. Day, and D. B. Fridsma. (2007 Mar.). Sharing detailed research data is associated with increased citation rate. *PLoS ONE.* [Online]. 2(3), e308. Available: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0000308

[7] Wikipedia. (2009). *Hwang Woo-Suk—Wikipedia, The Free Encyclopedia.* [Online]. Available: http://en.wikipedia.org/w/index.php?title=Hwang_Woo-Suk&oldid=260066075

[8] Wikipedia. (2009). *Scientific Misconduct—Wikipedia, The Free Encyclopedia.* [Online]. Available: http://en.wikipedia.org/w/index.php?title=Scientific misconduct&oldid=262353111

[9] H. Pearson. (2005 Apr.). Image manipulation: CSI: cell biology. *Nature.* [Online]. vol. 434, pp. 952–953, Apr. 2005.

[10] S. Fomel and G. Hennenfent, "Reproducible computational experiments using SCons," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 4, Apr. 2007, pp. 1257–1260. [Online]. Available: http://slim.eos.ubc.ca/Publications/Public/Conferences/ICASSP/2007/fomel07icassp.pdf

[11] R. Koenker. (1996). "Reproducible econometric research," Department of Econometrics, University of Illinois, Urbana-Champaign, IL, Tech. Rep. [Online]. Available: http://www.econ.uiuc.edu/~roger/research/repro/repro.html

[12] H. D. Vinod, "Care and feeding of reproducible econometrics," *J. Econometrics*, vol. 100, no. 1, pp. 87–88, Jan. 2001.

[13] C. Pouzat. (2006). Reproducible neurophysiological data analysis? [Online]. Available: http://www.biomedicale.univ-paris5.fr/physcerv/C_Pouzat/Compendium.html

[14] R. D. Peng, F. Dominici, and S. L. Zeger. (2006). Reproducible epidemiologic research. *Amer. J. Epidemiol.* [Online]. Available: http://aje.oxfordjournals.org/cgi/content/full/163/9/783

[15] M. Barni and F. Perez-Gonzalez. (2005 July). Pushing science into signal processing. *IEEE Signal Process. Mag.* [Online]. 22(4), 119–120. Available: http://www.gts.tsc.uvigo.es/~fperez/docs/pushing_science.pdf

[16] P. Vandewalle, G. Barrenetxea, I. Jovanovic, A. Ridolfi, and M. Vetterli. (2007 Apr.). "Experiences with reproducible research in various facets of signal processing research," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. [Online]. vol. 4, pp. 1253–1256. Available: http://infoscience.epfl.ch/record/97195

[17] J. Kovačević. (2007 Apr.) "How to encourage and publish reproducible research," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. [Online]. vol. 4, pp. 1273–1276. Available: http://www.andrew.cmu.edu/user/jelenak/Repository/07_04_Icassp_Kovacevic.pdf

[18] J. Vandewalle, J. Suykens, B. De Moor, and A. Lendasse. (2007 Apr.). "State of the art and evolutions in public data sets and competitions for system identification, time series prediction and pattern recognition," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. [Online]. vol. 4, pp. 1269–1272. Available: ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/ida/reports/06-167.pdf

[19] P. Marziliano, "Reproducible research: A case study of sampling signals with finite rate of innovation," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 4, Apr. 2007, pp. 1265–1268.

[20] M. Barni, F. Perez-Gonzalez, P. Comesaña, and G. Bartoli. (2007 Apr.). "Putting reproducible signal processing into practice: A case study in watermarking," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. [Online]. vol. 4, pp. 1261–1264. Available: http://www.gts.tsc.uvigo.es/gpsc/publications/wmark/icassp07reproducible.pdf

[21] S. Fomel and J. F. Claerbout (Guest Eds.). (2009). Special issue on reproducible research. *Comput. Sci. Eng.* [Online]. 11(1), pp. 5–40. Available: http://cise.aip.org/dbt/dbt.jsp?KEY=CSENFA&Volume=11&Issue=1

[22] P. Vandewalle, S. Süsstrunk, and M. Vetterli. (2003 July). "Superresolution images reconstructed from aliased images," in *SPIE/IS&T Visual Communication and Image Processing Conference*. [Online]. T. Ebrahimi and T. Sikora, Eds., vol. 5150, pp. 1398–1405. Available: http://infoscience.epfl.ch/record/33978

