Cataloguing and The Greatest Good

Arwen Caddy, RB Library and Information Supervisor

Introduction

One of my favourite lines from a Pixar film comes from a character you never see on screen. When a villainous robot begins attacking the city, the superhero Frozone is admonished by his wife, Honey, for paying attention to it, and not her. "*We are talking about the greater good!*" he protests. "*Greater good?*" comes the retort, "*I am your wife – I'm the greatest good you're ever going to get!*" [1]

I'll be honest, I like the line more for the timing and sassiness of delivery than the sentiment, but it raises the question of whether the greater good is always the most obvious good. When I heard the call for papers on Ethics and Cataloguing, my thoughts circled around the concept of the greater good. Is cataloguing always naturally perfect, or might there be conflicting ideas lurking within the principles and best practice?
This article, while not exploring an obvious ethical dilemma, is about matters of intellectual and practical ethics: "*Can we balance the competing goals of making the catalogue perfect, user-friendly, and fit-for-purpose?*"

Author's note: This piece is purposefully written to raise more questions than it offers answers; if you read it and it engenders contemplation, then I've probably got it right.

What follows does not aim to be comprehensive or conclusive, but to be illustrative of the topic. Here, then, are four of my favourites:

1. Efficiency or completeness?

Let's start with an easy one. I'm sure that all of you will flinch along with me if I ask you to imagine a manager with no sympathy for cataloguing who tells you to just '*get it on the system and then tidy it up later*' – especially if the undefined 'later' gets pushed further and further down your list of priorities – but take a deep breath…
What if you are launching a new catalogue and there is no way that you can create records in time. Do you catalogue five thousand perfectly and disappoint the first users with your selection, or do you try to create retrievable records for all ten thousand and hope they are functional?
This question has come up twice in my current role, and we have tried both options. Our LMS was launched in 2014 and we decided to create complete records as we went. Four years later approximately one third of our physical collection is still not included. The second time we decided to tidy as much metadata as possible, but ultimately to make all records available even if they were incomplete. Objectively it was the more successful project, but, after six months of overtime, launching an imperfect dataset made me feel like a weepy failure.
As you can probably tell, my instinct is certainly for completeness, but I can appreciate that this is sometimes unachievable. Is this an ethical issue? Perhaps not, but the impact it has on retrieval can make it an important one. And if you are going to pick five thousand, how do you decide what gets priority? Perhaps a populist answer does stray into social ethics.

2. Quantity or Quality? [apologies to all contributors to WorldCat]

I use WorldCat frequently, and it has much to recommend it, but oh! how I wish it had one version of the truth. Let's take a fairly standard publication: Journal of the Medical Association of Thailand. WorldCat has nineteen entries for this journal. They can't all be right.

---

1. The Incredibles. [film]. USA: Pixar Animation Studios/Walt Disney Pictures; 2004

It's unfair to snub. All those entries may be valid. Some are for the electronic edition, some paper – some records focus on pure metadata like ISSNs and dates, while others give more detailed notes on language or historical information. I often tell my assistants that it's better to give too much information than too little, and that every nugget you enter might one day be useful (or just fascinating).

Still, all I want in this case is the correct version of the title to transpose into my catalogue, and there are seven subtly different ones to choose from. We often think we're being clever by making individual entries for different formats, but how many users will become overwhelmed and give up? And don't assume it's obvious; I recently ordered a copy of Howards End from my local public library and inadvertently selected a 1960's critical appraisal, rather than the novel itself. If a cataloguer can do it, an 'ordinary' user certainly can.

3. Decisiveness or Perpetual Research?

Let's be honest, true cataloguers love a challenge. We like the warm glow that comes when we track down an ISBN from the deep depths of the internet and corroborate it with a brilliant flash of lateral thinking. But how long should we spend on these expeditions? Half an hour? Half a day? A week? Three months? Do we try and corroborate two pieces of evidence or more? Best of three? Best of five? If a decision is particularly difficult, should we document that too? When does pursuit of the truth become a waste of time? Does the result weigh in the balance? Is a day wasted if no answer is found, but a month of work worthwhile if the work is successful? This may seem instinctive, but there are issues of trust here. If you decide to go with your best guess, does the bastion of truth that is the library catalogue get slowly watered-down and slip into the sea?

