Academic citation practices need to be modernized
References should lead to full texts wherever possible

Researchers and academics spend a lot of time documenting the sources of the ideas, methods and evidence they have drawn on in their own writings. But I want to try and convince you that our existing citation and referencing practices are now woefully out of date and no longer fit for purpose in the modern world. The whole scholarly purpose of citing sources has changed around us, but our conventions have not recognized the change nor adapted yet. I first set out what’s wrong with what we do now, and then sketch a radical agenda for starting afresh.

(Just a word of warning—this is a long post. There’s a much shorter and punchier version here that some readers may prefer).
The current core details

There has been some past effort to standardize a certain core of things that all academic references will include. For books this includes

- author last name; plus at least one first name initial (or full first name);
- the same details for all other authors
- full title (including sub-titles, although they are often left off)
- place of publication; the name of the publisher; and the year date of publication

For reports, working papers, conference papers and ‘grey literature’ (not published by formal publishers) the core details are much as for books, but with the ‘publisher’ being whatever organization originated a report or working paper, or the name of a conference where a paper was presented. The name of any working paper series and any identification numbering should be recorded and the full date of publication or presentation (i.e. day, month and year of publication). These details are more numerous or demanding because grey literature is chronically harder to track down. Indeed a still incredibly common trait is for even serious and respectable organizations to publish reports without including even a year date on them, let alone other publication details.
For *journal articles* the standard core now includes

- author(s) last name, and initial(s) or first name, as for books.
- full title (including sub-titles)
- the journal name. This should be utterly straightforward. But instead journals often abbreviate their own names and the names of other journals in various completely unpredictable and unnecessary ways (e.g. writing ‘Jnl’ instead of ‘journal’ or using acronyms). Different sets of academic editors, professional associations and commercial publishers seem to take a perverse pride in recording the source details of publications in different ways. These are ultra-legacy elements, decipherable only by the cognoscenti and stemming from the days of letter press printing half a century ago, when saving characters also saved money. They have no current rationale at all.
- the year of publication, sometimes supplemented by a month date or a season name (e.g. Summer)
- the volume of publication. Actually this should always be just an esoteric re-coding of the year, with the number depending on when volume 1 first happened. However, some journals also have more than volume per calendar year, to make things difficult. An extra spice of incomprehensibility is added by some journals who record their volume numbers in Roman numerals.
(like LXIII)—a master stroke of one-upmanship.

- the page numbers range—i.e. the beginning page number to end page number.

Why these core details don’t work any more

Many elements of the list above are broken or have become pointless in the digital age.

1. *Author names* in many British and European sources (book publishers and journals alike) often still include just a single initial. Academics and professionals from these smaller nations have been remarkably slow to appreciate the globalization of knowledge, and hence the need for much more distinctive author names. They (and their journals) are still reluctant to go beyond a single initial (J.) to distinguish John Smith from Joan Smith. By contrast, American publishers and journals (more accustomed to a country with 300 million people in it) tend to give the first name in full, and sometimes a second initial as well. Clearly, in the era of global search engines the US practice needs to become universal, but there is still a long way to go.

2. In all the STEM sciences *the number of author names for journal articles* has tended to increase sharply in the last ten years. In some disciplines the proliferation of author names is beginning to cause some reductions to be made in which author names are included in
references, although the process is proceeding in an erratic and non-standardized way. Where once journals or book publishers might have tried to list all authors, increasingly some sources will only list the first ten or even just the first five authors in their reference lists. People who want a full author list hence need to actually go to the article itself, where the first page will still show everyone involved. With many physics papers listing 50+ authors, and some several hundred, this change has become inevitable.

