FUTURS FANTASTIQUES

Vendredi 10 décembre 2021

Grand Auditorium

Discussion conclusive autour des enjeux de l'IA en bibliothèque,
Laurence Engel, **Mike Keller**, **& Aslak Sira Myhre**

Emmanuelle Bermès:

So please let me introduce Laurence Engel who is the President of the National library of France, Mike Keller who is the Vice Provost and University Librarian and Director at Stanford University and online we have Aslak Sira Myrhe the Director of the National Library of Norway, who will be participating this debate remotely. Yes here he is, thank you for joining us I know it was a bit challenging for you so it's great to have you here for this panel. So I feel very lucky because it was complicated to organize things for this conference with the situation of many people being not able to travel and many uncertainties until the very last minute. So I'm very happy that actually for the closing panel I have the three directors of the three libraries who organized the first fantastic futures conferences, and I wanted to go back a little on that story of imagining these fantastic futures not only for libraries but also for archives and museums and maybe ask the three of you what prompted you to launch this initiative on artificial intelligence in our sector, in our fields. So maybe I would like to give the word to Aslak first, as he was the first to organize the conference if that's okay for you.

Aslak Sira Myhre:

Thank you. Thank you very much for having me here if you hear some sounds that are not my voice it's because I'm in a transit hall of an airport in Norway right now. Corona changes everything also the environment I'm here from. But to your question, I think back in 2019 2018 when we started working on this such national library of Norway AI, ML was happening all over the world so you couldn't open a newspaper you couldn't go online you couldn't see a TV show without hearing about AI and what happened around it.

And the first thing that prompted us, at national library of Norway at least, to start with this is that we know that AI and machine learning would be perfect for knowledge organizations and libraries. We know that there's a development in society of tools for this, but we also know that libraries and knowledge institutions will never have the funding, the money or the power within the market to be first in line for the development. So we should do something on our own, something that both can utilize the tools that are already being developed but also point a direction for developing our own tools, and we knew that the only way we could do that was in cooperation. We had the experience of the IIIF corporation with Stanford and many of you others already, and we also knew that at least at the national library of Norway we had capital, not in capitalized money but capital as a huge for us digital collection both of image, of sound, of living image and of documents that would be the playground and a good playground for AI. So we did two things, one thing was starting work on the AI lab of our own in Norway but secondly also reaching out to the library and knowledge community to find other institutions that was kind of the background I think primarily and of course the

background is always that we knew that if we did this we could create wonderful stuff, we could make wonders, because what we keep as we all know is the knowledge of the world and every time there's a new tool like AI and machine learning for humans to work with it can create more knowledge based on our institutions collections.

Emmanuelle Bermès:

Thank you for these insights and we can tell that you are actually in an airport. So Mike you were next in line and you hosted a very nice event in 2019 that I had a chance to attend so, what prompted you?

Mike Keller:

For a long time before Aslak and I got together, I had been thinking of a way of making use of AI to extract information that would be a substitute for the silliness of the MARC record fixed in variable fields, exquisite rules to make sure that everything was perfectly lined up which never of course happened, but in addition I had been on the board of a small Swiss biotech company that made use of AI to produce hypotheses of relevance between theories of action for human peptides and rare diseases. And I saw as soon as we had implemented that AI analysis and presentation to good effect to the company that there were some potential for us in the library business information management in a very specific way.

In addition of course we were aware of various AI activities going on around the Stanford campus, and Aslak actually was the catalyst for us getting together and we decided to put on this first conference in Oslo in this month in 2018 (and roughly the same days !) : it was a one-day event with about 350 people attending in person and thousands streaming. It was a great event and I must say the presentations there were admirable, some of them were practical, some of them were experimental in great contrast to what we see in this conference today which has had a lot of practical applications, meaningful experimentation, new thinking, new approaches, new architectures across a very wide range of contents and contexts. So I think in the few years since we started this thing, it's become really a much more interesting, much more obvious set of possibilities to all of us and Aslak thank you for being that catalyst and Laurence before I pass this over to you, thank you for continuing the effort. The fact is that the Bibliothèque nationale of France from really early 2000s has had marvelous technical people, really bright, really engaged, very confident and marvelously productive I would say. This series of activities in which we have been immersed, in addition to many others, proves the case all over again: this is a terrific staff here, very well supported, but also a very accomplished, so it's a real pleasure to be here with you.

Laurence Engel:

Thank you. Thank you very much, can I speak in French?

Emmanuelle Bermès:

Yes sure we have translation.

Laurence Engel:

It will be easier. Sorry because you don't have a translator.

