The SMART protocols checklist: guidelines for reporting experimental protocols in life sciences (#19630)

First submission

Please read the **Important notes** below, the **Review guidance** on page 2 and our **Standout reviewing tips** on page 3. When ready **submit online**. The manuscript starts on page 4.

Important notes

Editor
Todd Vision

Files	1 Latex file(s) 1 Other file(s) Please visit the overview page to **download and review** the files not included in this review PDF.
Declarations	No notable declarations are present

Review guidelines

Please read in full before you begin

How to review

When ready **submit your review online**. The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

To finish, enter your editorial recommendation (accept, revise or reject) and submit.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to **PeerJ standards**, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see **PeerJ policy**).

EXPERIMENTAL DESIGN

- Original primary research within **Scope of the journal**.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.
- Speculation is welcome, but should be identified as such.

The above is the editorial criteria summary. To view in full visit **https://peerj.com/about/editorial-criteria/**

7 Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that your international audience can clearly understand your text. I suggest that you have a native English speaking colleague review your manuscript. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Give specific suggestions on how to improve the manuscript	*Line 56: Note that experimental data on sprawling animals needs to be updated. Line 66: Please consider exchanging "modern" with "cursorial".*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

The SMART protocols checklist: guidelines for reporting experimental protocols in life sciences

Olga Giraldo Corresp., [1] , **Alexander Garcia** [1] , **Oscar Corcho** [1]

[1] Ontology Engineering Group. Campus de Montegancedo, sn, Boadilla del Monte, Universidad Politécnica de Madrid, Madrid, Spain

Corresponding Author: Olga Giraldo
Email address: ogiraldo@fi.upm.es

Experimental protocols are key elements when planning, doing and reporting research. Experimental protocols are central to research methodologies; however they vary in content, structure and metadata elements. This article presents the SMART Protocols checklist (SP checklist); this is a guideline of key content for reporting experimental protocols. We also describe the methodology that was followed in order to develop the SP checklist. The 18 data elements proposed in our checklist are represented in the SMART Protocols ontology. We focus on the content, what should be included. Rather that advocating a specific format for protocols in Life Sciences, the checklist includes a full description of what is planned; it does not propose how to design or conduct an experiment. By providing guidance for key content, the SP checklist aim to facilitate the documentation of high-quality protocols; thus making it easier for researchers to replicate the described experiments.

The SMART Protocols Checklist: Guidelines for Reporting Experimental Protocols in Life Sciences

Olga Giraldo[1], **Alexander Garcia**[1], **and Oscar Corcho**[1]

[1]**Ontology Engineering Group/Artificial Intelligence/ETSI, Universidad Politécnica de Madrid, Madrid, Spain**

Corresponding author:
Olga Giraldo[1]

Email address: ogiraldo@fi.upm.es

ABSTRACT

Experimental protocols are key elements when planning, doing and reporting research. Experimental protocols are central to research methodologies; however they vary in content, structure and metadata elements. This article presents the SMART Protocols checklist (SP checklist); this is a guideline of key content for reporting experimental protocols. We also describe the methodology that was followed in order to develop the SP checklist. The 18 data elements proposed in our checklist are represented in the SMART Protocols ontology. We focus on the content, what should be included. Rather ~~that~~ advocating a specific format for protocols in Life Sciences, the checklist includes a full description of what is planned; it does not propose how to design or conduct an experiment. By providing guidance for key content, the SP checklist aim to facilitate the documentation of high-quality protocols; thus making it easier for researchers to replicate the described experiments.

INTRODUCTION

Experimental protocols are fundamental information structures that support the description of the processes by means of which results are generated in experimental research (Giraldo et al., 2014). Experimental protocols describe how the data were produced, the steps undertaken and conditions under which these steps are to be carried out. Experimental protocols are crucial for the execution and reproducibility of any study, the reporting structure remains highly idiosyncratic. However, there is little consistency in the data elements that should be included; the required data elements to be reported within the same discipline vary from publisher to publisher.

Kilkenny et al. (2010a) evaluated 271 journal articles, they found that 4% did not report the number of animals used anywhere in the methods or the results sections. Assessing statistical significance requires ~~to know~~ the number of animals participating in an experiment. If experimental methods are to be reproduced, reused or adapted, then the size as well as other attributes of the population must be reported; comparing results and integrating data also require quality descriptions in the reported experimental methods. Very often, the studies reported in journals only contain a shortened summary of the methods due to the restriction in the number of pages Godlee (2001); this makes it difficult for researchers to understand how data were produced. The protocols are usually adapted for different purposes; however, keeping track of provenance across protocols is difficult. Once a study is completed, the protocol is usually archived; little care is paid to the derivations and/or versions of the protocols.

Several efforts are building data storage infrastructures, e.g. 3TU. Datacentrum (4TU, 2017), CSIRO Data Access Portal (CSIRO, 2017), Dryad (Dryad, 2017), figshare (figshare, 2017) and Zenodo (Zenodo, 2017). These data repositories make it possible to review the data and evaluate whether the analysis and conclusions drawn are accurate. However, they do little to validate the quality and accuracy of the data itself. Evaluating research implies being able to obtain similar, if not identical results. Journals and founders are now asking for datasets to be publicly available for reuse and validation. Fully meeting this goal requires datasets to be endowed with auxiliary data providing contextual information about the

dataset, e.g methods used to derive such data (Assante et al., 2016; Simmhan et al., 2005). If data must be public and available, shouldn't researchers be hold to the same principle when it comes to methodologies? Our work addresses the problem of adequate reporting for experimental protocols; our contribution is extensible and compatible with efforts such as those from BioSharing (McQuilton et al., 2016; Biosharing, 2017), the Resource Identification Portal (RIP, 2017), the Structured, Transparent, Accessible Reporting (STAR) initiative (Marcus, 2016; STAR, 2017) as well as with the ARRIVE (Kilkenny et al., 2010b) and NIH guidelines (NIH, 2017b).

The SMART Protocols Checklist is based on an exhaustive analysis of over 500 published and non-published experimental protocols, as well as guidelines for authors from journals publishing protocols. In the "Materials and Methods" section we present detailed information about the guidelines, protocols, standards and ontologies that we analyzed. Based on our analysis, we propose the SMART Protocols checklist (SP checklist). By providing a set of data elements and guidance for key content, the SP checklist aim to facilitate and improve the documentation of protocols. In the "Results" section, we present examples indicating how to report each metadata element. Our checklist provides a minimum set of metadata elements that could easily be reused, or adapted, by publishers and laboratories as part of their best practices for reporting experimental information.

MATERIALS AND METHODS

Materials

In this section we present the resources that were analyzed in order to identify the necessary and sufficient information for reporting a protocol in life sciences. The list of resources includes: i) ~~G~~guidelines for authors from journals publishing protocols, ii) a corpus of protocols and, iii) a set of minimum reporting structures and ontologies. From these sources we extracted and analyzed data elements related to the documentation of experimental protocols.

i) Instructions for authors from analyzed journals.

The instructions, ~~A.K.A.~~ guidelines for authors, describe the information that should be provided in the protocol. Publishers always have this information available for authors. In Table 1 we presented the list of guidelines that were analyzed.

Journal	Guidelines for authors
BioTechniques (BioTech)	(bioTechniques, 2017)
CSH protocols (CSH)	(CSH-Protocols, 2017)
Current Protocols (CP)	(Current-Protocols, 2017)
Journal of Visualized Experiments (JoVE)	(JOVE, 2017)
Nature Protocols (NP)	(Nature-Protocol, 2017)
Springer Protocols (SP)	(Springer-Protocols, 2017)
Methods X	(MethodX, 2017)
Bio-protocols (BP)	(Bio-protocol, 2017)
Journal of Biological Methods (JBM)	(JBM, 2017)

Table 1. Guidelines for reporting experimental protocols.

ii) Corpus of protocols.

Our corpus includes 530 published and unpublished protocols. Unpublished protocols (75 in total) were collected from four laboratories located at the International Center for Tropical Agriculture (CIAT) (CIAT, 2017). The published protocols (455 in total) were gathered from 12 journals, namely: BioTechniques, Cold Spring Harbor Protocols, Current Protocols, Genetics and Molecular Research (GMR, 2017), JoVE, Nature Protocol Exchange (NPE, 2017), Plant Methods (Plant-Methods, 2017), Plos One (Plos-One, 2017), Springer Protocols, MethodsX, Bio-Protocol and the Journal of Biological Methods. The analyzed protocols cover areas such as cell biology, molecular biology, immunology, neurosciences, and virology. The number of protocols from each journal is presented in Table 2

Journal	Number of protocols
BioTechniques (BioTech)	16
CSH protocols (CSH)	267
Current Protocols (CP)	31
Genetics and Molecular Research (GMR)	5
Journal of Visualized Experiments (JoVE)	21
Nature Protocols Exchange (NPE)	39
Plant Methods (PM)	12
Plos One (PO)	5
Springer Protocols (SP)	5
Methods X	7
Bio-protocols (BP)	40
Journal of Biological Methods (JBM)	7
non-published protocols from CIAT	75

Table 2. The corpus of protocols is available at: Giraldo (2017)

iii) Minimum information standards and Ontologies.
We analyzed minimum information standards from the BioSharing catalog (Biosharing, 2017), e.g. MIAPPE (MIAPPE, 2017), MIARE (MIARE, 2017) and MIQE (Bustin et al., 2009). See Table 3 for more information about the minimum information standards that we analyzed.

