

Σ.I.VL¹ Introduction

There are a number of ways to share and preserve scientific data. These range in data-density from a simple citation in a research article, to a relatively lightweight abstract, to the more informative journal article, through to vast databases of scientific data. An emerging area within this spectrum is the Supplement, enables further discovery by tying the published results to their underlying data and providing a level of structure to enable re-usability by the community. However, in the spectrum of scientific information density, supplements tend to fall far from the vernacular friendly journal article and more toward the somewhat unreadable raw data.

Research papers are limited by space constraints of journals. Often the more prestigious the journal, the more space constraints are forced on the authors. In a similar vein, the prestige of the journal can be used as leverage against an author to force them to otherwise present their data in a sub-optimal fashion. Scientific journals create standardized headings that are used by all papers, regardless of the field.

As a result of these and other limitations, and perhaps also in an effort to include research that has not yet been published elsewhere but needs to be presented in order to for the authors to claim their conclusions before their peers, many have turned to supplements as an alternative to publishing. Supplements they do not fit easily within the standards of the scientific publishing world, are not always peer reviewed, and often not indexed.¹

Thus, although much of scientific literature is highly structured and professionally indexed, supplements are a growing repository of scientific information that lack both the structure and the indexability of the standard literature. At minimum this results in lost and hard to find data. At worst, it allows for non-peer reviewed or even problematic data to leach into the scientific record through an unguarded back-door.

This is a problem that this paper set out to solve.

Σ.I.HL² Introduction

As supporting materials become regular components of this record, questions are emerging as to the archiving of these materials. The often take varied forms, such as data, software, algorithm descriptions, figures, or additional methods and text, just to name a few. Each of these modalities requires documentation, curation, archiving, and preservation, each of which may be unique to the medium.

As such, while these “mesoscale” journal article supplements have become an increasingly indispensable resource for research, both for presenting the full extent of their research, but also providing documentation and even repositories of scientific information and data, they are also failing to adequately do this job.

N b

Comment [BC-SSU1]: What do you think about just titling this “Introduction (General Audience)” instead of using the footnotes? We would want to make sure that the readers understand that there are two introductions here in the supplement.

- Deleted:** In general,
- Deleted:** there
- Deleted:** data-heavy
- Deleted:** which can serve to
- Deleted:** enable discoverability

- Deleted:** the
- Deleted:** Further in many cases, scientific

- Deleted:** have been termed gray literature, as
- Deleted:** typically not

- Deleted:** a lot of lost of hard to find data

Comment [BC-SSU2]: And then titling this “Introduction: Technical Level (or similar)”

¹ Vernacular language (General Audience)
² High Level

Given ~~the essential nature of the Supplement~~ for the progress of science and the ~~general lack of coherent organization~~, there are recent efforts to develop substantial journal oversight in this heretofore unstructured area of publication. In this article, we propose a novel structuring of the Supplement section to bring coherence to this vital information source, ~~and enable scholarly activities such as the verification of findings, re-use of data or software,~~ ~~and a more complete exposition of methods.~~

Deleted: their

Deleted: ir

Deleted: ,

Deleted: or

Σ.I.A The Supplement Section Today

This section, as denoted by the capital A, does not directly parallel the primary text. However, it does provide additional introductory material that might be of interest to the reader and is as such included within the supplement. This section is also of use to the reader as it introduces terminology that might be usefully elucidated in a glossary section, including the termination of integral content vs. additional and associated content.

Online publication supplements can provide an important and dedicated space for related and relevant information that simply won't fit within the limitations of a particular print or online publication.^{2 3} This includes information which may be *integral content*;⁴ for example, in addition to allowing for more text, supplements also provide space for oversized items such as tables, equations, figures, and high resolution images or even unconventional items such as multimedia.

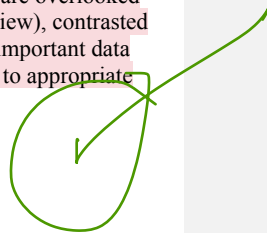
In other instances, the supplement further allows for the inclusion of *additional or associated content*; i.e., material that typically falls outside of standard presentation formats or their publication conventions and that serves to provide context and further relevant explanations or background. These materials may include clarifying notes, data, software and its accompanying notation, workflows, failed experiments or negative results, and additional multimedia content.^{5 6}

Given the massive scale of many current scientific efforts, the ability to reproduce and verify research results requires access to content not typically found in a journal article. This information could be contained in the supplemental materials, and could include supporting data and information relating to workflow and computational efforts. In particular, workflow and computational details frequently represent scientific decision making and assumptions that, if they were open to scrutiny, could improve the scientific process and allow follow-on researchers to better extend the results of the originally published research.

Deleted: Additionally, supplementary material potentially plays a growing role in the verification and reproduction process of published results. The veracity of science depends crucially on the availability and usability of the underlying raw data, and a sharing of documented computational steps that lead from the raw data through to the finished published output.