Je pense qu'on partage en tout cas tous cette idée d'une nécessité. Tu dis Mike qu'à la prochaine réunion, peut-être que partout il y aura cette évidence de l'utilisation de l'intelligence artificielle, mais ce qui nous a conduits à porter depuis de nombreuses années ces questions à la Bibliothèque, c'est cette conscience déjà d'une part que l'intelligence artificielle est là et qu'il ne s'agit pas d'inventer quelque chose mais de s'en saisir, d'autre part, comme Aslak l'a dit, que l'intelligence artificielle nous est utile pour pouvoir continuer nos missions quand on est confronté à des masses de documents et des masses de données entourant les documents comme celle auxquelles nous sommes confrontés et auxquelles on sera de plus en plus confronté avec le développement du dépôt légal numérique. Donc cette conviction que l'intelligence artificielle peut nous aider à accomplir nos missions mais cette conviction aussi que nous sommes un lieu où les développements de l'intelligence artificielle sont les bienvenus ou en tout cas parce qu'on a justement cette masse, on peut être à l'initiative ou donner envie de développements utilisant la technologie de l'intelligence artificielle qui soient utiles à nos missions et donc à des objectifs autres que ceux que portent les grandes entreprises commerciales. Donc il ne s'agit pas de s'opposer à elles, mais il s'agit de dire que l'intelligence artificielle, si elle est développée dans des institutions comme les bibliothèques, peut servir à autre chose, que l'on peut porter peut-être mieux que d'autres tous les enjeux qui sont associés à l'intelligence artificielle en termes d'efficacité, en termes de productivité, en termes de réalisation des missions, mais aussi pour tenir compte des inquiétudes qui sont celles des usagers en termes d'éthique, en termes de transparence des algorithmes, enfin toutes les questions auxquelles on est confronté. Et alors peut-être qu'en France on avait très envie de rentrer dans ce mouvement parce qu'on avait l'expérience d'interrogations de cette nature dès le développement du numérique, donc ce n'est pas exactement la même chose mais en tout cas c'est la même question qu'on se posait quand il y a 30 ans on a commencé à se poser la question de la numérisation des collections : la tentation – et on a tous été confronté à cette tentation – était de se dire nous on n'est pas capable de le faire, il faut que l'on ait des partenariats avec d'autres pour le faire, ce sera plus efficace, ça nous coûtera moins cher, etc. On sait 20 ans après que le choix que nous avions fait d'être dedans, « to be in and not to be out » a été le plus efficace pour nous parce que c'est ce qui nous a permis de rester à bon niveau, de continuer en effet à assumer pleinement nos missions, d'être indépendants, d'être plus indépendants et de ne pas abandonner le terrain à d'autres donc je pense que c'est cette conscience-là qui est essentielle.

Emmanuelle Bermès :

Merci beaucoup. You emphasize the challenges of ethics, transparency, and also to be in with the big players, the big private players like Google Amazon Facebook and others and that's definitely one of the challenges that we have.

Since we started this community, four years have passed and although Mike has touched on it a little, I wanted to ask you what change do you see during these years, where does your library stand in terms of development of that kind of technology? Do you want to follow up?

Mike Keller:

I do thanks. One development which has created a flood of content that otherwise would have been more or less invisible is the development under the leadership of Tom Cramer of the international image interoperability framework [IIIF] which has appeared on this screen

and in these talks many times in the last couple of days. We've created something that's creating a gigantic flow of new content, new in the sense of it not being on the net. It also has created some new responsibilities – I'll come back to that later. The point though is that there's a lot more data.

The second thing that has occurred has to do with the capabilities of our younger IT professionals and our older IT professionals who are learning quickly, and there is a community inside the Stanford libraries itself of some very bright people who have some very good ideas and are ready to put those ideas and their talent to good purposes. As Nicole Coleman has mentioned we have created something called the research data services which is essentially a customer facing outfit of about 15 or 20 people that we have assembled in a new organization from the sort of distributed community of practitioners in support of research by professors and students particularly graduate students and post-docs.

The other things that have happened: let me give you a little resume here. We are working on automated image descriptions on two collections the Andy Warhol contact sheets which Stanford's Cantor Arts Center acquired a bunch of years ago, and the field imagescaptured at the catholic archaeological site which is one of the oldest continually researched ancient towns or villages with several many many layers that Claudia Engel is deeply involved with and has been so for many many years.