Standards	Description
Minimum Information about Plant Phenotyping Experiment (MIAPPE)	A reporting guideline for plant phenotyping experiments.
CIMR: Plant Biology Context (Nikolau et al., 2006)	A standard for reporting metabolomics experiments.
The Gel Electrophoresis Markup Language (GelML) (Gibson et al., 2010)	A standard for representing gel electrophoresis experiments performed in proteomics investigations.
Minimum Information about a Cellular Assay (MIACA) (MIACA, 2017)	A Standardized description of cell-based functional assay projects.
Minimum Information About an RNAi Experiment (MIARE)	A checklist describing the information that should be reported for an RNA interference experiment.
The Minimum Information about a Flow Cytometry Experiment (MIFlowCyt) (Lee et al., 2008)	Recommendations about descriptions of the specimens and reagents included in the flow cytometry experiment, the configuration of the instrument used to perform the assays and the data processing approaches used to interpret the primary output data.
Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE)	This guideline guideline describes the minimum information necessary for evaluating qPCR experiments.
ARRIVE (Animal Research: Reporting of In Vivo Experiments)	Initiative to improve the standard of reporting of research using animals.

Table 3. Minimum Information Standards analyzed.

We paid special attention to the recommendations indicating how to describe specimens, reagents, instruments, software and other entities participating in different types of experiments. Ontologies available at Bioportal (Whetzel et al., 2011) and Ontobee (Xiang et al., 2011) were also considered; our analysis included, amongst others, OBI (Bandrowski et al., 2016), IAO (IAO, 2017), EXPO (Soldatova and King, 2006) and EXACT (Soldatova et al., 2008, 2014). We focused on ontologies representing bioassays, organisms, anatomical parts, reagents, chemical compounds, instruments and experimental actions; the list of analyzed ontologies is presented in Table 4.

Ontology	Description
The Ontology for Biomedical Investigations (OBI)	An ontology for the description of life-science and clinical investigations.
The Information Artifact Ontology (IAO)	An ontology of information entities.
The ontology of experiments (EXPO)	An ontology about scientific experiments.
The ontology of experimental actions (EXACT)	An ontology representing experimental actions.
The BioAssay Ontology (BAO) (Abeyruwan et al., 2014)	An ontology describing biological assays.
The Experimental Factor Ontology (EFO) (Malone et al., 2010)	The ontology provides a description of experimental variables available in EBI databases, and for external projects such as the NHGRI GWAS catalog. It combines parts of several biological ontologies, such as anatomy, disease and chemical compounds.
eagle-i resource ontology (ERO) (Torniai et al., 2011)	An ontology of research resources such as instruments, protocols, reagents, animal models and biospecimens.
NCBI taxonomy (NCBITaxon) (Federhen, 2015)	An ontology representation of the NCBI organismal taxonomy.
Chemical Entities of Biological Interest (ChEBI) (Hastings et al., 2013)	Classification of molecular entities of biological interest focusing on 'small' chemical compounds.
Uberon multi-species anatomy ontology (UBERON) (Mungall et al., 2012)	A cross-species anatomy ontology covering animals and bridging multiple species-specific ontologies.
Cell Line Ontology (CLO) (Sarntivijai et al., 2014, 2011)	The ontology was developed to standardize and integrate cell line information.

Table 4. Ontologies.

Methods for developing the SMART Protocols checklist

Developing the SP checklist entailed a series of activities; these were organized in the following stages: ~~i)Analysis~~ of guidelines for authors. ii) analysis of protocols. iii) analysis of standards and ontologies, and iv) evaluation of the data elements from our checklist. For a detailed representation of our workflow see Figure 1

Analyzing guidelines for authors

We reviewed instructions for authors from nine journals as presented in Table1. In this stage (step A in Figure 1), we identified bibliographic data elements suggested in the guidelines, see Table5.

Bibliographic data elements	Bio-Tech	NPE	CP	JoVE	CSH	SP	BP	MethodsX	JBM
title/name	Y	Y	Y	Y	Y	Y	Y	Y	Y
author(s) name	Y	Y	Y	Y	Y	Y	Y	Y	Y
protocol identifier (doi, url, iri, etc.)	DOI	DOI	DOI	DOI	DOI	DOI	DOI	DOI	DOI
protocol source (source, retrieved from)	N	Y	N	N	N	N	N	N	N
references/related publications	Y	Y	Y	Y	Y	Y	Y	Y	Y
categories or keywords	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 5. The "Y" means that the datum was suggested. The "N" means that the datum was not suggested.

In addition, we identified the rhetorical elements and determined which of these were required as either mandatory, suggested (to be included where applicable) or, optional (not required), see Table 6 for more details.

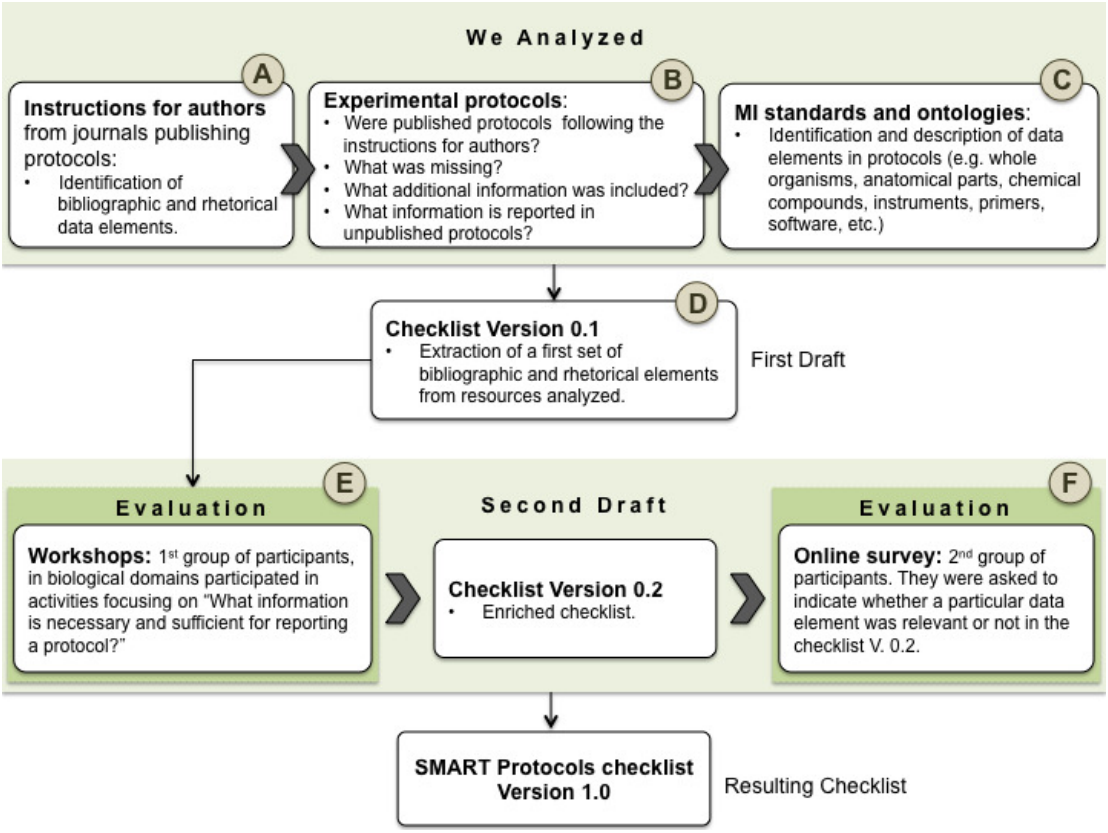

Figure 1. Methodology Workflow.

Analyzing the protocols.

In 2014 we started by manually reviewing 175 published and unpublished protocols; these were from domains such as, cell biology, biotechnology, virology, biochemistry and pathology. From this collection, 75 are unpublished protocols collected from four laboratories located at the CIAT. In 2015 our corpus grew to 530; we included 355 published protocols gathered from the 12 journals as listed in Table 2. Our corpus of published protocols is: i) ~~Open~~ access; ii) identifiable, each document has an identifier (DOI, URL, IRI, PURL); iii) multidisciplinary, the protocols cover several domains in life sciences, e.g. cell biology, developmental biology, neuroscience, microbiology, biochemistry, biotechnology, etc.; and iv) peer-reviewed.

In this stage, step B in Figure 1, we analyzed the content of the protocols; theory vs. practice was our main concern. We manually verified if published protocols were following the guidelines; if not, ~~*What was missing?, or,*~~ *what additional information was included?* We also reviewed common data elements in unpublished protocols.

