

## We Will Never Have Perfect Data

*An interview with Thelma Ross*

One of the most important thematic lines of the 2018 FIAF symposium was the question of metadata sharing. One of the three FIAF expert committees is dedicated to cataloguing, with the FIAF Moving Image Cataloguing Manual (2016)<sup>1)</sup> being its most important output in recent years. The cataloguing manual brings concepts from the library environment to the film archive environment to help unify the way we organize and describe film collections. Building on the international standard EN 15907 and a hierarchical view of what the cinematic work is — a four-tiered model of work, version, manifestation and carrier — at least for now appears to be the most promising solution for a comprehensive description of film collections. Thelma Ross of the Museum of Modern Art (MoMA) Film Center is Head of the FIAF cataloguing committee.

— — —

*As an introduction, please tell us a little bit about the Cataloguing and Documentation Commission (CDC) of FIAF, your mission, and aims.*

The general purpose of the CDC is to promote exchanging information and expertise and resources. We really mean concrete implementable resources, useful not just among FIAF colleagues

but for the wider community. We focus on the areas of cataloguing moving images but also of documentation related to cinema. Our goals are to get a sense of the working procedures in the areas of cataloguing and documentation in film archives and to help facilitate their standardization. We want to encourage schemas and philosophies that lead to interoperability of systems and the sharing of metadata through various communication protocols. As a commission we meet regularly, conduct workshops, and both create and use guidelines and standards in our institutions.<sup>2)</sup>

*You mentioned the wider community: how does it reflect the composition of the CDC, actually?*

One of the great strengths of the CDC membership is that we have a wide mix. We have academics, database system experts, cataloguers and metadata experts, documentation specialists... And we come from a variety of institutional environments, such as libraries, archives and museums. In terms of academic background, many of us come out of programmes that are designed to educate us in the cultural heritage institutions.

*Following on you mentioning the wider perspective, what is your experience with implementing the best practices in varied environments, such as*

1) See: <https://www.fiafnet.org/pages/e-resources/cataloguing-manual.html>.

2) See: <https://www.fiafnet.org/pages/E-Resources/Cataloguing-Documents-Commission-Resources.html>.

*yours — a museum which operates both a film collection and a library?*

That's what's been interesting about the publication of the Manual, because it came along as a tangible thing I could point to — in conjunction with the metadata schema, the CEN standard for moving images (<http://filmstandards.org/>). With these resources at hand, I became much more successful at advocating for moving image metadata and advocating for it being structured appropriately, so that it is not marginalised in a larger environment which has competing concerns. I have seen the same happen in other major institutions that our commission members come from, such as the BFI or the CNC. And even if within MoMA we have not adopted the one major system the others have (Adlib), we have been able to structure our data in a way that serves the needs of the collection. And again, in order to make my case for that, I relied on these existing standards.

*We can also see the effect of there being a book that you can put on people's desks. Could you comment on it still being a manual, not necessary a guideline?*

We've worked really hard on not being prescriptive, because we thought that would be unrealistic given all the differences in institutional environments, mission statements and collection sizes' budgets. We wanted to provide more than one approach. Even if you have a flat file system, we were conscious to provide guidance — if you can't do it this way, maybe there is another way. That doesn't mean you can't describe the entities, agents and relations that you need. A lot of what I did even at MoMA — my fight was not to make the system architecture comply to a four-level hierarchy system, my fight was to achieve that even within an existing system; all the entities I care about are represented. So, I have kept stressing that there is no pure approach here; the approach is flexible. Having the benefit of a varied membership of the CDC, we knew everyone would have to accommodate.

It might be frustrating, because people really like standardization to mean that we are all doing thing exactly the same way. To me it means we can take a variety of approaches to attain the same ends.

*What are the same ends — would that be the interoperability?*

I think that the end goal is that we have come to an agreement about certain concepts and how we understand them. So that when we are looking at anyone's data we are able to recognize them — whether we use the same field names or not is irrelevant to me. It facilitates the sharing and borrowing of data and making the information more accessible through tools like linked open data.

I believe that if we — like the library community which we have been trailing this whole time — make an agreement that these are the entities, relations and attributes of importance, it will facilitate sharing our data as well as making access to our collections easier.

One of your questions was about the needs of our users. One of the things we can do is invest in our data, which to me means making schemas that are appropriate to your materials a priority, just like we make investments in vaults and necessary storage. That is something we do without much discussion around it.

*What can be learned from the libraries, for example, in the field of authority files?*

The libraries community seems to have really figured it out. If you need to do name authority work, you just go to the Library of Congress name authority files or somewhere like the Getty Institute, and you don't have to create the record from scratch.

And with linked open data, you don't even need to copy that data; you just use the URI to take the user to the data. If I didn't have to research my constituents, I'd be saving so much time, and it would change the basics of cataloguing if we went down that road.

*I presume you are in touch with people looking for data and looking for films. Are you always able to explain to them the specifics of film collections' cataloguing? Are you often confronted with their expectations which might be coming from the library field?*

That's not actually a part of my everyday job. But just in my life at conferences and social events where there are academics, I seem to realise people often lack the context. We all have to understand that our profession is still in its infancy, while libraries have been around for much, much longer. I don't think it's surprising we lag behind, because in many countries we still try to legitimize ourselves as a profession, even. What's been helpful were the various study programmes, for example.

My personal experience is that when I tell people what I do, their response is often "I didn't even know a thing like that existed in the world". But if I had said I was a librarian, everyone would immediately know what that is. Here, digital has both complicated this and created an opportunity for advocacy, because people want everything streamed — and now!

*Would you be then using the term "film librarian" as a means of explaining your profession? Or would you rather get rid of the librarianship context and just say you work as a film cataloguer and that we're all film and moving image archivists?*

I always say I'm a film archivist first, because I feel like they need to know that exists as a thing in the world. And then I say what I do is what you would understand a librarian does; a librarian who describes books in a catalogue and you go down that system and you find the book that you want.

*How do you see the dilemma of openness versus archives as hierarchical keepers of knowledge?*

First of all, we all know that there are limits to openness when it comes to sensitive information about collections. We ethically are guardians of

information about donors or financial figures that we don't want to and should not share.

Among the various institutions, I have seen and experienced a willingness to solicit and crowd-source information while at the same time remaining a source of authoritative information about the items in their collection. So, this is not in favour of one or the other, but a kind of merging of the two, where users or the public get interested and want to engage, and the institution embraces that engagement.

*Any good examples here?*

For instance, when the Library of Congress published all of their still images as digital thumbnails, and they invited the community to tag it... That made search and discovery available for thousands and thousands of photos that had been previously inaccessible.

I also worked with Documentary Educational Resources ([www.der.org](http://www.der.org)), and they want to put together a union catalogue for ethnographic films. One of the things that they want to do is to make it possible for the community being represented in these films to be able to comment and provide information. And that is a wonderful marriage; I think we should engage our community and use their information. Also, there's some great information out there.

On the other hand, I know institutions invest in building authoritative data, and that shouldn't be dumped either. But I think the complication is, of course, how do you moderate incoming information? And how do you incorporate it within a system that has been standardized? That is still a challenge that I don't really have any answer to, but I think it's worth the fight. It's worth trying to use user-generated content. And that'll come with linked open data and with wiki data — things like that. There's obviously a lot of concern over what happens if we make our data open, and we have made mistakes.

But we will never have perfect data. And if we wait for that day, we will never make our data ac-

cessible. And I think that would be a shame. Mistakes will be more quickly identified and corrected if we make our data open. That might be a benefit that we don't talk about; we worry too much about not being perfect.

**Ladislav Cubr – Matěj Strnad**