Finally, the use of supplements to provide access to the underlying raw data will become ever more relevant as supplements are used to fulfill journal requirements for the disclosure of that underlying data.^{7 8}

Given all these potential benefits, the unfortunate degree to which supplements are overlooked by journals, (e.g., lacking editing, polishing and often even substantive peer review), contrasted with their extensive use by the scientific community as an additional source of important data and commentary, is particularly troubling in light of growing efforts by authors to appropriate



this space as an important component of the grey literature, and particularly given the intensifying use of text mining methods and machine learning algorithms to analyze ever-increasing amounts of data.⁹

Comment [BC-SSU3]: This sounds a bit repetitive to me. What would you think about removing this paragraph?

We believe that, with much of this information residing in its current unstructured state, vast amounts information stored in and represented by the publication may be undiscoverable, unusable, and unintentionally ignored. As such, supplements need to be elevated to a standard publication form of research dissemination.

Supplements remain a form of publication, and like the corresponding primary paper, supplements inherently need to be representable in the standard publishable form: on the printed paper. While some aspects proposed herein may seem to less than optimally presented in a printed vs digital format, they all remain printable.

Deleted: Relatedly, supplements, at the end of the day,

Deleted: its

Deleted: s,

Deleted: we believe

E.I.A.1 Glossary

This section provides a glossary for the terminology that is provided in the previous section.

Comment [BC-SSU4]: I wonder if there is a way to distinguish when you are describing the supplement from where you actually create it. Perhaps Italics?

Integral content relates to data that optimally should be included within the primary text if not for space limitations, additional or associated content is material that typically falls outside of standard presentation formats, including, data provenance, glossaries such as this one, background information, workflows and software-related information.

Deleted: .

Deleted: that may be of interest to the wholly uninitiated

Deleted: software

E.I.B Supplements are now an important component of modern science

Comment [BC-SSU5]: I'm a little confused by these next two sections of the introduction. Are A, B, and C meant to be three different "introduction" aspects? I think this section B is repetitive with previous sections, and I'm wondering if you would move section C to B, instead, and have 2 sections here. I think this would still show the structure of the supplemental introduction.

E.I.C Current Concerns with Supplements

Notwithstanding the many positive aspects of supplements, many journals find the size and nature of these supplements overwhelming. Some publishers are even calling for curbs in their use^{11 12}. Here, the supplement provides more citations than what are provided for the in primary paper.^{13 14 15 16 17} and it can also include atypical sources such as blogs and webpages.

Deleted: Their relegation to the metaphorical back of the journal notwithstanding, supplements ought to be an essential component of the modern scientific publication and data archiving efforts. Many, if not most articles in popular journals, and especially most genomics journals, include potentially useful if not necessary supplemental materials. ... [1]

Deleted: Notable for this supplement to the primary paper, it can also include citations to atypical (e.g., not scholarly sources, such as blogs and webpages) .

Deleted: To some degree this is expected: currently supplements

Supplements often contain a tremendous amount of data, facts, and analysis associated, sometimes tenuously, with their corresponding published papers. Standardization with help prevent this additional information from getting lost. In some instances, references have been made to the Wild West in characterizing the current status quo for supplemental material;¹⁹ for example, with some otherwise short papers including supplemental materials nearly 30 times their length.^{20 21} We believe that these and other issues can be addressed with a more considered approach to supplemental materials, to be described herein.

Deleted: Lior Pachter has elegantly described these missed opportunities in his "Stories from the Supplement" lecture¹⁸ where entire ideas, which can be quite deep, are often contained entirely within the supplement and difficult to find from the main text. In addition to the various scientific reasons described above, efforts

Comment [BC-SSU6]: Combine these two paragraphs?

Deleted: In addition to the various scientific reasons described

Deleted: and cabin

Deleted: ;

Deleted: given

Efforts to rein in supplements is necessary on a more practical level. As a result of the often disorganized nature of current supplements, authors regularly cram as much information as possible into the actual main text of the document. This neglect of supplements and their scientific potential can result in making the main text very unreadable through overloading the

Deleted: writers are often loathe to add important information down the rabbit hole

Deleted: Perhaps because of this,

Deleted: At the very least, this

limited space with too much data in lieu of vernacular text, and writing that limited vernacular text very tersely. However, supplements often lack extensive editing and mincing, making finding relevant data even more difficult to an outside reader.