Analyzing Minimum Information Standards and ontologies

Biomedical sciences have an extensive body of work related to minimum information standards and reporting structures, e.g. those from the BioSharing initiative. We were interested in determining whether there was any relation to these resources. Our checklist has the data elements that are common across these resources. We analyzed MIQE, used to describe qPCR assays; we also looked into MIACA, that provide guidelines to report cellular assays; ARRIVE, that provides detailed descriptions of experiments on animal models and MIAPPE, addressing the descriptions of experiments on animal models. See Table 3 for a list of the standards that we analyzed.

Metadata, data and, reporting structures in biomedical documents are frequently related to ontological concepts. We also looked into relations between metadata elements and biomedical ontologies available in BioPortal and Ontobee, such as OBI, CLO, CheBI, UBERON, ERO and EXACT. The complete list of the ontologies that we analyzed is presented in Table 4.

Rhetorical/Discourse Elements	Bio-Tech	NPE	CP	JoVE	CSH	SP	BP	Meth-odsX	JBM
Description of the protocol (objective, range of applications where the protocol can be used, advantages, limitations)	AR	AR	AR	S	AR	AR	AR	AR	S
Description of the sample tested (name; ID; strain, line or ecotype; developmental stage; organism part; growth conditions; treatment type; size)	NC	NC	S	NC	NC	NC	NC	NC	NC
Reagents (name, vendor, catalog number)	R	S	S	S	R	S	R	NC	S
Equipment (name, vendor, catalog number)	R	S	S	S	R	S	R	NC	S
Recipes for solutions (name, pH, final concentration, volume)	R	S	S	S	S	S	R	NC	S
Procedure description	R	R	R	S	R	R	R	R	S
Include alternatives to performing specific steps	NC	NC	S	S	NC	S	NC	NC	NC
Critical steps	R	NC	S	NC	NC	NC	NC	NC	NC
Pause point	R	NC	NC	O	S	NC	NC	NC	NC
Troubleshooting	R	O	R	O	S	S	NC	NC	S
Caution/warnings	NC	NC	R	O	NC	S	NC	NC	S
Execution time	NR	O	S	NC	NC	S	NC	NC	NC
Storage conditions (reagents, recipes, samples)	R	NC	R	S	S	S	NC	NC	NC
results (figure, tables)	R	NC	R	R	S	R	S	NC	S

Table 6. R= Required; NC= Not Considered in guidelines; S= Suggested; O= Optional; AR= Required in the Abstract.

Output of resources analyzed.

The first draft version of the checklist is the main output from the analysis~~;~~ as illustrated in Figure 1 (A, B, and C).

Evaluation of data elements by domain experts

This stage entailed two activities. The first activity was carried out at the CIAT with the participation of 19 domain experts in areas such as virology, pathology, biochemistry and plant biotechnology. The input of this activity was the checklist V. 0.1 (see step E in Figure 1). This evaluation focused on "*What information is necessary and sufficient for reporting an experimental protocol?*". The result from this activity was the version 0.2 of the checklist. Domain experts suggested to use an on-line survey for further validation. This survey was designed to enrich and validate the checklist V. 0.2. We used a Google survey that was circulated over mailing lists; participants did not have to disclose their identity (see step F in Figure 1). The output of this activity was the checklist V. 1.0. The survey and its responses are available in Annex2 and Annex3 at https://smartprotocols.github.io/annex/

RESULTS

Bibliographic data elements

In addition to the data elements presented in Table~~5~~, domain experts proposed **"author identifier"** and **"provenance of the protocol"**. These two data elements were not often represented in either published or unpublished protocols. For unpublished protocols the following bibliographic data elements were added to the checklist "**laboratory-validation scientist**" and "**version of the protocol**". A complete description of the bibliographic metadata elements proposed is presented in the results section under *"The SMART Protocols checklist"*.

Discourse data elements

Here we present the descriptive elements that domain experts considered necessary to understand the suitability of a protocol. For instance, ~~the~~ 100% of the domain experts (from a total of 19 people)~~,~~ agreed to include the **"objective"**, ~~the~~ 79% of them agreed to include the **"applications"**, ~~the~~ 49% of them agreed to include **"advantages"** and ~~the~~ 84% of them agreed to include the **"limitations"**. See Figure 2.

	Purpose	Applications	Advantages	Limitations
# of experts	19	15	9	16
%	100	79	47	84

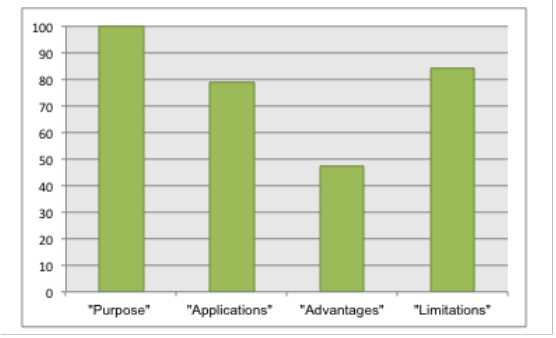

Figure 2. Descriptive data elements.

The domain experts agreed to include data elements that could facilitate the planning and execution of protocol instructions. ~~The~~ 79% of the experts included the datum **"alternatives to performing specific steps"**~~,~~ and ~~the~~ 84% of them included the data elements **"critical steps"**, **"troubleshooting"**, **"pause point"** and **"timing"** see Figure 3.

	alternatives to performing specific steps	Critical steps	Pause point	Timing	Troubleshooting
# of experts	15	16	16	16	16
%	79	84	84	84	84

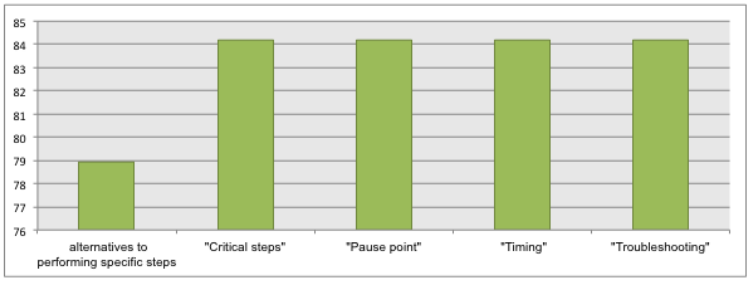

Figure 3. Data elements for describing protocol instructions.

All participants agreed to include material entities such as: Samples, instruments (including conventional and ~~H~~High-throughput equipment), consumables, reagents (those purchased ready-to-use), recipes for solutions (mixtures prepared in the lab), kits, primers and software. Attributes representing each of these materials were also considered because these entities participate in one or more steps through the protocol.

The sample is the input of the protocol; it needs an accurate description that varies depending on the domain. Our checklist demands the description of the sample and recommends some attributes to make this description more accurate. We included all the attributes suggested by the domain experts. Interestingly, only one expert (5%) considered **"amount"** and **"identifier"** to be important. Data elements such as, **"strain or line"**, **"developmental stage"** and **"organism part"** were considered very important. See Figure 4. The description of the sample should be revised and adapted depending on the application context of the protocol.

	Strain or line	Developmental stage	Organism part	Growth conditions and pre-treatment of the sample	Amount	ID
# of experts	18	17	17	3	1	1
%	95	89	89	16	5	5

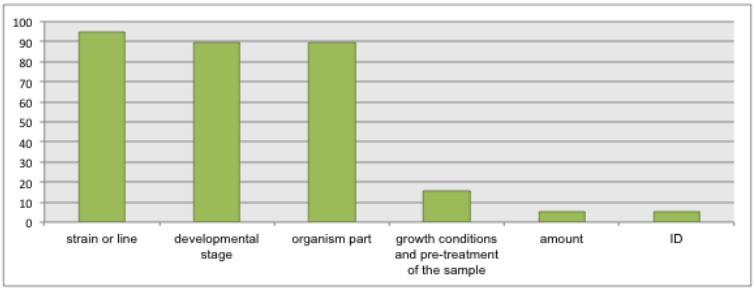

Figure 4. Data elements for describing the sample.

Domain experts working in assays involving standard PCR, Construction of cDNA Libraries, cloning, sequencing, and gene detection helped us to identify a minimum set of data elements useful to describe primers. The items selected by experts were reused from the MIQE guidelines to describe Primers/Oligonucleotides (Bustin et al., 2009). Only one expert (5%) included the attribute "**expected PCR product size**". Attributes such as **"primer name"**, **"primer sequence"** and **"manufacturer or vendor"** ranked high in our evaluation. See Figure 5.

	name	primer sequence	vendor	expected product size
# of experts	15	18	8	1
%	79	95	42	5

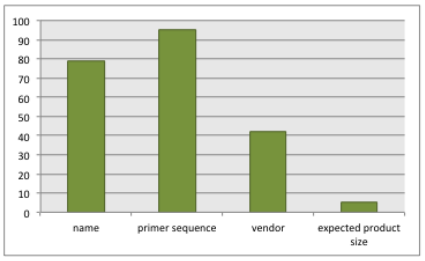

Figure 5. Data elements for describing primers.