In addition, we are imagining how we might use AI through a third party as a company to understand what additional descriptors and concepts we might make use of from the digitally readable collection items that we have, and they're very large number of them: 2.5 million e-books, 75 or 80 000 journals in digital form, many of them having had their back sets digitized so those are actually readable, lots and lots of datasets, lots and lots of other tech reports and so forth in digital form, and how can we best exploit the capability of AI to help us understand the concepts in these books and give better and more precise ideas to our readers of what is in these collection items? Our new head of our Hopkins Marine Station Library (the Miller Library), practicing oceanographer, took her job a couple of years ago and realized that there were lots and lots of biodiversity data and other kinds of data in the lab books that had been placed in the Miller Library down in Monterey. She's decided to digitize them, and as a result we're building backwards of a big collection of digital information collected over the last, I would say many decades, that are highly relevant to climate changes in the way the oceans are exploited and lived in and so forth, this project is a very personal one for her but we've supported it in part because we know that the combination of the converted data from these notebooks which often are data collected on long voyages across the pacific and back are not commonly known and yet are quite important with regard to the environment, particularly the pacific facing environment.

Some of you may know we have had several years of good support from the A.W. Mellon Foundation in New York on the use of linked data (RDF triples and RDF quads, and so forth) in developing ways to more quickly and accurately understand the possibilities of the items that we own, and how we process them and how we make them available, and how we relate other items that have linked data representations, so that they're more easily discoverable. This work has produced more great ideas and there's an element of machine learning in all of this. Thankfully, we are working closely with our colleagues and friends at the Casalini Libri operation on this project along with a few dozen other institutions and

there's a very nice example of a linked data catalog which bears remarkable resemblance to data.bnf.fr which by itself is a marvelous development of some age now.

There are a couple of research projects that we're supporting that are interesting: a group of humanists are working with a couple of linguists and I think one computer scientist, to study the social lives of concepts. What is the meaning of a 17th century term to a trial, a legal process, what is the 19th century concept of what is a trial in some detail, what is a 20th century post-World War II sense of what a trial is, and so on and so forth, the idea being to try to understand the vagaries of time and place and context over these centuries with regard to lots and lots of concepts. It's not an easy task, this project's been going on for a while but especially Nicole Coleman (who by the way has provided me with these notes) is working closely with this group.

There's another activity at Stanford that we're supporting called the Omics Project which is examination of the big data projects in genomics proteomics phenomics, projects that are quite separate in some respects, but in other ways will be more, how to say this, more interesting if we provide them with other digital resources in the form of demographic information, climatic functions, migration and even (God help us) epidemics. We are working with some other libraries around the world to understand what we can produce for these people and we are beginning to produce it.

We have over 3.5 million books that have been digitized by Google. We've downloaded many of them but not quite all of them yet, but they have already become an object or a collection for research, and there's a particular scholar at Stanford who is trying to operate on about 3 million of these books. Such a large corpus, there is no capacity at Stanford, which has a great many servers and a lot of memory, to support the research! So we went back to our friends, colleagues and competitors at Google and they are supporting this with a lot of cloud storage and especially cloud computing.

So more to follow on that, but it's an example of what the future is, we individual institutions will not be able to support the biggest use of big data we have to go to cloud compute and cloud storage. But the good news is that we can do it. And we can do well with it. I think the idea that we're going to be moving away from our dependence upon local iron to distant and virtual memory is really powerful.

We have a final thing that I want to mention because it's of such importance in America and particularly in my heart: we have had as you know terrible murders in America very publicly known, publicly prosecuted (some of them to good effects, some of them not so good effect), but we've decided in the libraries that we have to take up a public position on these, and we have appointed one of our librarians to be essentially a social values librarian curator, and she and some of her colleagues have started something called “no systematic racism” which incorporates attention both to these terrible murders and their implications over the course of the last 70 years really, and recently, not just with regard to murders though but also with regard to racist behaviors by various officials and even citizens, on the basis of complaints or basis of mistaken identities of people of color. I have to say that, the San Francisco bay area is a marvelous melting pot, with lots of tolerance among the various communities and entities around the bay and around the northern California, but there are some truly outstanding evils that we're intending and are confronting. And when we get the data that we are owed by our own government agencies,. we will be able to do more in the way of presenting the range of

difficulties and possibly through our friends at the law school the range of solutions. We're not doing this just for public purposes understanding which is enough to do, but also to see if we can't help correct these attitudes which are unworthy of the American dream and indeed of the French dream.

So that's my recital. Oh there's another one we're working on: that's entity extraction, offers some promise in addition to the new concepts and just extracting lots and lots of entities that get affiliated with particular metadata records in whatever form, better discovery will be the result. Thanks I've talked enough.