[23] P. Vandewalle, S. Süsstrunk, and M. Vetterli. (2006). A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP J. Appl. Signal Process.* (Special Issue on Super-Resolution Imaging). [Online]. vol. 2006, p. 14. Available: http://rr.epfl.ch/3

[24] A. Ridolfi and M. Z. Win. (2006). Ultrawide bandwidth signals as shot-noise: A unifying approach. *IEEE J. Select. Areas Commun.* (Special Issue on UWB Wireless Commun. Theory Appl.) [Online]. 4, 899–905. Available: http://rr.epfl.ch/11

[25] A. Ridolfi, "Power spectra of random spikes and related complex signals, with application to communications," Ph.D. dissertation, School of Comput. and Commun. Sci., Ecole Polytechnique Fédérale de Lausanne–EPFL, 2004. [Online]. Available: http://infoscience.epfl.ch/record/33626

[26] GNU. (1991). GNU General Public License. [Online]. Available: http://www.gnu.org/copyleft/gpl.html

[27] L. Meylan, D. Alleysson, and S. Süsstrunk. (2007). A model of retinal local adaptation for the tone mapping of color filter array images. *J. Opt. Soc. Amer. A (JOSA A).* [Online]. 24(9), 2807–2816. Available: http://rr.epfl.ch/10

[28] P. Vandewalle, L. Sbaiz, J. Vandewalle, and M. Vetterli. (2007). Super-resolution from unregistered and totally aliased signals using subspace methods. *IEEE Trans. Signal Process.* [Online]. 55(7), 3687–3703. Available: http://rr.epfl.ch/4

[29] L. Sbaiz, P. Vandewalle, and M. Vetterli. (2008). Groebner basis methods for multichannel sampling with unknown offsets. *Appl. Computat. Harmonic Anal.* [Online]. 25(3), 277–294. Available: http://rr.epfl.ch/18

[30] P. Vandewalle, "Super-resolution from unregistered aliased images," Ph.D. dissertation, School of Comput. and Commun. Sci., Ecole Polytechnique Fédérale de Lausanne, Switzerland, July 2006, no. 3591. [Online]. Available: http://rr.epfl.ch/6

[31] A. Chebira, Y. Barbotin, C. Jackson, T. Merryman, G. Srinivasa, R. F. Murphy, and J. Kovačević. (2007). A multiresolution approach to automated classification of protein subcellular location images. *BMC Bioinformatics.* [Online]. 8(210). Available: http://www.andrew.cmu.edu/user/jelenak/Repository/07_ChebiraBJMSMK/07_ChebiraBJMSMK.html

[32] IEEE. (2008). Electronic Information Dissemination, IEEE PSPB operations manual [Online]. Available: http://www.ieee.org/web/publications/rights/policies.html

[33] V. Stodden. (2009). The legal framework for reproducible scientific research: Licensing and copyright. *Comput. Sci. Eng.* [Online]. 11(1), 35–40. Available: http://www.stanford.edu/~vcs/papers/ReproducibleResearchLegal08172008.pdf

[34] M. Schwab, M. Karrenbach, and J. Claerbout. (2000 Nov.). Making scientific computations reproducible. *Comput. Sci. Eng.* [Online]. 2(6), 61–67. Available: http://sepwww.stanford.edu/doku.php?id=sep:research:reproducible:cip

[35] F. Leisch, "Sweave: Dynamic generation of statistical reports using literate data analysis," in *Compstat 2002 — Proc. Computational Statistics*, W. Härdle and B. Rönz, Eds. Physica Verlag, Heidelberg, 2002, 575–580. [Online]. Available: http://www.statistik.lmu.de/~leisch/Sweave/

[36] C. Geyer. (2006). An Sweave demo. [Online]. Available: http://www.stat.umn.edu/~charlie/Sweave

[37] D. Scharstein and R. Szeliski. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision.* [Online]. 47(1), 7–42. Available: http://vision.middlebury.edu/stereo

[38] T. D. Brody, "Evaluating research impact through open access to scholarly communication," Ph.D. dissertation, School of Electronics and Comput. Sci., Univ. Southampton, May 2006. [Online]. Available: http://eprints.ecs.soton.ac.uk/13313/

[39] S. Hitchcock. (2009). The effect of open access and downloads ('hits') on citation impact: A bibliography of studies [Online]. Available: http://opcit.eprints.org/oacitation-biblio.html

[SP]