The protocols should include the recipes that describe the preparation of solutions in the laboratory. All the attributes suggested by domain experts were included. See Figure 6.

	name	chemical compound	initial concentration	final concentration	storage conditions	cautions	hints
# of experts	15	15	10	12	13	10	6
%	79	79	53	63	68	53	32

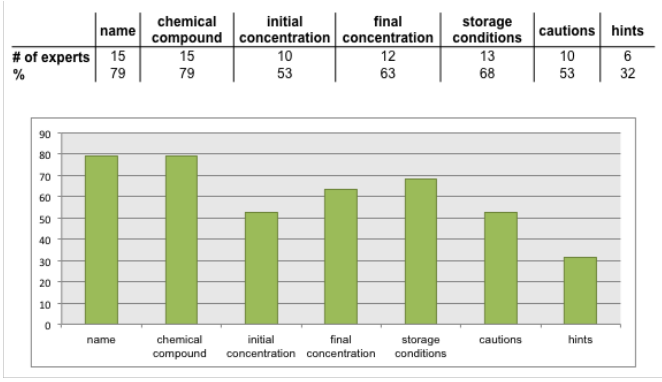

Figure 6. Data elements for describing solutions prepared in the laboratory.

Domain experts strongly suggested that consumables had to be included in the checklist. It was a general surprise not to find these data elements in the guidelines for authors that we analyzed. They shared bad experiences caused by the lack of information about the type of consumables that should be used in different assays. Some of the incidents that may arise from the lack of this information include: i) cross contamination, when no information suggesting the use of filtered pipet tips is available; ii) misuse of containers, when no information about the use of containers resistants to extreme temperatures and/or impacts; iii) misuse of containers, when a container made of a specific material should be used, e.g. glass vs. plastic vs. metal. This is critical information because researchers need to know if reagents or solutions prepared in the laboratory require some specific type of containers in order to avoid unnecessary reactions altering the result of the assay. The attributes for representing consumables are illustrated in Figure 7.

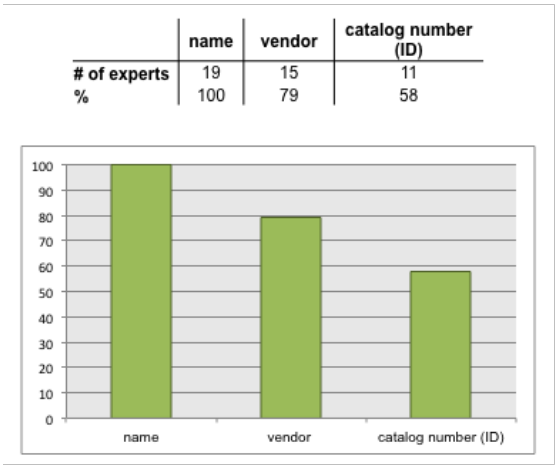

Figure 7. Data elements for describing consumables.

Some protocols include the use of software. In these cases the software should include the **"name"**, **"version"** and **"homepage"**. See Figure 8.

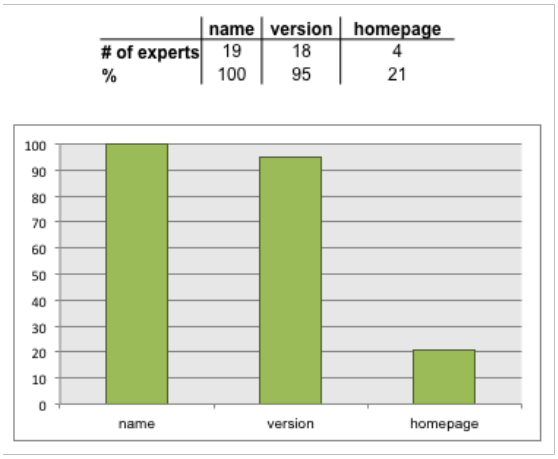

Figure 8. Data elements for describing software.

The SMART Protocols checklist.

The description of each metadata element is given below; examples illustrating good and bad reporting are also included. A checklist presenting the metadata proposed for reporting an experimental protocols is available at https://github.com/oxgiraldo/SMART-Protocols/tree/master/SP

Bibliographic data elements

Title. The title of the protocol should be informative, explicit, and concise (50 characters or fewer). The use of ambiguous terminology and trivial adjectives or adverbs (e.g., novel, rapid, efficient, inexpensive, or their synonyms) should be avoided. The use of numerical values, abbreviations, acronyms, and trademarked or copyrighted product names is discouraged. This definition was adapted from BioTechniques (bioTechniques, 2017). In Table7 we present examples illustrating how to define the tittle.

Bad example	A **single*** protocol for extraction of **gDNA‡** from bacteria and yeast.	Protocol available at (Vingataramin L, 2015)
Good example	Extraction of nucleic acids from yeast cells and plant tissues using ethanol as medium for sample preservation and cell disruption.	Protocol available at (Linke et al., 2010)

Table 7. *Use of ambiguous terminology, ‡use of abbreviations.

Author. The full name(s) of the author(s) with an author ID (e.g. ORCID (ORCID, 2017) or research ID (ResearcherID, 2017)) is required. The role of each author is also required; depending on the domain there may be several roles. It is important to use a simple word that describes who did what. We have identified two roles that are common across our corpus of documents.

- ***Creator of the protocol:*** This is the person or team responsible for the development or adaptation of a protocol.

- ***Laboratory-validation scientist:*** Protocols should be validated in order to certify that the processes are clearly described; it must be possible for others to follow the described processes. If applicable, statistical validation should also be addressed. The validation may be procedural (related to the process) or statistical (related to the statistics). According to the Food and Drug Administration (FDA) (FDA, 2017), validation is "*Establishing documented evidence which provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes*" (Das, 2011).

Version of the protocol. The following guidelines for version control of protocols were adapted from the version control guidelines proposed by the National Institute of Health (NIH) (NIH, 2017a).

- ***Document dates:*** The date indicating when the protocol was generated should be indicated in the first page and, whenever possible, incorporated into the header or footer of each page in the document.

- ***Version numbers:*** The current version number of the protocol is identified in the first page and, when possible, incorporated into the header or footer of each page of the document.

  - ***Draft document version number:*** The first draft of a document will be Version 0.1. Subsequent drafts will have an increase of "0.1" in the version number, e.g., 0.2, 0.3, 0.4, ... 0.9, 0.10, 0.11.

  - ***Final document version number and date:*** The author (or investigator) will deem a protocol final after all reviewers have provided final comments and these have been addressed. The first final version of a document will be Version 1.0; the date when the document becomes final should also be included. Subsequent final documents will have an increase of "1.0" in the version number (1.0, 2.0, etc.).

- ***Documenting substantive changes:*** A list of changes from the previous drafts or final documents will be kept. The list will be cumulative and identify the changes from the preceding document versions so that the evolution of the document can be seen. The list of changes and consent/assent documents should be kept with the final protocol.

Good example	*"This protocol was adapted from "How to Study Gene Expression," Chapter 7, in Arabidopsis:A Laboratory Manual (eds. Weigel and Glazebrook). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, 2002."*	Protocol available at (Blazquez, 2007)

Table 8

Provenance of the protocol. This indicates whether the protocol results from modifying a previous one or, if it is the result from reusing steps in other protocols. See Table 8

Protocol availability. This indicates whether the protocol comes from a repository, e.g. protocols.io or, if was published in a journal like JoVE, MethodsX or Bio-Protocols. This data element should not be confused with the provenance of the protocol. The former indicates where was the protocol obtained from. The latter refers to adaptations of the protocol.

Descriptive metadata

This set of data elements should make it easier for the readers to decide on the suitability of the protocol for their experimental problem.

Overall objective or Purpose. The description of the purpose should make it possible for readers to decide on the suitability of the protocol for their experimental problem. See Table 9

Good example	*"Development of a method to isolate small RNAs from different plant species (. . .) that no need of first total RNA extraction and is not based on the commercially available TRIzol® Reagent or columns."*	Protocol available at (Rosas-Cárdenas et al., 2011)

Table 9

Application of the protocol. This information should indicate the range of techniques where the protocol could be applied. See Table 10.

Good example	*"DNA from this experiment can be used for all kinds of genetics studies, including genotyping and mapping."*	Protocol available at (Lu, 2011)

Table 10

Advantage(s) of the protocol(s). The advantages of a protocol compared to other alternatives should be discussed. See Table 11. Where applicable, references should be made to alternative methods that are commonly used to achieve the same result.

Good example	*"We describe a fast, efficient and economic in-house protocol for plasmid preparation using glass syringe filters. Plasmid yield and quality as determined by enzyme digestion and transfection efficiency were equivalent to the expensive commercial kits. Importantly, the time required for purification was much less than that required using a commercial kit."*	Protocol available at (Kim and Morrison, 2009)

Table 11

Limitation(s) of the protocol(s). The limitations of the protocol should be discussed. It should be clear in which situations the protocol could be unreliable or unsuccessful. See Table 12.