Emmanuelle Bermès:

Thank you it really demonstrates the diversity of the realm of AI, the potential of AI in libraries and also the social impact and thanks for emphasizing that.

Maybe I will ask Aslak to follow on because I do remember a few years ago we invited at the BnF Svein Arne and the team to present a proof of concept very convincing that they had realized but I understand they've worked a lot since. So, any insights on what you're doing or what you're planning currently?

Aslak Sira Myhre:

First of all a big thank you to Mike for what he told us and informed us about, now I think what they work on and Stanford actually show is the extreme good of cooperation because Stanford being a university library, national library of Norway of course being a national library, we work in different directions, we work with different aspects of library life, and when we put all this knowledge together as you do at these conferences we really get a feeling of the vast palette of colors, of ideas, of knowledge that we can work with AI and library collections.

For our part I think the biggest change since 2018 is that, in 2018, we called it “fantastic futures” because it was some kind of a science fiction (like dream) to work with AI and libraries. Today, in 2021, we have AI all around the library. It's no longer an AI lab thing, this development has gone extremely fast so if I'm going to say something about how we work concretely with it, I would try to structure it somehow.

First of all we are working with AI in traditional metadata creation. We are working on image, photos, where we use AI to extract information from the photo and from the text around the photo and add it automatically on this metadata on the digitized photo, this is something we need because we are digitizing about one million positive photos (and more than that when we come to negatives and slides) every year and we have no chance to put on metadata manually, and the metadata that are there are scarce, so we have to have this: it's not the question of “should we” it's a question of “how can we do it”, because we need to.

We're also working on the Norwegian Sami bibliography where we are trying to extract, as Mike was talking about, insights from the content of articles to see if they have some irrelevance. This is the minority population donate the population of northern Norway, and we create a bibliography where we now are using AI tools to make this bibliography better.

So those are two examples on traditional library knowledge replacing manpower, humans, with the AI algorithms. But under is the new knowledge, the knowledge that we never made

with human effort and I think some of you have seen this on the fantastic futures conference in Paris now the Maken project this is a project where we have vectorized all our books and all our photos, and then we have created the like algorithm that finds similarities (not based on the way humans see content but the way on the computer or the AI sees content), and then we created a public service, a service for our users: it's mainly for browsing, it's mainly for going around the shelves of a library finding new ideas looking into our material, but then using the AI as a tool to create a user experience and maybe make people open more books and see different kind of aspects of the library.

And then thirdly we showed you something in 2018 which was how we, by using AI, could extract entities, places and people and also connections between entities places and people from our media archive. aAt that point we had taken one month, we had transcribed all the radio from that month, we did TV transcribed and also did imagery analysis and the newspapers. Now we are planning to create a national media archive: this is in our pipeline now for one of the largest projects of the national library for the next years, where all the newspapers (we have now digitized about 80 percent I think of all newspapers produced in Norway), all the radio and all the broadcast, all the TV really put into a service. This service will of course be navigable on day, date and which publisher, but the content is totally impossible to navigate for humans, so we will use the method we showed in 2018 (this is our plan) on the whole volume and, as Mike says, we will have to go into the sky, we will need processors and the capacities that we don't have in our hardware right now, do the analysis and then create a user experience for researchers, for journalists, but also for the public in Norway.

So this is one way we're using it. But I have to mention some more of the ways AI actually is in use for big and large and for small. For big and large, we have since 2018 of the past years created a language model for AI on Norwegian, Norwegian being not French, not English, but a very small language not never the first language for Google or anybody else to create good models for AI or anything else. So we did it ourselves and the AI team has done this with marvelous results, so good results so that the Danes will be using it for their language model as well even though the language is Norwegian, and also, according to researchers and published reports, better than what Google uses for English language today. So this is something we're really proud of and this is basically also what we can do because we have, as Mike mentions he has 3 million books, we have all the Norwegian public life digitized, we have the data that can create the groundwork for AI or for machine learning not only for our institution but also for in our case in Norwegian community. So now we are creating the groundwork for a lot of the AI work in different sectors in Norway by using our collection as the basis.

And then there are all the other things all the add-ons. Just mentioning that for instance now: we just had a breakthrough of the reading of handwriting (the OCR of handwriting : not OCR but reading of that). We have trained – it's Transcribus I think, I think it's a pretty common tool, but we trained it on Norwegian handwriting for the 19th century and we are now at an 80% accuracy on a generic version, which means that…And this wasn't possible when we met in San Francisco two years ago: at that point that was science fiction and we dreamed about it. Now it's possible, we are doing it right now, it's going out into our service right now, and I don't even have the scope of all the places we are applying the generic AI tools that exist now in our organization.