3. *Volume and issue numbers* no longer make sense for journals that have moved to continuous publication. Even for journals that retain volume and issue numbers almost all articles are now being published online before (often months or year before) they get a print volume and issue number. The gap between online publication and print issue publication can be substantial. In the social sciences and humanities other authors may be very reluctant to cite such online pieces, because online papers are often harder to find on dated publisher websites, and they know that any interim reference they make will become obsolescent. Most libraries, electronic depositories and publishers insist on rewriting the citation of an early online article, so as to use instead the print on paper volume and issue version of the reference, even when this effectively falsifies the timing of the work. For instance, one of my recent papers spent nearly two years in this limbo of ‘early online’ status, and at the
end of it had changed from an autumn 2011 piece to a summer 2013 one. In other words the current core convention deliberately introduces inaccuracy and falsified details into academic referencing, the opposite of their supposed purpose.

4. *Page numbers* are also irrelevant for many sources now. In ‘early online’ articles the page numbers all start at 1, *until* the article gets incorporated into a specific volume and issue number, where suddenly the page numbers are changed completely—thereby invalidating any previous page-specific citations. Similarly many ebooks now often do not have page numbers, since they re-size automatically to fit the screen size of the device that readers are using, and to adjust to readers’ preferences for font sizes. Hence pagination in the digital age makes no sense at all.

5. *Place of publication for books* is also a mostly pointless piece of information. Many big international publishers issue the same books in two or more places at the same time—for instance, in the USA and in the UK or Europe—yet these identical books will be referenced as if they were different. For many smaller publishers it is sometimes a bit of a job to find out where the place of publication actually is—even by searching their websites. This is often the last piece of information that I have to include in my reference lists, precisely because it actually matters very little in a digital era. Every publisher of any importance worldwide is now
on the internet and the web and almost all will be accessed via Google Books or Amazon—home even to self-publishers nowadays.

'Legacy' referencing marginalizes open access texts

What is the essential purpose of academic referencing? What is its ‘be all and end all’ rationale, such that we devote so many hours to it? A completely out of date answer dominates current practice—namely that referencing and citing is about showing (acknowledging) your sources, in a way that can be followed up by another researcher. Your referencing should direct them to the same precise sources and pages that you yourself used in constructing an argument or a case. In this sense referencing is about replicability (ascertaining that a cited source actually exists and says what you says it says), as well as about correctly assigning credit, or (far less commonly) criticizing inadequate work.

But in the digital era this is too limited an ambition. Referencing should instead be about directly connecting readers to the full text of your sources, ideally in a one-stop way. Using URL referencing of the kind I employ in this blog, or other innovative methods, readers should be able to go directly (in a single click and in real time) to the specific part of the full text of source that is being cited. In other words, modern referencing is not about pointing to some source details for books that cost a small fortune
and are buried away in some library where the reader is not present; still less about pointing to source details for an article in a pay-wall journal to which readers do not have access. That is legacy referencing, designed solely to serve the interests of commercial publishers, and 90% irrelevant now to the scholarly enterprise. If that is the best that we can do in connecting readers to our source texts, then it will have to do. But let’s face it, it’s not much use in today’s world.

With open access spreading now we can all do better, far better, if we follow one dominant principle. Referencing should connect readers as far as possible to open access sources, and scholars should in all cases and in every possible way treat the open access versions of texts as the primary source. Versions of the text that depend upon paid for access (buying the book, or subscribing to the journal) should be relegated to the status of secondary sources, supplementary information for status conscious academics (or their promotion committees), but not forming part of the core information about a text. This may seem revolutionary but it is actually just a reflection and slight extension of the new rules that the British government’s research funding body has already introduced for the next ‘research excellence framework’ (REF) exercise, expected in 2020. For any academic’s or researcher’s journal articles to be considered as part of a university case for REF funding support they will either need to be available in open access form in the journal
(free to any reader), or the university must show an immediate pre-publication version of the paper on their e-depository.

So the primary version of a journal article, the version that we reference in our own work and provide links to, should be in one of four forms:

1. In a wholly open-access journal. This is probably the best option because a well-known journal is easy to find, and most readers in the field will already know that in this source they can click through to any paper, maximizing their incentives to do so.