Metadata for materials

The metadata related to the sample, reagents, recipes for solutions, instruments, consumables, kits, software etc. is described below. These are the necessary materials for carrying out the steps of the protocol.

Good example	*"A major problem faced both in this and other safflower transformation studies is the hyperhydration of transgenic shoots which result in the loss of a large proportion of transgenic shoots."*	Protocol available at (Belide et al., 2011)

Table 12

Sample. This is the role played by a biological substance; the sample is an experimental input to a protocol. The information required depends on the type of sample being described and the requirements from different communities. Here, we present the data elements for samples commonly used across the protocols and guidelines that we analyzed.

- **Bio-source properties:** Information about the physical object that will be analyzed.

  Strain or genotype: Subspecies information such as ecotype, cultivar, accession, line. In the case of crosses or breeding results, pedigree information should also be provided.

  Starting material: Whole organism, organism part (tissue, organ, corporal bodily fluids), tissue culture, cell culture, protoplasts, nucleic acids, proteins, etc.

  Amount of Bio-Source: Mass (mg fresh weight or mg dry weight), number of cells or other measurable bulk numbers (e.g. protein content).

  Developmental stage: This includes age and gender (if applicable) of the organism.

- **Growth conditions:**

  Growth substrates: Such as, hydroponic system (type, supplier, nutrients, concentrations), soil (type, supplier), agar (type, supplier), cell culture (media, volume, cell number per volume).

  Growth environment: Including but not limited to, controlled environment such as greenhouse (details on accuracy of control of light, humidity and temperature), housing conditions (light/dark cycle). Not controlled environment such as location of the field trial.

  Growing time: This refers to the growing time of the biomaterial prior to a treatment.

- **Sample pre-treatment or sample preparation:** This includes information about collection, transport, storage, preparation (e.g. drying, sieving, grinding, etc.) and preservation of the sample.

Laboratory equipment. The laboratory equipment includes apparatus and instruments that are used in diagnostic, surgical, therapeutic and experimental procedures. In this subsection, a list of all necessary equipment should be listed; manufacturer name or vendor (including the homepage), catalog number (or model) and configuration of the equipment should be included. See Table 13

Good example	**Name / manufacturer / model:** *"Inverted confocal microscope, PC and image acquisition software / Zeiss / LSM 780."*	**equipment configuration:** *"Configure a four-channel microscope with appropriate excitation light sources and emission filters: FITC-488 excitation, 490–560-nm emission; ..."*	Protocol available at (Lee et al., 2015)

Table 13

- **Laboratory equipment name:** Name of laboratory equipment (e.g. FocalCheck fluorescence microscope test slide.)

- **Manufacturer name:** A person, company, or entity that produces finished goods (e.g. Life Technologies, Zeiss)

- **Laboratory equipment ID (catalog number):** This is a identifier provided by the manufacturer or vendor (e.g. F36909 –catalog number for FocalCheck fluorescence microscope test slide from Life Technologies).

- **Equipment configuration:** This is the configuration of the equipment. The parameters that make it possible to carry out an operation, procedure, or task (e.g., the configuration of a inverted confocal microscope).

Laboratory consumables or supplies. The laboratory consumables include, amongst others, disposable pipettes, beakers, funnels and test tubes for accurate and precise measurement, disposable gloves and face masks for safety in the laboratory. In this subsection, a list of all the consumables necessary to perform the experiment should be listed with manufacturer name (including the homepage) and catalog number. See Table 14.

Bad example	Filter paper	Protocol available at (Zhang et al., 2008)
Good example	Filter paper (GE*, catalog number: 10311611)	Protocol available at (Cao et al., 2014)

Table 14. *name of the manufacturer.

- **Laboratory consumable name:** Names of laboratory consumable (e.g. Cryogenic Tube, sterile, 1.2 ml.)
- **Manufacturer name:** A person, enterprise, or entity that produces finished goods (e.g. Nalgene, Thermo-scientific, Eppendorf, Falcon)
- **Laboratory consumable ID (catalog number):** This is a identifier provided by the manufacturer or vendor (e.g. 5000-0012 (catalog number for Cryogenic Tube, sterile, 1.2 mL from Nalgene).

Recipe for solutions. A recipe for solutions is a set of instructions for preparing a particular solution, media, buffer, etc. The recipe for solutions should include the list of all necessary ingredients (chemical compounds, substance, etc.), initial and final concentrations, pH, storage conditions, cautions and hints. Ready-to-use reagents do not need to be listed in this category; all purchased reagents that require modification (e.g. a dilution or addition of β-mercaptoethanol) should be listed. See Table 15 for more information.

Bad example	See in the section recipes, the recipe 1 (PBS)	Protocol available at (Cao et al., 2014)
Good example	Phosphate-buffered saline (PBS) recipe	Protocol available at (Chazotte, 2012)

Table 15. Good and bad practices~~,~~ ~~about~~ the presentation of solution recipes.

- **Solution name:** This is the name of the liquid preparation that contains at least 2 chemical substances; one of them playing the role of solvent and the other playing the role of solute. If applicable, the name of the solution should include the following information: Solution concentration, final volume and final pH (e.g. Ammonium bicarbonate (NH4HCO3) (50 mM, 10 ml, pH 7.8))
- **Chemical compound name:** This is the name of a drug, solvent, chemical, etc., with a property that can be measured such as concentration (e.g. agarose, dimethyl sulfoxide (DMSO), phenol, sodium hydroxide).
- **Initial concentration of a chemical compound:** This is the first measured concentration of a compound in a substance.
- **Final concentration of chemical compound:** This is the last measured concentration of a compound in a substance.
- **Storage conditions:** This includes, amongst others, shelf life (maximum storage time) and storage temperature for the solution (e.g. “Store the solution at room temperature”, “maximum storage time, 6 months”). Specify whether solution must be prepared fresh.

- **Cautions:** Toxic or harmful chemical compounds should be identified by the word 'CAUTION' followed by a brief explanation of the hazard and the precautions that should be taken when handling (e.g. "CAUTION: NaOH is a very strong base. Can seriously burn skin and eyes. Wear protective clothing when handling. Make in fume hood").

- **Hints:** Commentaries or hints that help the researcher to correctly prepare the recipe (e.g. "Add NaOH to water to avoid splashing") should be provided.

Reagents. A reagent is a substance used in a chemical reaction to detect, measure, examine, or produce other substances. ~~Please list~~ all the reagents used when performing the experiment, the vendor name (including homepage) and catalog number. Reagents that are purchased ready-to-use should be listed in this section. See Table 16.

Bad example	Dextran sulfate, Sigma-Aldrich	Protocol available at (Karlgren et al., 2009)
Good example	Dextran sulfate sodium salt from *Leuconostoc spp.*, Sigma-Aldrich, D8906-5G	Protocol available at (Javelle et al., 2011)

Table 16. Good and bad ~~practices, about~~ the presentation of reagents.

- **Reagent name:** Name of the reagent (e.g. "Taq DNA Polymerase from Thermus aquaticus with 10X reaction buffer without MgCl2").

- **Reagent vendor or manufacturer:** This is the person, enterprise, or entity that produces chemical reagents (e.g. Sigma-Aldrich).

- **Reagent ID (catalog number):** This is a identifier provided by the manufacturer or vendor (e.g. D4545-250UN (catalog number for Taq DNA Polymerase from Thermus aquaticus with 10X reaction buffer without MgCl2 from Sigma-Aldrich)).

Kits. A ~~K~~it is ~~a~~ gear consisting of a set of articles or tools for a specific purpose. ~~Please list~~ all the kits used when carrying out the protocol, the vendor name (including homepage) and catalog number.

- **Kit name:** This is the name of the kit (e.g. Spectrum^{TM} Plant Total RNA Kit, sufficient for 50 purifications).

- **Kit vendor or manufacturer:** This is the person, enterprise, or entity that produces the kit (e.g. Sigma-Aldrich).

- **Kit ID (catalog number):** This is a identifier provided by the manufacturer or vendor (e.g. STRN50 (catalog number for Spectrum^{TM} Plant Total RNA Kit, sufficient for 50 purifications)).

Primer/Oligonucleotide. A primer is a synthetic short strand of nucleic acid that serves as a starting point for DNA or RNA synthesis. All primers used when performing the experiment, as well as vendor names (including homepage) should be listed.

- **Primer/Oligonucleotide name:** This is the name of a primer/oligonucleotide (e.g. CII-fw, CII-rv).

- **Primer/Oligonucleotide sequence:** Primer/oligonucleotide sequences are determined from known sequences, there must be a match to the region of DNA, mRNA or cDNA to be amplified (e.g. "5-GAATGATGTACCACCTTTG-3").

- **Primer/Oligonucleotide vendor or manufacturer:** This is the person, enterprise, or entity that produces primers (e.g. Operon).