So I think going back to your question Emmanuelle this is the biggest change is that 2018 we were imagining, 2021 we are using and the question is not if we should use it but do we have the resources to, what do we have the resources to actually make use of right now. So I think this is the short version I guess from what we do.

Emmanuelle Bermès:

Thank you very much. At the BnF we're a bit younger on that topic. But we have been thinking about this lately.

Laurence Engel:

Quand je vous entends, il y a effectivement peut-être une maturité un peu moins forte à la Bibliothèque mais le même constat qu'absolument tous les champs sont concernés, sont déjà concernés et sinon vont l'être très vite par l'intelligence artificielle. On travaille beaucoup évidement sur le dépôt légal du web mais aussi sur le dépôt légal des œuvres numériques, c'est ce qui est dans une certaine mesure devant nous encore mais cela veut dire, à chaque fois qu'on lance un programme nouveau (c'est ce qu'on fait depuis quelques années sur le catalogage, c'est ce qu'on fait avec le CNC sur l'entrée des œuvres audiovisuelles), on le fait en introduisant d'emblée des éléments d'intelligence artificielle parce que c'est la condition pour réaliser les choses utilement. On voit aussi que tous les champs de la recherche sont concernés, et c'est la raison pour laquelle dans les projets réalisés ça prend du temps. Mais maintenant c'est ouvert, on a mis en place un DataLab – BnF DataLab – qui part de cette idée que les objets de la recherche (c'est tous les exemples qui ont été donnés par Mike) évoluent et se nourrissent des possibilités offertes par l'intelligence artificielle. Le DataLab, c'est la possibilité, dans les murs, d'utiliser les données associées à des catalogues, de constituer des corpus de données au-delà des corpus d'œuvres elles-mêmes, et de les constituer pour pouvoir y travailler, donc d'offrir la possibilité à des chercheurs qui ont besoin de ces corpus de le faire de manière protégée, en protégeant le droit d'auteur, en étant accompagnés de spécialistes de la manipulation des données pour pouvoir réaliser leurs propres recherches. Et donc ça c'est un service qui a été ouvert cet automne, donc c'est vraiment tout récent : l'aménagement d'espaces, la possibilité de se retrouver, la possibilité de travailler sur des outils numériques et qui sont susceptibles de mobiliser bien sûr l'intelligence artificielle.

Donc voilà pour le passé. Et puis des projets en cours qui touchent vraiment aux missions fondamentales de la bibliothèque : le catalogage (je le disais), mais aussi un outil en cours de construction, d'aide à la décision, d'identification des problèmes, d'analyse de l'état de conservation des collections comme aide à la décision pour les travaux que nous engageons en matière de conservation ou d'intervention de restauration, donc un outil qui s'appelle DALGOCOL et qui utilise l'intelligence artificielle puisqu'il s'agit d'avoir des algorithmes qui permettent de mieux identifier l'état de conservation des collections.

Donc voilà, une grande quantité de projets déjà en cours, et peut-être parce qu'on a un esprit cartésien qui nous amène à tenter de formaliser les choses (mais je crois que ça a été évoqué déjà ce matin), la volonté de formaliser cet ensemble dans une feuille de route, de construire un programme d'une certaine manière. Pourquoi ? Parce qu'on a besoin de faire en sorte que tout le monde soit bien conscient que l'intelligence artificielle est déjà là, que l'intelligence artificielle peut intervenir dans tous les champs de développement des activités de la bibliothèque. Pouvoir aussi fédérer les équipes, les amener à se former (il y a, tu

l'évoquais [Mike], la question de la formation qui est en jeu), donc formaliser l'ensemble des projets, identifier aussi les projets pour pouvoir mieux les financer simplement en les ayant mis dans une feuille de route. Donc ça c'est ce qui est en cours, Emmanuelle en parlerait bien mieux que moi parce qu'elle est à la manœuvre, et l'objectif est de vraiment, on a déjà une première version de cette feuille de route en début d'année, est de ne pas attendre longtemps, d'arriver d'ici la fin de l'année ou à peu près à rendre publique cette feuille de route pour pouvoir identifier les priorités dans chacun des champs que j'évoquais tout à l'heure. Donc effectivement on voit à l'évocation de tous les projets tout ce qui est devant nous encore.

Emmanuelle Bermès:

So yes now the DataLab is open and we have this Roadmap that we want to publish very soon and of course it will be in French, but we will translate it.