Deleted: ,
Deleted: to some degree, these fears are founded:
Deleted: this
Deleted: ; they tend to be poorly edited and often make

Even with all these concerns, many journals support and even promote the extensive usage of supplements.²² Broad efforts, such as this one, continue to be made to establish a set of best practices to address a number of aspects related to supplemental material.²³

Σ.II. Proposal

Supplements have the potential to provide substantial clarity to the published text, not only by providing much needed annotation, but by also providing additional information and data. Even though the supplement will likely never be as precise or as defined as the main text, considerable improvements need to be made across the board. In particular, without the constraints of space, online supplemental material can afford to be clearly written, better organized, and well-documented, allowing for an expanded and useful representation of the research and its results.

Comment [BC-SSU7]: Do you mean to have this section the same as the primary paper? I think you can probably take out this section. I also wonder if you want to move sections II.TL and II.CP to the end of this supplement section, and start with section II.A so that it more closely follows the primary text?

To this end, universally accepted structures and standards will substantially expand the usefulness of supplemental materials. For example, with an indexed, searchable, and useful supplement, authors need not jam as much into the main text of the paper – resulting in a more coherent and readable main text. Notably, both the published paper and its supplement can benefit from tying each section in the main text to its corresponding expanded and expounded supplement section; i.e., tethering together text with its corresponding raw data and related information through an established, logical, and linked hierarchy.

Σ.II.TL³ Proposal

1. DOIs, micro-referencing, and hierarchical headings.
2. Supplements should fit within the research stack an archived of scientific record.
3. Supplements should include workflows, data verification and provenance and should be curated.
4. Supplements should be designed to be findable, crawlable and readable by indexes such as Web of Science, Pubmed and Google Scholar.
5. Supplements should follow FAIR Standards

Comment [BC-SSU8]: Again, I wonder if we can write this out instead of using the footnotes here.

Σ.II.CP⁴ Proposal

1. Attribution: DOI; micro-referencing;
2. Design: Hierarchical headings.
3. Elements: workflows; data verification and provenance; subsections
4. Search: findable; crawlable; readable; indexed
5. Qualities: FAIR; readable; structured;printable

Σ.II.A Our Proposal: Standardizing the Supplemental Materials Section for Genomic Research Articles

Comment [BC-SSU9]: Should this be the same title as in the primary MS?

³ Technical language

⁴ Computer Parsable language

Best practices for supplements ought to be designed to deal with the above-mentioned concerns, as well as other pertinent issues particular to supplemental material. These best practices should include guidelines relating size and format (including documentation), scope, persistence, and accessibility of supplemental material.

Additional best practices should relate to the curatorial responsibility of journals, focusing on remedying the general lack of peer review, lack of discoverability, and inability to cite substantial portions of the information found only in the supplementary materials.

Ours is not the first effort to suggest better administration of supplements.^{24 25 26 27}

However, a number of concerns specific to genomic-oriented journals have been overlooked, particularly in the areas of interoperability, interpretability, reusability, organization, versioning, granularity in large dynamic data sets, and overall standards. With the growing relevance and importance of supplemental materials in genomic research, we propose a number of additional changes that can be employed in publishing supplements to help make the data and information published therein more useful for the researcher.

With a recognized and useful supplement, such as this one, authors need not jam as much raw data and tenuously related information into the paper, and as such, the main text can be made all the more readable. This is particularly the case if each section and subsection in the main text can be directly tied to the corresponding expanded section or subsection, through an established, logical, and linked hierarchy.

Even though the supplement will likely never be as refined a document as the main text, improvements can be made.

Σ.II.a The FAIR Standards: Findable, Accessible, Interoperable, and Reusable

As provided in the primary paper, the recent FAIR approach for scientific information that relates to both human and machine analysis of presented data.²⁸ Noteworthy of this supplement, we provide additional citations than what have been provided in the primary text. For example with regard to the FAIR standards, see, also^{29 30 31 32 33} See, also similar ideas:^{34 35}

Succinctly, under this paradigm, scientific data in supplementary material should be: Findable, Accessible, Interoperable and Reusable.

Data should be findable both for human researchers as well as computers, requiring unique and persistent identifiers for example, as provided by groups such as CASRAI³⁶ for the data and its associated components, such as metadata and documentation. Data ought to also be accessible. Here accessibility relates mainly to good data stewardship, and in particular, data, code, and workflow information should be stored long term and legally accessible via appropriate open licensing and other methods necessary for non-inhibiting access.^{37,38}

Comment [BC-SSU10]: What do you mean by this here?

Deleted: g:

Deleted: <#>size and format (including documentation) ... [2]

Formatted: Font:(Default) +Theme Headings CS (Times New Roman), 12 pt

Deleted: genomic

Deleted: As such, with

Deleted: This supplement provides an expanded version of the propose structural suggestions from the primary text as well as additional information relating to suggested standardizations that will be valuable in optimizing the usefulness of supplemental materials. [3]

Deleted: , for among other things, said raw data and tenuously related information, in the supplement,

Deleted: In particular, without the constraints of space on a published page, online supplemental material can afford to be clearly written and well-documented, allowing for an expanded and better defined representation of the research and results. .