- **Expected PCR product size:** This is the size of the amplicon that the primers will generate. The total size of the PCR product also includes the regions of primers.

Software. Software is composed of a series of instructions that can be interpreted or directly executed by a processing unit. In this sub-section, please list software used in the experiment including the version as well as where to get it.

- **Software name:** Name of the software (e.g. LightCycler 480 Software).
- **Software version:** A software version number is an attribute that represents the version of software (e.g. Version 1.5).
- **Software availability:** This should indicate where ~~could the software~~ be downloaded from. If possible, license information should also be included (e.g. https://github.com/MRCIEU/ariesmqtl, GPL3.0).

Metadata for the procedure

Some recommendations that improve the description of experimental instructions are listed below.

- Recommendation 1. Whenever possible, list the steps in numerical order; use active tense. For example: "Pipette 20 ml of buffer A into the flask" NOT: "20 ml of buffer A are/were pipetted into the flask" (Nature-Protocol, 2017).
- Recommendation 2. Whenever there are two (or more) alternatives, these should be numbered as sets of consecutive steps Current-Protocols (2017). For example: choose procedure A (steps 1-10) or procedure B (steps 11-20); then continue with step 21 and so on.
- Recommendation 3. For techniques that comprise a number of separate major procedures, organize these in the exact order in which they should be executed (Nature-Protocol, 2017).

Useful auxiliary information should be included after protocol steps in the form of "alert messages". The goal is to remind or alert the user of a protocol. These messages may cover special tips or hints for performing a step successfully, alternate ways to perform the step, cautions regarding hazardous materials or other safety conditions, time considerations (e.g., pause points, speed with which the step must be performed) and storage information (temperature, maximum duration) (Current-Protocols, 2017).

- **Critical steps:** Highlight critical steps in the protocol and give indications that help to carry these out in a precise manner e.g., if time and temperature are crucial, or the use of RNase free solutions is required, information should be provided in order to indicate how these steps are critical and how to overcome the issues. "Critical Steps" should help the user to maximize the likelihood of success; use the heading 'CRITICAL STEP', followed by a brief explanation. See Table 17.

Good example	**Step:** *"Remove dirt from the surface of the specimen with a tissue. If necessary, moisten the tissue with ..."*	**CRITICAL STEP:** *"Dirt may introduce a variety of inhibitory substances (...); these substances may interfere or even completely block subsequent enzymatic manipulations of the DNA extracts."*	Protocol available at (Rohland and Hofreiter, 2007)

Table 17

- **Pause point:** If the protocol naturally breaks into separate stages, include subheadings and resume the numbered list. Subheadings are particularly appropriate after steps in the protocol where the procedure can be stopped, i.e., when the experiment can be stopped and resumed at a later point in time. Any PAUSE POINTS should be indicated with a brief description of the options available. See Table 18.
- **Timing:** If possible, include the approximate time of execution of a step, or set of steps. Timing could also be indicated at the beginning of the protocol. See Table 19.

Good example	**Step:** *"Weigh out no more than 500 mg of sample powder and transfer it to a 15 ml tube."*	**PAUSE POINT:** *"The sample powder can be stored at room temperature, but should be subjected to the extraction as soon as possible."*	Protocol available at (Rohland and Hofreiter, 2007)

Table 18

Good example	**Procedure:** *"Preparation of the bone or tooth sample"*	**TIMING:** *"15–30 min per sample"*	Protocol available at (Rohland and Hofreiter, 2007)

Table 19

Good example	**Hint:** *"We tested several commercial thermostable DNA polymerases. In our hands, the most consistent results were obtained using Advantage 2 PCR Polymerase Mix (Clontech, Mountain View, CA)."*	Protocol available at (Varkonyi-Gasic et al., 2007)

Table 20

- **Hints:** Provide any commentary, note or hints that will help the researcher correctly perform the experiment. See Table 20.

- **Troubleshooting:** if known, list common problems, possible causes, and solutions/methods of correction. This can be submitted as a 3-column table or listed in the text. See Table 21.

Good example	See "Table 1.Troubleshooting table."	Protocol available at (Rohland and Hofreiter, 2007)

Table 21

DISCUSSION

We have described 18 metadata elements that can be used to improve the reporting structure of protocols. Our work is based on the analysis of 530 published and non published protocols, guidelines for authors, suggested reporting structures and available protocols in on line journals. Our guidelines had extensive input from a group of researchers whose primary interest is having reproducible protocols.

The quality of the information reported in experimental protocols and methods is a cause for concern; poorly described methods generate poorly reproducible research. In a study conducted by Flórez-Vargas et al. (2014) in Trypanosoma experiments, they reported that no article met all criteria that should be reported in these kind of experiments. The study reported by Kilkenny et al. (2009) has similar results leading to similar conclusions; key metadata elements are not always reported by researchers. Publishers, researchers and funders have started to address issues around adequate reporting; some journals are introducing the use of checklists and reporting structures (Nature, 2017); BioSharing, STAR, ARRIVE and other initiatives also illustrate this positive trend. Currently, the protocol is treated as any other scientific publication. However, the protocol is a particular type of publication, slightly different from any other scientific article. An experimental protocol is a document that is kept "alive" after it has been published. The protocols are routinary used in laboratory activities, researchers often improve and adapt them; for instance, by extending the type of samples that can be tested, reducing timing, minimizing the quantity of certain reagents without altering the results, adding new recipes, etc. The issues found in reporting methods probably stem, at least in part, from the current structure of scientific publishing, which is not adequate to effectively communicate complex experimental methods (Flórez-Vargas et al., 2014).

In laboratories, experimental protocols are released and, they periodically undergo revisions until they are released again. These documents follow the publication model put forward by Carole Goble, "*Don't Publish, release*" with strict versioning, changes and forks (Goble, 2017). Experimental protocols are

essentially executable workflows for which identifiers for equipment, reagents and samples need to be resolved against the Web. These workflows are currently followed by humans but in the future robots may be executing experiments (Yachie et al., 2017); it makes sense to investigate for these documents other publication paradigms. In the development of our checklist we considered the SMART Protocols ontology (Giraldo et al., 2014), OBI, EXACT and many other ontologies; our metadata elements can easily be mapped to ontologies and resources on the web such as PubChem (Kim et al., 2016) (Wang et al., 2017). Our checklist does not cover aspects inherent to each possible type of experiment such as those available in the Minimum Information for Biological and Biomedical Investigations (MIBBI) (MIBBI, 2017); these are based on the minimal common denominator for specific experiments. Both approaches complement each other; where MIBBIs offer specificity, the SP guidelines provide a context that is general enough for facilitating reproducibility and adequate reporting without interfering with records such as those commonly managed by Laboratory Information Management Systems. By the same token, our approach complement the ISA tools effort (Sansone et al., 2012).

Reporting guidelines are not an accepted norm in biology (MIBBI, 2017); however, experimental protocols are part of the daily activities for most biologists. They are familiar with these documents, the benefits of standardization are easy for them to understand. From our experience at the CIAT, once researchers were presented with a standardized format they could extend and manage with minimal overhead, they adopted it. The experience with the SP reporting structure was a gateway into MIBBIs that were applicable for the kind of experiments they were working on. By analyzing reporting structures and guidelines for authors we are contributing to the homogenization of data elements that should be reported as part of experimental protocols.

CONCLUSION

Improving reporting structures for experimental protocols requires collective efforts from authors, peer reviewers, editors and funding bodies. There is no "one size fits all"; the improvement will be incremental. Our guidelines are a step in this direction. Improving the reporting of experimental protocols will add a necessary layer of information that should accompany the data that is currently being deposited in data repositories. Moreover, experimental protocols should be mandatory supplementary material in biomedical publications; as these documents are workflows they should be machine readable so that software agents can automatically process them. Authors should be aware of the importance of experimental protocols in the research life cycle; experimental protocols ought to be reused and modified, derivative works are to be expected. This should be considered by authors before publishing their protocols; the terms of use and licenses are the choice of the publisher but, where to publish is the choice of the author. Terms of use and licenses forbidding "reuse", "reproduce", "modify" or "make derivative works based upon" should be avoided ~~-particularly for publishing this specific type of content.~~ Protocols represent the actual "know how" in the biomedical domain. Similarly, publishers should adhere to the principle of encouraging authors to have the protocols available; for instance as preprints or in repositories for protocols or journals. Publishers **SHOULD NOT** enforce the use of a particular repository or journal. For data, publishers require or encourage it to be available, same principle should be applied for protocols. Experimental protocols are ~~imprescindible~~ when reproducing or replicating an experiment; data is uncontextualized unless the protocols used to derive the data are available. Restrictions like the one mentioned here are an impediment for researchers to use the protocols in their most natural way, that is adapting and reusing them for different purposes –not to mention sharing, which is a common practice amongst researchers.

ACKNOWLEDGMENTS

OXG has been funded by the EU project DrInventor FP7-ICT-2013.8.1, AGC has been funded by the KOPAR project, H2020-MSCA-IF-2014, Grant Agreement nr: 655009.