Laurence Engel:

We did for the Digital Roadmap.

Emmanuelle Bermès:

Yeah exactly.

We have a digital Roadmap that is already translated into English and Spanish so it's definitely an easy task.

Laurence Engel:

It's an invitation for me to speak in English?

Emmanuelle Bermès:

No, I'm thinking that I'm just translating because I'm not sure Aslak gets what you say. So I'm just summarizing so that he has a sense of what you are saying.

I have a last question and I'm going to ask you first, to you Laurence, because we are currently projecting ourselves in the future with the "Contrat de performance", this exercise that we are doing to try to say what we are going to do in the forthcoming years. So very shortly because we are running out of time: where do you see us or what important thing would you like to have achieved in a few years?

Laurence Engel:

That's why we want to finish our roadmap for artificial intelligence also because it's necessary to put it inside our contract with the Government. I come back to the French.

Alors peut-être plutôt les défis auxquels on est confronté (et c'est pour ça qu'on essaie de formaliser les choses) : c'est l'industrialisation, c'est de passer de l'expérimentation au développement général, à l'utilisation normale de l'intelligence artificielle dans nos projets, et donc d'arriver à le financer aussi, donc la contractualisation n'est pas complètement anodine. Et dans ce cadre, on a effectivement mis en avant des projets qui peuvent nous permettre de créer les priorités. Ce qui, du côté de la relation avec les usagers, touche à la fouille de

données ou à la fouille d'images fait évidemment partie des projets prioritaires qu'on essaie de faire financer en premier. Le projet que j'évoquais, relatif à l'aide à la décision en matière de conservation, est évidemment très important aussi, et le catalogage l'est tout autant, donc j'aurais tendance, dans la longue liste des projets qu'on a mis sur la table, à mettre ceux-là en priorité, avec la conviction, concernant le financement, que justement, parce que tout le monde est préoccupé d'intelligence artificielle, parce qu'on apporte un certain nombre de réponses sur les questions que se posent les citoyens, on a aussi l'opportunité de trouver des financements qui ne sont pas forcément nos financements habituels : en France on a toute une série de projets (projets d'investissements d'avenirplan de relance…), dans lesquels tous les sujets numériques et tous les sujets relatifs à l'intelligence artificielle sont en bonne place. Il est vraiment essentiel qu'on arrive à faire valoir nos projets pour réussir à les financer. Et peut-être pour revenir sur tout ce qu'on a fait ensemble, le fait d'être ensemble est là aussi, je pense, quelque chose d'essentiel. C'est de toute façon le propre des bibliothèques de travailler ensemble mais cela témoigne de la pertinence des projets que l'on porte et cela nous aide à les déployer aussi vis-à-vis de nos financeurs, chacun dans notre pays. Donc travailler ensemble ce sera une priorité.

Emmanuelle Bermès:

Merci Laurence.

Aslak, would you like to say a few words about where you would like to be in three or five years from now?

Aslak Sira Myhre:

In a perfect place! But expect for that, I would say that in three or five years, we will be (it's not a question of what I want or not), we will be in a position where AI tools will be integrated in library system, in library tools of other things, we will be using AI as a "shelf ware" as we call in Norway (it's a part of the tools we have in our toolbox). That will not be a problem, we will not need an extra effort for that, we will not have to be meeting in conferences like this to talk about it.

But there will be the second need, also in three and five years, and that's where we have to be inventive, where we have to be a high mind, where the commercial producers of tools will not produce the tools that are precise enough for our use. And I hope for that part we will be using AI in almost every aspect of our digital collections at least, we will be using it to recreate like Mike precisely said the stupid MARC fields as long as we still need them. So we'll spend less and less human manpower or human power to create MARC fields or that kind of structured metadata, particularly because most of those metadata will be born already with a digital object. But I hope we also be using it to extract knowledge that humans never have had the possibility to extract. The fact that we can read a million newspapers in a couple of days and extract knowledge from it is marvelous, it's still fantastic, it's still a fantastic future and that we are using this and all the other possibilities for analysis of content that AI gives us in a direction that is not just compatible to our collection but good for our collections.

So this is something where we have to be inventive I think, where we have to cooperate and where we have to share.