Comment [BC-SSU11]: If we reorder the primary text as suggested, we should reorder these sections as well.

MORE

Accessibility also relates to making the underlying software code also accessible. However, while supplemental material should always strive to provide all the relevant information in one place, including a snapshot of the version of the software code used for the analysis, subsequent and further evolving versions of the code should be linked to, perhaps even indexed, but stored separately, perhaps on a specialty site such as GitHub or BitBucket, provided that adequate metadata is included at these sites such that discovery of relevant software is not inhibited.³⁹ Versioning is vital for reusable and changeable objects like data and software, and Digital Object Identifiers (DOIs) should be assigned to all data, code, and workflow information associated with the published findings.⁴⁰

Deleted: , often necessary for evaluating the paper's analysis,

Deleted: -

Data stored in supplements should also be interoperable, as human readers need to clearly understand the connection of the data to the main text. Readers should be able to appreciate the nature of the data from the presentation of the data and how it can be combined or compared with other datasets. Interoperability requires that the data also be easily digested by computational systems, using a standard that allows for straightforward data manipulation.

Deleted: ;

Deleted: Further, readers

Deleted: i.e.,

Deleted:

Deleted: able to

Deleted: e.g., in

Deleted: For example, both

Deleted: either

Deleted: or

Finally, data needs to be reusable. Both humans and machines should be able to apply the data to follow-up research and additional computational analysis.

Σ.II.b Provenance

There remains a need for veracity and verifiability of research data. The provenance of data refers to a complete description of the origins of the data, as well as the process by which that data arrived in its current database and current form (by conversions, normalizations, etc.) and data should be tracked as they are collected and repackaged in subsequent research.⁴¹ Provenance is highly relevant to assessing data quality, which can often be estimated based on the source of that information, providing an audit trail that will allow for an appreciation of the resource usage in putting together the dataset as well as the locating the potential source of any errors in the data, providing the location of all the data relevant for replication of the results, and attribution of the resulting data and conclusions. This is an important issue for assessing ownership, copyright rights, license limitations, any privacy restrictions, and liability, if any, ascribed to erroneous data.

Deleted: Associated with being able to replicate the original workflow, there

Deleted: , in some instances, provenance of data, i.e.,

Deleted: e.g.,

Deleted: ect

Deleted: : (i)

Deleted: ;

Deleted: (ii)

Deleted: ; (iii)

Deleted: ; and (iv)

Deleted: ,

Deleted: in

Σ.II.c Workflows

Supplementary material should be designed to incorporate workflow-related information. For example, the Supplement can outline in depth the individual and collective workflows that resulted in the eventual dataset and the published conclusions.⁴² Workflows are especially relevant for *in-silico* analyses, as the exact particulars and parameters employed in a workflow can make all the difference between reproducible and non-reproducible data. In this regard, supplemental data should include both abstract versions of workflows, as well as flowcharts or similar representations of the actual executed workflows as they relate to the particular code and execution infrastructure of the lab conducting the research.⁴³

Deleted: workflow

Deleted: outlining

Workflows should be directly linked to specific figures and files associated with the paper so that subsequent researchers can review and analyze what transformations, analysis, or other manipulations have already been done to a data set. Similarly, subsequent researchers can understand the implemented processes that resulted in the figure, from raw data to processed data, to a supplemental table of the processed table, to a figure in the primary paper, and finally to the text describing that figure.

- Deleted: , and
- Deleted: s
- Deleted: so
- Deleted: For example, a workflow should be able to trace the process from raw data to processed data,

Workflows should also have their own standardized identifiers, such that those identifiers include references to the relevant datasets associated with the workflow, any relevant software applied to the workflow, dates that further help to describe the version of the data and the software, and any other relevant information that could be used to cross-reference different datasets and their associated workflows. In the alternative, third party solutions such as Galaxy could be used to organize workflows.⁴⁴ The supplement can include links to such solutions.⁴⁵

5.II.d. Language in the Supplement

A key aspect of scientific writing is language. The nature of scientific progress and the evolution of myriad micro-disciplines have resulted in scientific writing that can be difficult to understand. To some degree, this jargon-filled language can be justified as it offers the necessary precision to properly present research, reproduce a result, and for effectively automatically parsing through text. On the other hand, the broader scientific community would likely appreciate a simpler, more vernacular language that's easier for a more generalized audience to understand and potentially more communicative; allowing for cross-disciplinary fertilization and perhaps better reflecting the multidisciplinary nature of many current scientific efforts.