REFERENCES

4TU (2017). 4tu, centre for research data. Retrieved on 07/07/2017 from http://researchdata.4tu.nl/en/home/.

Abeyruwan, S., Vempati, U. D., Küçük-McGinty, H., Visser, U., Koleti, A., Mir, A., Sakurai, K., Chung, C., Bittker, J. A., Clemons, P. A., Brudz, S., Siripala, A., Morales, A. J., Romacker, M.,

Twomey, D., Bureeva, S., Lemmon, V., and Schürer, S. C. (2014). Evolving BioAssay ontology (BAO): modularization, integration and applications. *Journal of Biomedical Semantics*, 5(Suppl 1 Proceedings of the Bio-Ontologies Spec Interest G):S5.

Assante, M., Candela, L., Castelli, D., and Tani, A. (2016). Are Scientific Data Repositories Coping with Research Data Publishing? *Data Science Journal*, (15):p.6.

Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M. H., Bug, B., Chibucos, M. C., Clancy, K., Courtot, M., Derom, D., Dumontier, M., Fan, L., Fostel, J., Fragoso, G., Gibson, F., Gonzalez-Beltran, A., Haendel, M. A., He, Y., Heiskanen, M., Hernandez-Boussard, T., Jensen, M., Lin, Y., Lister, A. L., Lord, P., Malone, J., Manduchi, E., McGee, M., Morrison, N., Overton, J. A., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R. H., Schober, D., Smith, B., Soldatova, L. N., Stoeckert, C. J., Taylor, C. F., Torniai, C., Turner, J. A., Vita, R., Whetzel, P. L., and Zheng, J. (2016). The Ontology for Biomedical Investigations. *PLOS ONE*, 11(4):e0154556.

Belide, S., Hac, L., Singh, S. P., Green, A. G., and Wood, C. C. (2011). Agrobacterium-mediated transformation of safflower and the efficient recovery of transgenic plants via grafting. *Plant Methods*, 7(1):12.

Bio-protocol (2017). Bio-protocol LLC, Instructions for Authors. Retrieved on 07/07/2017 from https://goo.gl/oCP2YG.

Biosharing (2017). Biosharing. Retrieved on 07/07/2017 from https://biosharing.org/standards/.

bioTechniques (2017). The International Journal of Life Sciences Methods, instructions for authors. Retrieved on 07/07/2017 from https://goo.gl/is9aM1.

Blazquez, M. (2007). Quantitative gus activity assay in intact plant tissue. *Cold Spring Harbor Protocols*, 2007(2):pdb.prot4688.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., and Wittwer, C. T. (2009). The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry*, 55(4):611–622.

Cao, J., Zhu, X., and Yan, X. (2014). Fluorescence microscopy for cilia in cultured cells and zebrafish embryos. *Bio-protocol*, 4(14).

Chazotte, B. (2012). Labeling golgi with fluorescent ceramides. *Cold Spring Harbor Protocols*, 2012(8):pdb.prot070599.

CIAT (2017). International Center for Tropical Agriculture (CIAT). Retrieved on 07/07/2017 from https://ciat.cgiar.org/.

CSH-Protocols (2017). Cold Spring Harbor Protocols,Instructions for Authors. Retrieved on 07/07/2017 from http://cshlpress.com/cshprotocols/.

CSIRO (2017). The commonwealth scientific and industrial research organisation data access portal. Retrieved on 07/07/2017 from https://data.csiro.au.

Current-Protocols (2017). Current Protocols, The Fine Art of Experimentation. Instructions for Authors. Retrieved on 07/07/2017 from http://www.currentprotocols.com/WileyCDA/Section/id-810273.html.

Das, B. (2011). Validation protocol: First step of a lean-total quality management principle in a new laboratory set-up in a tertiary care hospital in india. *Ind J Clin Biochem*, 26(3):235–243.

Dryad (2017). Dryad. Retrieved on 07/07/2017 from http://datadryad.org/.

FDA (2017). Food and Drug Administration, White paper: FDA Guidance for Industry Update – Process Validation. Retrieved on 07/07/2017 from https://goo.gl/SRf6YV.

Federhen, S. (2015). Type material in the NCBI Taxonomy Database. *Nucleic Acids Res*, 43:D1086–98.

figshare (2017). figshare. Retrieved on 07/07/2017 from http://figshare.com.

Flórez-Vargas, O., Bramhall, M., Noyes, H., Cruickshank, S., Stevens, R., and Brass, A. (2014). The quality of methods reporting in parasitology experiments. *PLoS ONE*, 9(7).

Gibson, F., Hoogland, C., Martinez-Bartolomé, S., Medina-Aunon, J. A., Albar, J. P., Babnigg, G., Wipat, A., Hermjakob, H., Almeida, J. S., Stanislaus, R., Paton, N. W., and Jones, A. R. (2010). The gel electrophoresis markup language (gelml) from the proteomics standards initiative. *PROTEOMICS*, 10(17):3073–3081.

Giraldo, O. (2017). Corpus of experimental protocols. Retrieved on 07/07/2017 from https://goo.gl/88H71c.

Giraldo, O., Garcia, A., and Oscar, C. (2014). SMART Protocols: SeMAntic RepresenTation for Experimental Protocols. In *Proceedings of the 4th Workshop on Linked Science 2014*, Riva del Garda,

Trentino, Italy. 4th Workshop on Linked Science 2014- Making Sense Out of Data (LISC2014).
GMR (2017). Genetics and Molecular Research. Retrieved on 07/07/2017 from http://www.geneticsmr.com/.
Goble, C. (2017). DON'T PUBLISH. RELEASE!. Retrieved on 07/07/2017 from https://www.force11.org/presentation/dont-publish-release.
Godlee, F. (2001). Publishing study protocols: making them visible will improve registration, reporting and recruitment. In *BMC News and Views*.
Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., and Steinbeck, C. (2013). The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res*, 41:D456–63.
IAO (2017). The Information Artifact Ontology (IAO). Retrieved on 07/07/2017 from https://goo.gl/5E8myD.
Javelle, M., Marco, C. F., and Timmermans, M. (2011). In situ hybridization for the precise localization of transcripts in plants. *Journal of Visualized Experiments : JoVE*, (57):3328.
JBM (2017). Journal of Biological Methods, Instructions for Authors. Retrieved on 07/07/2017 from https://goo.gl/yFB2tY.
JOVE (2017). Journal of Visualized Experiments, Instructions for Authors. Retrieved on 07/07/2017 from https://goo.gl/FjAEv4.
Karlgren, A., Carlsson, J., Gyllenstrand, N., Lagercrantz, U., and Sundström, J. F. (2009). Non-radioactive in situ hybridization protocol applicable for norway spruce and a range of plant species. *Journal of Visualized Experiments : JoVE*, (26):1205.
Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., and Altman, D. G. (2010a). Improving bioscience research reporting: the arrive guidelines for reporting animal research. *PLoS Biol*, 8(6):e1000412.
Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., and Altman, D. G. (2010b). Improving bioscience research reporting: The arrive guidelines for reporting animal research. *PLoS Biology*, 8(6):e1000412.
Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, M. F. W., Cuthill, I. C., Fry, D., Hutton, J., and Altman, D. G. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS ONE*, 4(11):e7824.
Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., and Bryant, S. H. (2016). PubChem Substance and Compound databases. *Nucleic Acids Research*, 44(D1):D1202–D1213.
Kim, Y.-C. and Morrison, S. L. (2009). A rapid and economic in-house dna purification method using glass syringe filters. *PLoS ONE*, 4(11):e7750.
Lee, J. A., Spidlen, J., Boyce, K., Cai, J., Crosbie, N., Dalphin, M., Furlong, J., Gasparetto, M., Goldberg, M., Goralczyk, E. M., Hyun, B., Jansen, K., Kollmann, T., Kong, M., Leif, R., McWeeney, S., Moloshok, T. D., Moore, W., Nolan, G., Nolan, J., Nikolich-Zugich, J., Parrish, D., Purcell, B., Qian, Y., Selvaraj, B., Smith, C., Tchuvatkina, O., Wertheimer, A., Wilkinson, P., Wilson, C., Wood, J., Zigon, R., Scheuermann, R. H., and Brinkman, R. R. (2008). Miflowcyt: The minimum information about a flow cytometry experiment. *Cytometry Part A*, 73A(10):926–930.
Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Ferrante, T. C., Terry, R., Turczyk, B. M., Yang, J. L., Lee, H. S., Aach, J., Zhang, K., and Church, G. M. (2015). Fluorescent in situ sequencing (fisseq) of rna for gene expression profiling in intact cells and tissues. *Nat. Protocols*, 10(3):442–458.
Linke, B., Schröder, K., Arter, J., Gasperazzo, T., Woehlecke, H., and Ehwald, R. (2010). Extraction of nucleic acids from yeast cells and plant tissues using ethanol as medium for sample preservation and cell disruption. *BioTechniques*, 49(3):655–657.
Lu, Y. (2011). Extract genomic dna from arabidopsis leaves (can be used for other tissues as well). *Bio-protocol*.
Malone, J., Holloway, E., Adamusiak, T., Kapushesky, M., Zheng, J., Kolesnikov, N., Zhukova, A., Brazma, A., and Parkinson, H. (2010). Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics*, 26(8):1112–1118.
Marcus, E. (2016). A STAR Is Born. *Cell*, 166(5):1059–1060.
McQuilton, P., Gonzalez-Beltran, A., Rocca-Serra, P., Thurston, M., Lister, A., Maguire, E., and Sansone, S.-A. (2016). Biosharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences. *Database: The Journal of Biological Databases and Curation*, 2016:baw075.
MethodX (2017). Method X, Instructions for Authors. Retrieved on 07/07/2017 from