2. An open-access article in a generally pay-walled journal. Readers will still get the full text if they click through, because the authors or university have paid to secure that. But current estimates suggest that less than 5% of articles in paywall journals are open access, so this status needs to be clearly communicated e.g. by putting [Open access] at the end of the reference. Otherwise readers may see this as just another legacy source.

3. The immediate pre-publication version available on the university e-depository. Essentially this is the author’s final manuscript version, so that the text and Figures etc are completely identical to those in the formally published version—but, of course, the pagination is not the same.

4. The immediate pre-publication version available on
another widely accessible and well-used access open access site, such as the brilliant Research Gate, or perhaps academia.edu. (If you don’t know about these sites already, please read my post on not being an academic hermit).

From now on therefore, any version of a journal article behind a paywall should only be cited as the primary source if none of the four options above is available and nothing better than source details alone can be produced. However, if an open access version of a text is available, this must always be treated as the primary text. Here the commercial version of the text becomes the secondary version and it should always be cited second and in a manner that makes this completely clear. For instance, after the primary reference to the full text, you could write: ‘Also available as: ….’

Updating legacy practices for digital
Perhaps you are sceptical about (or do not agree with) the argument for prioritizing open access versions above. Whatever your stance here, hopefully in this section I can convince you of the need to make some other fundamental changes in our concept of the core details for journal articles and books. These absolutely essential elements, the details that should be universally given, need to be expanded in the digital era so as to cover:

- The shortened URL for the university e-depository version
of the text. All commercial publishers and many journals still hate including URLs in reference lists because in the past academics and researchers would just copy long URL addresses off the open web, or ordinary Google, with many defects. Often the links were strictly temporary (and so often became broken links). The URLs cited also included all kinds of ‘rubbish text’ elements, and so they were overly long and looked ugly, out of place and disruptive in reference lists. These problems are behind us now. No one needs to include long URLs anymore — simply go to bitly.com (or another alternative site) and type in a long URL, to get a compact and neat looking version that fits easily into any reference list. All university e-depositories should now issue permanent URLs for any item that they store, links which are guaranteed not to break or change in future. The best depositiories now have their own very short but well-branded permanent URLs — for instance the LSE’s Research Online service gives URLs that look like this: http://eprints.lse.ac.uk/56492/
In the remote contingency that anything goes wrong with the reproduction of such a link (e.g. because of misprints), the URL includes LSE’s web address and the staff will also be able to help readers to find the right source, and can issue corrections or arrange redirects. I hope that all depositories are geared up now to work in this way?

- **The DOI permanent URL for the source** (which covers both journal articles and also books). DOI here stands
for ‘digital object identifier’, which is a unique code number issued by commercial publishers for each individual journal article or book that they publish. This identification number will never change. And if you add the prefix http://dx.doi.org/ to the front of it then you get a permanent URL. So take my paper on ‘Analysing party competition in plurality rule elections’ which has: doi: 10.1177/1354068811411026. Combining the prefix and the number gives an invariant URL http://dx.doi.org/10.1177/1354068811411026 that takes you reliably to the commercial publisher’s site for the article. In this case it is a paywall site, so it is not much good beyond a legacy source link unless a reader has access to that journal. But if the article is in an open access journal or is an open access piece in a paywall journal, then the DOI gives readers a second permanent URL. However, it should be clear that the publishers’ DOI is a lot less attractive and is more easily messed up or mis-recorded (because it has so many numbers) than a shortened permanent URL from a good university online depository. This is an inevitable product of a DOI system that now includes some 85 million separate items.

We need to junk page references in favor of short source quotes that readers can search and find

Academic references should wherever possible be precise, and so a hallmark of past good practice has been that
citations tell readers exactly where to look in otherwise long and baffling texts for the provenance of what is being said. But if digital and open access texts are now our primary sources, the ones that we know the vast majority of readers will use, then pagination becomes irrelevant. The page numbers on a e-depository text, on the ‘early online’ version of an article, and in the ‘final’ article included in a journal issue and volume will all be different. This way only confusion lies.