- Deleted: , themselves within numerous sub-disciplines,
- Deleted: for the uninitiated
- Deleted: or

The 'Goldilocks problem' of finding the level of jargon just necessary for accuracy, but without alienating the broader readership could potentially be overcome through the effective use of supplements. Overall, in terms of language, the supplement allows for multiple 'languages', allowing it to be both easily understood by human researchers, as well as being machine-readable. In some instances, this might be reflected in a standardized hierarchy and standard terminology, in other cases it may necessitate otherwise awkwardly composed machine readable text juxtaposed to human readable text. The supplement can contain a section that provides a jargon-free schematic of the research, or it can include easy-to-understand presentations or graphics that an author might use in a scientific or lay presentation. While the basic information provided in these types of graphics is likely not suitable for the main text, they remain extremely valuable in terms of communicating the ideas to broader audiences.

Moved (insertion) [1]

For example, This merger of presentation material with publication material has obvious benefits: a standard conference talk where the paper might be presented contain important background information and even historical or scientific context often not included within the actual introductory sections of the published paper. The inclusion of this information is likely to be of substantial value to researchers from other fields. Further, providing additional components of the slide deck from a talk or a number of related presentations could effectively merge a dynamic presentation of the data with the heretofore more static published presentation of the

- Moved (insertion) [2]
- Deleted: t
- Deleted: Additionally, the supplement
- Deleted:
- Deleted:
- Deleted: , or even

Moved up [1]: The 'Goldilocks problem' of finding the level of jargon just necessary for accuracy, but without alienating the broader readership could potentially be overcome through the effective use of supplements.

Moved up [2]: the supplement can contain a section that provides a jargon-free schematic of the research. Additionally, the supplement can include easy to understand presentations or graphics that an author might use in a scientific, or even lay, presentation. While the basic information provided in these types of graphics is likely not suitable for the main text, they remain extremely valuable in terms of communicating the ideas to broader audiences.

data. As with data and other digital artifacts, a DOI should be assigned to an associated slide deck.

We believe it is essential that vocabularies, taxonomies, and metadata be standardized such that data can be easily read and manipulated across labs, fields and time. To this end, the supplement could also have a very precise glossary, translating language used in the paper into precise database identifiers and standardized names so that machine text miners can learn for each supplement how to easily parse through that supplement and relate it to a database entry.

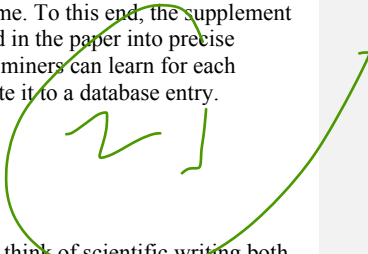
(perhaps here you can include the two sections IITL and IICP)

2.II.e Hierarchical Information Structures

To understand the overall structure of the supplement, one has to think of scientific writing both in terms of a hierarchy and in terms of parallel passes at increasingly greater levels of detail. Supplements can be both. They can provide a hierarchy in the sense that they divide the information into discrete chunks to allow readers to avoid reading through a tremendous amount of highly detailed albeit potentially irrelevant (to their present interests) text. Additionally, a hierarchy provides a roadmap: reading a scientific text can be seen as analogous to an information retrieval task, wherein a reader first peruses an introductory idea section and then jumps into a more detailed version of that section. The current structure of a standard scientific manuscript implements a version of this idea. A vague yet still informative title, a more detailed abstract, a somewhat expanding introduction, a detailed result section with even more detailed tables, and then moving back out, a conclusion that applies the details therein more broadly. The proposed supplement guidelines would expand on this age-old structure, building onto this preexisting hierarchy and providing even more detail.

This hierarchical structure would operate in a parallel fashion to the main text. Essentially, it would be a shadow text that directly tracks and corresponds to the main text, providing more detailed explanations for each heading. A reader looking for more detail on a particular part of the main text could easily find and then consult the analogous part of the supplement, which would be similarly situated within the hierarchical structure. Using a literary metaphor, the published paper can be viewed as the primary classical text. The supplement reflects the annotation, gloss and other editorial content on that text adding both integral and associated, tangentially relevant content and context. Here, however, the author and the editor are one and the same.

This hierarchical mirroring can be readily extended to the figures and tables, which can have more detailed contents in the supplement. This idea, of course, of both a hierarchy with increasing level of details and a parallel shadowing flow to it can be extended beyond that of a single paper to a whole collection of papers, as often the case in a large multi-group project where a coauthored high level paper describes the overall structure of the project, and a succession of more detailed papers often across multiple journals describe single, specific, drilled down ideas. With big consortia science project publishing multiple interconnected papers, a global hierarchy for all the papers can be developed, with that global hierarchy then corresponding to various supplements associated with individual papers published in conjunction



Deleted: , concurrently,

Deleted: translate into the details of its construction

Deleted: : i.e.,

Deleted: so that it essentially operates as

Deleted: part of the main text

Deleted: Effectively a shadow paper, this

with a primary roll-out, or even later subsequent papers. This system would also provide a clearer picture of the interconnectivity of the individual papers.