https://www.elsevier.com/journals/methodsx/2215-0161/guide-for-authors.
MIACA (2017). Minimum Information About a Cellular Assay (MIACA). Retrieved on 07/07/2017 from http://miaca.sourceforge.net/.
MIAPPE (2017). Minimum Information about Plant Phenotyping Experiment. Retrieved on 07/07/2017 from https://goo.gl/oVEhhe.
MIARE (2017). Minimum Information About an RNAi Experiment. Retrieved on 07/07/2017 from http://miare.sourceforge.net/HomePage.
MIBBI (2017). Minimum Information for Biological and Biomedical Investigations. Retrieved on 07/07/2017 from https://biosharing.org/collection/MIBBI.
Mungall, C. J., Torniai, C., Gkoutos, G. V., Lewis, S. E., and Haendel, M. A. (2012). Uberon, an integrative multi-species anatomy ontology. *Genome Biology*, 13(1):R5.
Nature (2017). Reporting checklist for life sciences articles. by nature publishing group. Retrieved on 07/07/2017 from https://goo.gl/YykR51.
Nature-Protocol (2017). Nature Protocol, Instructions for Authors. Retrieved on 07/07/2017 from http://www.nature.com/nprot/info/gta.html.
NIH (2017a). National Institute of Dental and Craniofacial Research from NHI, Version Control Guidelines. Retrieved on 07/07/2017 from http://www.nidcr.nih.gov/Research/ToolsforResearchers/Toolkit/VersionControlGuidelines.htm.
NIH (2017b). Principles and guidelines for reporting preclinical research. Retrieved on 07/07/2017 from https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research.
Nikolau, B., Fiehn, O., the participants of the 2006 Plant, Conference, M., Rhee, S., Dickerson, J., Lange, M., Lane, G., Roessner, U., Ward, J., Last, R., and Chapple, C. (2006). CIMR: Plant Biology Context Metabolomics Standards Initiative (MSI). Retrieved on 07/07/2017 from http://cosmos-fp7.eu/system/files/presentation/plant.pdf.
NPE (2017). Nature Protocol Exchange. Retrieved on 07/07/2017 from http://www.nature.com/protocolexchange/.
ORCID (2017). ORCID — Connecting Research and Researchers. Retrieved on 07/07/2017 from https://orcid.org/.
Plant-Methods (2017). Plant Methods. Retrieved on 07/07/2017 from http://plantmethods.biomedcentral.com/.
Plos-One (2017). Plos One. Retrieved on 07/07/2017 from http://journals.plos.org/plosone/.
ResearcherID (2017). ResearcherID. Retrieved on 07/07/2017 from http://www.researcherid.com/.
RIP (2017). Resource identification portal. Retrieved on 07/07/2017 from https://scicrunch.org/resources.
Rohland, N. and Hofreiter, M. (2007). Ancient dna extraction from bones and teeth. *Nat. Protocols*, 2(7):1756–1762.
Rosas-Cárdenas, F., Durán-Figueroa, N., Vielle-Calzada, J.-P., Cruz-Hernández, A., Marsch-Martínez, N., and de Folter, S. (2011). A simple and efficient method for isolating small rnas from different plant species. *Plant Methods*, 7(1):4.
Sansone, S.-A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, C., Hofmann, O., Fang, H., Neumann, S., Tong, W., Amaral-Zettler, L., Begley, K., Booth, T., Bougueleret, L., Burns, G., Chapman, B., Clark, T., Coleman, L.-A., Copeland, J., Das, S., de Daruvar, A., de Matos, P., Dix, I., Edmunds, S., Evelo, C. T., Forster, M. J., Gaudet, P., Gilbert, J., Goble, C., Griffin, J. L., Jacob, D., Kleinjans, J., Harland, L., Haug, K., Hermjakob, H., Sui, S. J. H., Laederach, A., Liang, S., Marshall, S., McGrath, A., Merrill, E., Reilly, D., Roux, M., Shamu, C. E., Shang, C. A., Steinbeck, C., Trefethen, A., Williams-Jones, B., Wolstencroft, K., Xenarios, I., and Hide, W. (2012). Toward interoperable bioscience data. *Nat Genet*, 44(2):121–126.
Sarntivijai, S., Lin, Y., Xiang, Z., Meehan, T. F., Diehl, A. D., Vempati, U. D., Schürer, S. C., Pang, C., Malone, J., Parkinson, H., Liu, Y., Takatsuki, T., Saijo, K., Masuya, H., Nakamura, Y., Brush, M. H., Haendel, M. A., Zheng, J., Stoeckert, C. J., Peters, B., Mungall, C. J., Carey, T. E., States, D. J., Athey, B. D., and He, Y. (2014). Clo: The cell line ontology. *Journal of Biomedical Semantics*, 5:37–37.
Sarntivijai, S., Xiang, Z., Meehan, T., Diehl, A., Vempati, U., Schurer, S., Pang, C., Malone, J., Parkinson, H., Athey, B., and He, Y. (2011). Cell line ontology: Redesigning the cell line knowledgebase to aid integrative translational informatics. *Neoplasia*, 833:25–32.
Simmhan, Y. L., Plale, B., and Gannon, D. (2005). A survey of data provenance in e-science. *SIGMOD*

Rec., 34(3):31–36.

Soldatova, L. and King, R. (2006). An ontology of scientific experiments. *Journal of the Royal Society Interface*, 3(11):795–803.

Soldatova, L. N., Aubrey, W., King, R. D., and Clare, A. (2008). The EXACT description of biomedical protocols. *Bioinformatics*, 24(13):i295–i303.

Soldatova, L. N., Nadis, D., King, R. D., Basu, P. S., Haddi, E., Baumlé, V., Saunders, N. J., Marwan, W., and Rudkin, B. B. (2014). EXACT2: the semantics of biomedical protocols. *BMC Bioinformatics*, 15(Suppl 14):S5.

Springer-Protocols (2017). Springer Protocols, Instructions for Authors. Retrieved on 07/07/2017 from goo.gl/7qgWUE.

STAR (2017). Structured, transparent, accessible reporting, the cell star methods guide for authors. Retrieved on 07/07/2017 from http://www.cell.com/star-authors-guide.

Torniai, C., Brush, M., Vasilevsky, N., Segerdell, E., Wilson, M., Johnson, T., Corday, K., Shaffer, C., and Haendel, M. (2011). *Developing an application ontology for biomedical resource annotation and retrieval: Challenges and lessons learned*, volume 833, pages 101–108. CEUR Workshop Proceedings.

Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E. F., and Hellens, R. P. (2007). Protocol: a highly sensitive rt-pcr method for detection and quantification of microrna. *Plant Methods*, 3(1):12.

Vingataramin L, F. E. (2015). A single protocol for extraction of gdna from bacteria and yeast. *BioTechniques*, 58(3):120–125.

Wang, Y., Bryant, S. H., Cheng, T., Wang, J., Gindulyte, A., Shoemaker, B. A., Thiessen, P. A., He, S., and Zhang, J. (2017). Pubchem bioassay: 2017 update. *Nucleic Acids Research*, 45(D1):D955.

Whetzel, P. L., Noy, N. F., Shah, N. H., Alexander, P. R., Nyulas, C., Tudorache, T., and Musen, M. A. (2011). BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic acids research*, 39(Web Server issue):W541–5.

Xiang, Z., Mungall, C., Ruttenberg, A., and He, Y. (2011). Ontobee: A Linked Data Server and Browser for Ontology Terms. In *2nd International Conference on Biomedical Ontologies (ICBO)*, pages 279–281, Buffalo, NY, USA. 2nd International Conference on Biomedical Ontologies (ICBO), July 28-30.

Yachie, N., Consortium, R. B., and Natsume, T. (2017). Robotic crowd biology with maholo labdroids. *Nat Biotech*, 35(4):310–312.

Zenodo (2017). Zenodo. Retrieved on 07/07/2017 from https://zenodo.org/.

Zhang, W., Nilson, S. E., and Assmann, S. M. (2008). Isolation and whole-cell patch clamping of arabidopsis guard cell protoplasts. *Cold Spring Harbor Protocols*, 2008(6):pdb.prot5014.