4. Imports or In-house?

"Many a time and oft" I have heard advocates speak about the benefits of importing records from the British Library or another great catalogue. [2] It's time-saving, efficient, embracing of technology, collaborative, consistent, and do you really think you can do better yourself?
Well, yes, actually. I'm sorry, but I think I can catalogue my own library better than someone who doesn't work here. Controversial? Let me explain (and anyone who has heard me speak will know I always use Durex in my examples)…
I work at the scientific library for the company that makes products including Nurofen, Strepsils, E45, and Durex. Unsurprisingly we catalogue a lot of journal articles on topics around sexual health, pregnancy and contraception. However, when my team catalogues these, we don't use the term 'sexual health'. We use 'sexual wellbeing', because that is the language used by the company. My record will be more retrievable than yours, because I tailor it to my users.

So, when we start to delve a little deeper, we can quickly see conflicts between ideals that won't always be resolved in the same way. For me, this is clearest when I think about the conflicting 'greater goods' at the heart of my current role. They revolve around the paradox between consistency and perfection, and this is my fifth and final illustration.

5. Consistency and Perfection?

Back in 2014 I was privileged to present a session at the CIG Conference. Some people might remember it was about the software we used for cataloguing our main repository of scientific articles, and the fact that a record was locked down on creation against any alteration. The bottom line was if we made a mistake in the record, we could not correct it.

---

2. The Merchant of Venice. William Shakespeare

Here's a simple question: if you make a mistake on record A (that you can't fix), do you correct it on record B, or do you copy the mistake? Sounds easy, doesn't it? Who would choose to copy the mistake and end up with two bad records?

Well, you might if you were trying to group things together, to make them retrievable under the same search. Perhaps it's difficult to visualise with just two records. Let me give you an example.

There is a UN/World Health Organisation committee called The Joint Expert Committee on Food Additives; its acronym is JECFA. For years, when speaking aloud, my team called it JEFCA, because "jeff-ka" is easier to pronounce than "jeck-fa". At some point, we realised we were saying it wrong and started making a conscious effort to correct ourselves. Then we realised that a casual oral mispronunciation had spilled into our written catalogue. Between 2012 and 2015 we had added 7 records to our repository with JEFCA, not JECFA, as both publisher and part of the title. If we added a new record correctly as JECFA, no-one would be able to find either it or the others. For consistency, it makes sense to go with the greater number, but we're usually only adding one at a time, so this becomes self-perpetuating.

Another example is the one I shared at CIG 2014. The author whose initials are R. C. Rowe, and so half the records are under Rowe, R and the other half under Rowe, C. These at least are retrievable if you only search for the surname, but when was the last time you told a user to search with the bare minimum of data for the best results!

These examples, where the mistake is in the most fundamental metadata, are painful and polarising. Ultimately a decision must be made and, in both cases, we opted to correct the mistakes in all future records as soon as they were spotted. However, there are also times where it's not so much about a mistake as an improvement. This usually happens with keywords.

We often get little flurries in our adding to the catalogue where we'll suddenly get a dozen papers on a subject that one of our researchers is working on. Imagine you are cataloguing twelve papers and on about the seventh you spot a brilliant term that should clearly be included as a keyword. It's not on the first six records because you've only just come across it, and those six are already locked. Do you 'improve' the second half of the records? Now what if that term is a synonym, and you wish you could not add to but replace the term used in the first six. Or, horror upon horror, what if the term you have been happily using previously turns out to be misspelt, misinterpreted, or incongruous?

I'm not a medic, and neither is anyone in my team. Our backgrounds are in librarianship, records management, or archives, not medicine or chemistry. The consequence of this is that we often catalogue papers we don't fully understand (and occasionally we catalogue the type of paper where we know what all the words in the title mean but have absolutely no idea what they are doing together in that order!) We use many techniques to ensure our keywording is of the finest quality, but sometimes we simply don't have enough knowledge to understand the subtle differences in distinguishing between adverse events and side effects, or knowing what drug class a particular raw ingredient is associated with. So, do we make the 'improvement' going forwards, or do we stick with the term we thought of first?

Conclusion

In this article, I've examined the related issues of:

1. Efficiency or Completeness? Is the cataloguing priority to list every title or to complete as many entries as possible in the time available?
2. Quantity or Quality? Are consistent entries better than varied entries for varied purposes when relating to one work?
3. Decisiveness or Perpetual Research? What is the impact of decisions due to time pressure on the finished resource?
4. Imports or In-house? Should your catalogue reflect one recognised as standard or be tailored to your user group?

Consistency or Perfection? Should practice be constantly improving or is it more important to have clear direction for categorisation and retrieval?

Perfection may be idealistic, but I think most cataloguers would strive for a value of 'correctness' as the greatest good. If we don't, where is the line drawn before we stray from interpretation to misinformation? On the other hand, context gives perspective to the need to be correct all the time. What is utterly correct, may not be best for our users.