Further extending this literary metaphor. The author of the supplement can act as a compiler and editor of a collection of works, providing relevant information to, for example, draw perhaps unseen connections across the body of work. See, e.g. Σ .Fig.A.

Comment [BC-SSU12]: I think this metaphor works here, in the supplement.

This proposed hierarchy would include standardized headings for easy human and machine readability, with the structured headings directly corresponding to headings in the primary paper. Additionally the supplementary material should be designed to include ample indexable metadata relating various elements within the paper's hierarchy.

Σ .II.f Proposed hierarchy

Comment [BC-SSU13]: Again, I think this can be combined with the previous section.

In this proposed hierarchy the primary text sits atop the supplement, synthesizing the entirety of the supplemental information in broad strokes. Local links point to more detailed descriptions of methods and data located further within the supplemental materials.

The detailed description expanding upon the top level primary text should be logically divided such that each division addresses one coherent aspect of the analyses. The order of these divisions would map onto the order of appearance within the top-level primary text. Additionally, the divisions would also map onto the actual published paper, allowing researchers to easily move between the supplement and the original paper. As a bonus a clearer the hierarchy that can be easily mapped onto the original published paper will make adding, editing or modifying these links by internally and externally that much easier.

In a secondary hierarchical structure, each of these individual divisions may relate to its own huge amount of supplementary calculations and data sets. These calculations and datasets, would be further linked such that they relate back to each division within the top-level primary text. Moreover, to promote machine readability of the data sets, data should be provided in a standard tabular format, for example CSV format. Charts, graphs and other pictorial representations of the data should be decomposable, for example accompanied by machine readable files comprising the underlying the images.

Deleted:

Practically speaking, all data falling within the hierarchy should be localized to a single digital location. When necessary hyperlinks can be provided to outside sources, but all supplemental data should fall within the scope of the journal's supplement section.

Deleted:

In some cases, the sheer size of intermediate or non-essential data sets may require that some data might reside in an off-site website, provided that the authors guarantee viability to the links. Here, usage of standard widely accepted repositories, an institutionally supported and persistent website, a commercial cloud, or even a shared community repository might be best.

Σ .II.g Citation Standards

With an established hierarchy, different components of the paper and its supplement can be referenced intelligently, including, for example, distinct digital object identifiers (DOIs) for portions of the paper itself, as well as related identifiers, through the clever use of prefixes and suffixes for related portions within the supplement. The use of these DOIs need not be limited to just text, but can be expanded to include suffixes for related figures, tables, data sets and other related information. DOIs, or similar systems would also be useful given the nature of the supplement, allowing for the insertion or deletion of information without otherwise complicating the finding of other information. This use of DOIs is especially important in overwhelmingly large supplements that would be too time consuming to actually peruse through to find the desired section, text, figure or other source of information. Here, simply directing the reader to see the supplement, as is unfortunately, all too common, would effectively be a fool's errand without micro-referencing.

Unlike the published text, authors can take advantage of the nature of the supplementary section to provide for micro referencing of micro-authorship, utilizing ORCID IDs or other persistent unique identifiers for unique identification of authors and noting which specific authors from the original publication, as well as perhaps, authors not included in the original publication contributed to each individual portion of the paper. Not only would this provide a more realistic accreditation of authors than standard author listings, but would provide interested readers with direct access, perhaps through published email addresses, for each author for the particular area, text, figure of interest.

Figures would not only include captions and links to relevant parts of the text, but might also include additional information related to the relevant contact individuals for each figure, and access to the source code and data that generated the figure. Again, this would be particularly important with the growing trend to have tens if not hundreds of authors on biological papers.

Supplementary material will also include an expand bibliography. This bibliography can be designed to provide contextual information both with regard to the paper itself as well as the supplementary material. Additionally, the bibliography can be annotated to provide substantive information as to how each source relates to the presented information.

Σ.II.h Micro-referencing

Citation standards should be broadened to allow for pinpointed referencing between the primary text and the supplemental text such that readers of the primary text will be directed from the main text to the relevant section in the supplemental material and readers of the supplemental material will be directed back to the relevant portion of the main text, e.g., perhaps with micro DOIs or other reference systems. To some degree, this micro-referencing can be accomplished through an elegant hierarchical structure in the main text that would be shadowed in the supplemental text and/or vice versa. This should be further simplified through a standardized numbering system, allowing for sections, subsections, and even further divisions if necessary.

Further, this citation standard can include additional information relating to super-sections, tying together published papers across multiple journals and even disciplines. Optimally, publication

Comment [BC-SSU14]: You already mention a lot of this information in the previous sections. I wonder if you want to keep this as a separate section or just move some of these sentences to other sections where relevant.

NOT

Ref of Main

databases would provide identifiers to not only the main published paper, but would at minimum list the other identifiers associated with the paper.