And if we want to URL-link readers directly to text passages, then incorporating pagination is simply not viable for ordinary academics and authors. We can only URL link to a whole source, taking readers to the top of the source but not to a specific point inside it. We cannot do more specific URLs at present.

But if a URL link is for a quotation the problem disappears. Readers follow the link to the top of the sources, copy a few words of the quoted passage into the Control+F search box, and go directly to the passage cited. (If the source is a book with preview, then readers could also try their luck with the facility to search inside a book on Google Books). So the solution for modern scholars must be to expand our use of very short ‘source quotes’ in what we write. Previously I might write in this style: “Bastow et al 2014, p. 21) argue that social scientists study mainly systems where humans are dominant”. Now I need to write in a a slightly different way so as to get rid of the
need for obtrusive reference, thus: “Simon Bastow and colleagues argue that the social sciences focus most on ‘the study of human-dominated systems’”. Source quotes replacing page references do not have to be memorable, nor especially salient bits of text, nor very long—they can be very short so long as they are unique. The six words that form this particular link are enough to identify without ambiguity a single sentence in a book of 300+ pages. [If in doubt, check that your source quote words are unique using Control+F].

Other citation problems
Everything to do with referencing and citations is made a hundred times worse by the completely pointless proliferation of different referencing and citation styles and systems, one that commercial publishers have facilitated in a desperate effort to prove their responsiveness to academic demands and show their ‘value added’. There are now some 5,000+ different styles, and every month some bozo or other adds another one. This co-ordination fiasco is exhaustively documented in some of the most esoteric and pointless books ever produced by humankind—of which Turabian’s Chicago style guide is perhaps the most over-blown example. Literally thousands of editing, proof-reading and librarian/information science jobs worldwide are now dedicated to coping with the pointless complexity of so many different academic referencing protocols.
Open access pressures to cut publishing prices and costs should help reduce the mess a bit. Leading profit-makers Wiley and Elsevier still charge $2,700 to make paywall journal articles open access, whereas the sustainable lrbl is less than 25% of this, around $600 for a top journal—should should force even legacy publishers to create a ‘low cost airline’ version of their current, pointlessly high cost offerings. As more people use brilliant modern text editors, like that on Medium.com which offer just a stripped down set of editing options, they will also see that the publishers’ over-refinements add no value.

(Digression: When you write on Medium your text is auto-saved every few seconds, and you can only use bold, italics, underline or URL links to vary it. Every Medium article uses the same font size throughout. There are two levels of heading only, and no sub- or super-scripts (that I’ve found yet). You can’t do formulae, tables or charts (although you can load in pictures of formulae, tables or charts). When you log on to Medium for the first time it prompts you: ‘Bang out some text!’ It’s just the best way of writing, and makes going back to MS Word, or coping with idiotic journal style sheets, seem as antiquated as using a typewriter).

Quick recap — what authors should do now
Let me sum up the many different strands of argument above by summarizing what things a ‘born digital’, pro-open access scholar should be putting in their references
now, wherever feasible. And definitely these are the details that you must be recording for the future in EndNote, Mendeley or any other reference-management software that you are using.

**For books:**

- Author first name, second initial and last name. Repeat for other authors up to a limit—e.g. first five or ten authors only, and then use *et al*.

- Full title and sub-title, all in words (don’t even use ‘&’ for instance)

- Year of publication

- Shortened permanent URL to the full text in open access online-depository—preferably the author’s home university e-depository version, or if not there some other open access version, e.g. at Research Gate.

- The DOI permanent URL to the commercial text, that is, [http://dx.doi.org/](http://dx.doi.org/) followed immediately by the DOI number. If an open access version of the source is available make clear that the commercial version is a secondary source. Put: ‘Also available as …’

- The name of the commercial publisher.

**For journal articles:**

- Author first name, second initial and last name. Repeat
for other authors up to a limit—e.g. first five or ten authors only.

- Full title and sub-title, all in words, exactly as printed in the journal

- The date of publication—give Year alone if that is all that is available, but for continuously published and early online articles give date, month and year, for instance: 26 April 2014.