And then the third thing which is my favorite and maybe it's a bit childish but I also still hope that we have a playground, a sandbox, somewhere where we can experiment with things that are not necessarily that utilistic. They're not there to solve a problem that a researcher has or a problem that the catalog librarian has, but more to see can we do something fun, can we do something great, can we create something spectacular… I hope we'll be there too of two reasons: first of all fun is fun. You don't need to legitimize fun, fun is fun. But also I think we need that kind of playfulness around our collections because more and more we will create digital knowledge that we put out on the web in amounts that nobody will be able to digest. So if you want people to go into Stanford's collection, into the BnF's collection, the national library of Norway's collection or other library collections, we have also to make it fun. We also have to be somewhere in the market of entertainment and creating our educational entertainment if you like so.

So I hope those three things will be there and that we still have time to have some fun with the AI.

Mike Keller:

Well said Aslak as ever.

So for Stanford, I've been talking with Tom Cramer about hiring our own data scientist. We have excellent support for data scientists in this new division that has been mentioned, but we don't have anybody who's looking internally at what we do and how we do it and with what we do it, we're going to have such a person on staff in the next year I would say.

In addition, I know that because of our work with the share-VD [Virtual Discovery] folks at Casalini Libri, that we have, I think, all of our bibliographic records converted to linked data and I expect that will give us a whole bunch of new opportunities, not just for that collection of triples but also for reconciling them with other triples out there and using AI to organize and understand the relationships and begin to understand how far out we have to go to match the needs of people who have a thought, who have a requirement for some sort of information or other.

But is hidden, not in libraries but elsewhere: how can we exploit the contents of these national collections of moving pictures and audio-visual materials, how can we understand how the (in the music world for instance Peter Broadwell's wonderful presentation), how can we understand the influences or the instructions that we get from a recording made by WC in the early 20th century? That presents interpretation that perhaps has been forgotten! In any case there's lots of new opportunities. I think the combination of AI + machine learning + linked data and I hope a discovery environment for the triple if compliant, image sets will be completed in three to five. We'll see. But there's a lot of work to be done and I don't know about you Aslak, but I find satisfaction in getting the work done or seeing the work get done. That's a kind of play in itself I think.

Emmanuelle Bermès:

Great conclusion.

Aslak Sira Myhre:

Agree on that Mike.

Emmanuelle Bermès:

Do we have any questions in the room for our speakers although it's a bit late but, yes I see a question here at the front.

Mike Keller:

Stand up and speak there's nobody here with a mic.

Emmanuelle Bermès:

Please Aurelia, wait a minute. Lucie arrive avec le micro. Here she comes.

I'm advising to use the microphone because otherwise the people online won't hear.

Mike Keller:

Ah, Online! I'm sorry I forgot.

Question:

Ma question s'adresse aux trois intervenants, on sait que l'intelligence artificielle consomme beaucoup d'électricité, de capacité de calcul. Je voudrais savoir si dans le cadre d'une stratégie d'établissement concernant l'intelligence artificielle, l'impact environnemental est évalué, pris en compte. Pour poser la question autrement, est-ce qu'on peut savoir par exemple quel est l'impact carbone du traitement par l'intelligence artificielle de la collection (par exemple du dépôt légal) par rapport au fait d'envoyer quelqu'un dans la station spatiale internationale ? Merci.

Laurence Engel:

The question is about the cost of artificial intelligence for the environment.

Emmanuelle Bermès:

So the question was about environmental impact of AI and do we have any way to measure this impact and compare it with other things like international travel into space or things like that.

Laurence Engel:

So I can begin if you want because it's quite easy to answer but I don't have any figures. But it's obvious.

Je pense qu'il y a 10 ans, on ne se posait pas cette question, mais on ne se posait pas davantage la question du coût environnemental d'autres modalités de production comme on se la pose aujourd'hui. Donc je n'ai pas de chiffres parce que je ne les connais pas, en revanche dans notre feuille de route cela est absolument évidement parce que l'actualité nous permet de nous poser cette question, on ne peut pas ne pas se la poser. Il ne s'agit pas de se dire que l'intelligence artificielle comme le digital d'une manière générale est la réponse à tout. Mais comme on l'a dit depuis le début, il y a des choses qu'on ne peut pas faire sans l'intelligence artificielle, donc dans ce cas-là, il faut qu'on s'autorise l'utilisation de

cette technologie, ne serait-ce que pour éviter que d'autres ne l'utilisent à notre place parce que cela répondrait mieux aux attentes des usagers.