Σ.III Conclusions

Supplements have become a necessary part of conducting regular scientific business, both from the original researcher's standpoint of presenting their research in its entirety, and also from the follow-up researcher to effectively use the original research.

Although we provide a comprehensive wish list for a supplement to deal with the many issues inherent in current supplementary materials, one outstanding concern relates to editing and peer reviewing of these unwieldy behemoths. To the extent possible, review of the supplements will be increasingly necessary as they become an integral part of the scientific process, however, given their large size, one approach may be to review random samples of the supplement or utilize a trusted third party such as ResearchCompendia.org to verify computational results.

The popularity of consortia science and the deluge of data that it brings has created an ever-growing need for more structured supplemental data. This is necessary not only for providing FAIR access to important datasets, but particularly for the increasing use of machine learning tools to mine scientific literature. The proposals herein represent only some of the changes necessary to maintain the usefulness of supplemental data.

Deleted: for the

Deleted: on

Deleted: this

-
- ¹ Hopewell, Sally, Mike Clarke, and Sue Mallett. "Grey literature and systematic reviews." *Publication bias in meta-analysis: Prevention, assessment and adjustments* (2005): 48-72.
- ² Borowski, Christine. "Enough is enough." *The Journal of experimental medicine* 208.7 (2011): 1337-1337.
- ³ Lior Pachter, Stories from the Supplement, Presentation at Genome Informatics, CSHL, November 1, 2013 available online at <https://liorpachter.wordpress.com/2013/11/02/stories-from-the-supplement/>
- ⁴ Flanagan, Annette, et al. "Recommended practices for online supplemental journal article materials." 2014-05-07]. http://www.niso.org/apps/group_public/download.php/10055/RP-15-2013_Supplemental_Materials.pdf.
- ⁵ Kenyon, Jeremy, and Nancy R. Sprague. "Trends in the Use of Supplementary Materials in Environmental Science Journals." *Issues in Science and Technology Librarianship* (2014).
- ⁶ Stodden, V., "[Resolving Irreproducibility in Empirical and Computational Research](#)" IMS Bulletin, Nov 2013. <http://bulletin.imstat.org/2013/11/resolving-irreproducibility-in-empirical-and-computational-research/>
- ⁷ Taichman, Darren B., et al. "Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors." *The Lancet* (2016).
- ⁸ Brooks Hanson, Andrew Sugden, Bruce Alberts, "Making Data Maximally Available," *Science*, Vol. 331 no. 6018 p. 64, 2011. doi:10.1126/science.1203354 <http://www.sciencemag.org/content/331/6018/649.full>
- ⁹ Ratner, Mark. "IBM's Watson Group signs up genomics partners." *Nature biotechnology* 33.1 (2015): 10-11.
- ¹¹ Maunsell, John. "Announcement regarding supplemental material." *The Journal of Neuroscience* 30.32 (2010): 10599-10600.
- ¹² Marcus, E. 2009. Taming supplemental material. *Cell* 139(1):11-11.
- ¹³ Pop, Mihai, and Steven L. Salzberg. "Use and mis-use of supplementary material in science publications." *BMC bioinformatics* 16.1 (2015): 237.
- ¹⁴ Seeber, Frank. "Citations in supplementary information are invisible." *Nature* 451.7181 (2008): 887-887.
- ¹⁵ Weiss, Manfred S., et al. "Citations in supplementary material." *Journal of Applied Crystallography* 43.6 (2010): 1285-1286.
- ¹⁶ Rafferty, Anthony R., Bob Wong, and David G. Chapple. "An increasing citation black hole in ecology and evolution." *Ecology and evolution* 5.1 (2015): 196-199.
- ¹⁷ Kenyon, Jeremy, Nancy Sprague, and Edward Flathers. "The Journal Article as a Means to Share Data: a Content Analysis of Supplementary Materials from Two Disciplines." *Journal of Librarianship and Scholarly Communication* 4 (2016).
- ¹⁹ Carpenter, Todd. "Standards Column-Taming the World of Data: Pressures to Improve Data Management in Scholarly Communications." *Against the Grain* 22.6 (2014): 44.
- ²⁰ Pop, Mihai, and Steven L. Salzberg. "Use and mis-use of supplementary material in science publications." *BMC bioinformatics* 16.1 (2015): 237.
- ²¹ Newton-Cheh, Christopher, et al. "Genome-wide association study identifies eight loci associated with blood pressure." *Nature genetics* 41.6 (2009): 666-676.
- ²² Hanson, B., Sugden, A., and Alberts, B. 2011. Making data maximally available. *Science* 331:649.
- ²³ Schwarzman, Alexander B. "NISO/NFAIS Supplemental Journal Article Materials Working Group: A progress report." (2010).