- Shortened permanent URL to the online-depository open access full text (again preferably the author’s home university e-depository version, or other open access version, like academia.edu)

- The DOI permanent URL to the commercial text, that is, http://dx.doi.org/ followed immediately by the DOI number. If an OA version is available make clear that the commercial version is a secondary source. Put: ‘Also available as …’

- The full name of the journal, without any abbreviations.

- The start page and end page numbers (in full). Write in a transparent way, for instance: pp. 150-179.

- Other legacy elements can be added to taste, such as the largely pointless volume and issue numbers if they are still in use. Write in a transparent way, without codes, brackets etc. For instance: vol. 67, no. 4.
For conference papers, working papers, reports, media sources and grey literature generally

- Author first name, second initial and last name. Repeat for other authors up to a limit—e.g. first five or ten authors only, and then *et al.*
- Full title and sub-title, all in words
- The date of publication—give date, month and year in full wherever they are available, for instance: 26 October 2012. Give the year alone if that is all you can document.
- Shortened *permanent* URL to any online-depository or open access full text (preferably the author’s home university e-depository version, or another open access version).
- A shortened, non-permanent URL linking to the publishing organization’s website or other location where the item was accessed. Few business firms or government organizations have yet wised up to the need to guarantee permanent URLs. But some heavy publishers like the World Bank or the UK’s National Audit Office now do so. Wherever you cite non-permanent URLs, you also need to add: ‘Accessed on [give full date]’.
- The place of publication and the name of the organization publishing the item.
- If the item is in a series of working papers or reports,
Achieving modernization

Nobody can change citation systems on their own. We are just at the start of what will inevitably be a long process of driving out legacy citation systems from scholarship and science and replacing them bit by bit with a modernized approach that treats digital and open access sources as always the primary sources, and commercially published versions behind paywalls as radically inferior secondary sources. No doubt that there are also things missing from my lists above that should be there, but I am too dumb to spot, or that matter to different disciplines in ways I can’t see yet. So this is just a starting proposal, something to get the process of discussion started.

And I know from discussion with publishers and librarians that there are a range of other possible ideas in the offing, some of which strike me as a bit bonkers. E.g some people want to insist on recording ORCID numbers for authors in references—but these are also hard to remember and easy to get wrong. All we need is for authors to choose a distinctive author name and stick with it (as suggested above), and for journals and publishers to record these names properly. Undoubtedly converting legacy systems, plus fighting off this and other ‘distractor’ ideas, will take time and have many other aspects that need to be considered.
But if you are an academic, researcher or PhDer there is a very real payoff to immediately beginning to record all of the elements that I have set out above. For a start your own access to full text sources will radically improve by starting down this route. And you will be future-proofing your accumulated references against having to meet novel demands for these new digital and open access details that otherwise you may neglect. Also librarians and information scientists and PhD supervisors across the world need to radically update the advice that they give PhD students and early career researchers on how to do referencing. If your advice does not currently cover the suggestions made here, then you need to at least alert those you advise to this menu of ideas.

Finally, let me plead directly with any readers who exercise power in the scholarly publication process—because you edit a journal or a book series, you work for a publisher or undertake editing for authors, or you are a powerful person in your professional academic association. Look hard at the rules, conventions and styles that you are currently applying. Can you honestly say that your legacy requirements meet contemporary needs and have adapted to the digital dissemination of scholarship? Are your citation rules designed in a ‘born digital’ way? Do they foster open access and replicability by getting the full texts of source materials in front of readers in the most direct and simple way possible? If not, why not think about changing, discuss the issues with colleague, circulate
this post to the rest of your team or firm, and try to build a critical mass for change? We need to accomplish a very difficult, collective modernization, but your contribution can make a difference.

I’d welcome comments or alternative suggestions via email to: p.dunleavy@lse.ac.uk. Alternatively you can comment on the (shorter) LSE Impacts blog version of this post.

For new update materials please see my Twitter account @Write4Research and the many commentators on the LSE’s Impact blog.