Et vous parliez du dépôt légal : pour le dépôt légal numérique, si on veut poursuivre notre mission de conservation, de construction d'une mémoire et de conservation d'une mémoire collective, on ne peut pas faire l'économie de l'intelligence artificielle pour être en capacité de conserver les archives du web durablement, donc c'est en ces termes qu'il faut raisonner. En revanche, bien sûr, on ne peut pas considérer que l'intelligence artificielle est la réponse à tout, et d'ailleurs tout le monde, pour toutes les recherches, on n'a pas besoin d'utiliser l'intelligence artificielle. C'est plutôt maîtriser la technique pour précisément ne pas s'enfermer dans cette technique.

Emmanuelle Bermès:

Aslak or Mike any ideas on environmental impacts of AI?

Mike Keller:

Yeah, so the truth is that there is plenty of data out there in various directions, various topics. It's not just about global warming, it's also about the effects of global warming, it's about what is causing this, how can it be addressed or how are we going to learn to live with it, as the glaciers melt what is happening to the seas... And as you know, particularly in my country in Alaska, there are plenty of people who don't believe in climate warning, warming even, the same people who don't believe in vaccinations, they think it's a matter of personal liberty. Well when these people get their feet wet because the homes on the periphery of the oceans are going to be overrun by very strong high tides and new levels of coastal levels of water at all, there may be some recognition. It is possible for us to now using AI, I believe, to connect the various symptoms of this terrible tragedy that we're inflicting upon ourselves, and so many different ways, and it's everything from air pollution to methane release to just misunderstanding the implications of hot weather becoming the norm, and thus the ability to understand how we become, in my part of the world, a very dry climate where formally it was difficult and now it's getting impossible. We need to connect all those dots and begin to take a global coordinated approach to dealing with these problems: they're not going to go away, we are going to experience them and for those of us who live near the coasts, there will be really really bad trouble. The information that we're getting out of the notebooks I mentioned in the Hopkins Marine station in Monterey has to do with ocean currents and temperatures, also a lot to do with the fisheries, and of course the fisheries are terribly depleted now and the ocean temperatures are rising and the currents are changing as well so we're now experiencing a La Niña year a very strange one, that's bringing a little bit of rain but hardly enough, and the average temperature in the Sierra Nevada mountains which is where we had formerly snow melt with snow depths of up to 25 feet at the end of a good year, there is no more snowmelt. There is no more snow up there. They may have a six or eight inch snowfall and they'll be able to ski downhill but within days that melt is gone and formally it would take months for the deep deep snowpack to melt and go down into our reservoirs and ultimately to our homes and to our businesses, that's over. We don't know how we're going to cope with it, and I think it's very important that we all understand the various relationships of the elements of this disaster and quickly, otherwise we're all just going to be afflicted and not know what to do and where to go. It's really a dangerous time.

Add to that the notion of fracking for oil and gas in our lands which ruins the water table, poisons the water really, is a terrible tragedy as well and we are the food basket (we in California particularly in northern California), the food basket for lots of America if we can't get the water out of the ground because it simply isn't there, a lot of people are going to lose their businesses but a lot of other people are not going to get the food they thought they might get, it's going to be a very primitive situation in about a decade, maybe sooner. I think that AI offers us the opportunity to connect a lot of dots, and at least understand the demolition of our environment in ways that may enable us to take concerted action and maybe to survive.

Emmanuelle Bermès:

Aslak would you have anything to add and it will be the closure I guess of our panel.

Aslak Sira Myhre:

I think if I understand the question correctly, I don't think we have numbers now for what AI cost and harm and damage when it comes to the question of energy use, which is the most relevant I would say: how much energy do we spend on this partly on producing the processor power and the solar power but mainly to the energy that we use on doing an AI search? But I think if we are going into the computer industries the world wide web into the use of AI and processor power and looking for where we could cut to create a right environmental benefit., I don't think we are the starting point, I think the bitcoin mining is where we start. I think it's the instagrams and the snapchats and all these other “less (I would say) important” processes to mankind as compared to us. I think if we come to a point where the human society of the world has to limit the use of data power from environmental reasons (and we can come to that point, because we use a huge amount of energy on compute the processing right now), but if we come to that point where we have to as humans of the world decide where do we use our computing power, I think AI for knowledge institutions and libraries will be one of those who will be standing the longest before it's been cut down. Because what we create is of so big importance for us, as Mike says, connecting the dots not just in contemporary life but also through human history. This ability to connect the dots and create knowledge is also, in general, the ability to create something like a map for the future, so I think we are not the ones who should flog ourselves about the use of computing power on AI: there are other places you should look for environmental benefits.

Emmanuelle Bermès:

Thank you. So we are looking into futures that are not always fantastic but we still have to work on AI to figure them out and probably we will have the need for future, future fantastic conference. On that, I would like to thank our speakers of this panel. Thank you.