-
- ²⁴ Carpenter, Todd. "Outside the core: working towards an industry recommended practice for supplemental journal materials." *Serials* 23.2 (2010).
- ²⁵ Borowski, Christine. "Enough is enough." *The Journal of experimental medicine* 208.7 (2011): 1337-1337.
- ²⁶ Maunsell, John. "Announcement regarding supplemental material." *The Journal of Neuroscience* 30.32 (2010): 10599-10600.
- ²⁷ Nature America. (2012). Editorial: Moderating supplementary data. *Nature Neuroscience*, 15(3).
- ²⁸ <https://www.force11.org/node/6062>
- ²⁹ Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3 (2016).
- ³⁰ Rodríguez Iglesias, Alejandro, et al. "Publishing FAIR Data: an exemplar methodology utilizing PH-base." *Frontiers in Plant Science* 7 (2016): 641.
- ³¹ da Silva Santos, Luiz Olavo Bonino, et al. "FAIR Data Points Supporting Big Data Interoperability."
- ³² Wilkinson, Mark D., et al. "Interoperability and FAIRness through a novel combination of Web technologies." *PeerJ Preprints* 4 (2016): e2522v1.
- ³³ Rodríguez Iglesias, Alejandro, et al. "Publishing FAIR Data: an exemplar methodology utilizing PH-base." *Frontiers in Plant Science* 7 (2016): 641.
- ³⁴ McQuilton, Peter, et al. "BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences." *Database* 2016 (2016): baw075.
- ³⁵ Nichols, Thomas E., et al. "Best Practices in Data Analysis and Sharing in Neuroimaging using MRI." *bioRxiv* (2016): 054262.
- ³⁶ Haak, Laure, David Baker, and Thorsten Höllrigl. "CASRAI and ORCID: Putting the pieces together to collaboratively support the research community." *Procedia Computer Science* 33 (2014): 284-288.
- ³⁷ Stodden, Victoria. "Enabling reproducible research: Licensing for scientific innovation." *Int'l J. Comm. L. & Pol'y* 13 (2009): 1.
- ³⁸ Donoho, David L., et al. "Reproducible research in computational harmonic analysis." *Computing in Science & Engineering* 11.1 (2009): 8-18.
- ³⁹ da Veiga Leprevost, Felipe, et al. "On best practices in the development of bioinformatics software." *Frontiers in genetics* 5 (2014).
- ⁴⁰ Stodden, V. & Miguez, S., (2014). Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. *Journal of Open Research Software*. 2(1), p.e21. DOI: <http://doi.org/10.5334/jors.ay>
- ⁴¹ Bechhofer, Sean, et al. "Why linked data is not enough for scientists." *Future Generation Computer Systems* 29.2 (2013): 599-611.
- ⁴² Donoho, D., and Stodden, V., "Reproducible Research in the Mathematical Sciences," in *Princeton Companion to Mathematics*, Edited by Nicholas J. Higham Mark R. Dennis, Paul Glendinning, Paul A. Martin, Fadil Santosa & Jared Tanner, Princeton University Press 2016.
- ⁴³ Garijo, Daniel, and Yolanda Gil. "A new approach for publishing workflows: abstractions, standards, and linked data." *Proceedings of the 6th workshop on Workflows in support of large-scale science*. ACM, 2011.
- ⁴⁴ <https://galaxyproject.org/>
- ⁴⁵ Deelman, Ewa, et al. "Workflows and e-Science: An overview of workflow system features and capabilities." *Future Generation Computer Systems* 25.5 (2009): 528-540.

Their relegation to the metaphorical back of the journal notwithstanding, supplements ought to be an essential component of the modern scientific publication and data archiving efforts. Many, if not most articles in popular journals, and especially most genomics journals, include potentially useful if not necessary supplemental materials.

Supplemental material may be seen as necessary both in terms of preserving and allowing subsequent access to structured and raw data, as well as cataloguing promising (and failed) ideas and directions for further research. Data archiving is also essential in the important goal of reproducing reported findings¹ and also, in developing follow-on research efforts and tangential, or even unrelated research efforts.

To the degree that journals can enforce standards in their supplements, supplements represent untapped potential as usable data archives, particularly for information that might forever be lost from the scientific record. For example, whereas negative results are typically not publishable, they can nevertheless become part of the scientific record through their inclusion in supplements.

size and format (including documentation);
scope;
persistence;
and accessibility of supplemental material.

This supplement provides an expanded version of the propose structural suggestions from the primary text as well as additional information relating to suggested standardizations that will be valuable in optimizing the usefulness of supplemental materials.

As described above, and in the primary paper, scientific papers tend to become convoluted in their sometimes ineffective efforts toward conciseness. Supplements have the potential to provide substantial clarity of the published text, not only by providing often much needed annotation, but by also providing important and additional information.

¹ Alberts, Bruce, et al. "Self-correction in science at work." *Science* 348.6242 (2015): 1420-1422.