

CIDOC NEWSLETTER No. 01/2009

<i>Christian-Emil Ore, Chair</i>	
Greeting	1
<i>Nick Poole</i>	
Documentation is dead. Long live Documentation!	4
<i>Regine Scheffel</i>	
Digital curation from the perspective of museum documentation	7
<i>Jörn Sieglerschmidt</i>	
The spell of ubiquitous knowledge: Europeana, a portal to European cultural and scientific knowledge	11
<i>Regine Stein</i>	
The cultural object and its image: some thoughts on frequent confusions	14
<i>Karl-Heinz Lampe and Siegfried Krause</i>	
Challenges in museum documentation for transdisciplinary information integration: From field names to events	17
<i>Øyvind Eide</i>	
The CIDOC Co-reference Working Group	19

<i>Christian-Emil Ore, Président</i>	
Salutation	1
<i>Nick Poole</i>	
La Documentation est morte. Vive la Documentation!	4
<i>Regine Scheffel</i>	
La conservation numérique dans la perspective de la documentation muséographique	7
<i>Jörn Sieglerschmidt</i>	
Le charme de l'ubiquité des connaissances: Europeana, un portail au savoir scientifique et culturel européen	11
<i>Regine Stein</i>	
L'objet culturel et son image: quelques réflexions sur des confusions fréquentes	14
<i>Karl-Heinz Lampe et Siegfried Krause</i>	
Défis dans la documentation des musées pour l'intégration d'informations transdisciplinaires: Des noms de champs aux événements	17
<i>Øyvind Eide</i>	
Le Groupe de Travail CIDOC sur le Co-référencement	19

Dear CIDOC member,

The present newsletter is the fourth in the sequence after we reintroduced the annual CIDOC newsletter in 2005, and the second with articles and not just abstracts. CIDOC's editor, Maja Sojat-Bikic, has done an excellent job in developing our newsletter. As the Chair of the CIDOC Board, I use this column to inform you briefly about the activities in CIDOC. In this issue I will also share with you some thoughts on why we are doing all this and the role of CIDOC.

On our web page CIDOC's objective is expressed as »to provide the museum community with advice on good practice and developments in museum documentation«. To be able to do this, there are several requirements that need to be satisfied. First of all we need people with first-class competence in museum documentation, secondly we need papers, presentations, tutorials and material on the subject and thirdly we need to make all this available for the museum community.

Cher membre du CIDOC,

Cette newsletter est la quatrième depuis la réintroduction de la newsletter CIDOC annuelle en 2005, et la deuxième contenant des articles plutôt que de simples résumés. L'éditeur de CIDOC, Maja Sojat-Bikic, a fait un excellent travail dans le développement de notre newsletter. En tant que président de CIDOC, j'utilise cette colonne pour vous tenir brièvement au courant des activités de CIDOC. A cette occasion, je partagerai aussi avec vous des pensées sur pourquoi nous faisons tout ceci et sur le rôle de CIDOC.

Sur notre page web, l'objectif de CIDOC est décrit comme »d'offrir des conseils concernant les meilleures pratiques et les nouveautés dans le domaine de la documentation muséale«. Afin d'être en mesure de faire ceci, il y a plusieurs conditions qui doivent être satisfaites. Tout d'abord, nous avons besoin d'un personnel ayant des compétences de premier ordre en documentation de musées, deuxièmement, nous avons besoin d'articles, de présentations, de tutoriels

CIDOC has currently over 450 members via ICOM and several hundreds more are associated via the working groups and participation in the annual conferences. Thus, we have easy access to hundreds of highly skilled professionals. However, CIDOC is what we, all the members and the elected Board members, make it. The work of CIDOC is traditionally produced by the working groups. A working group is simply a group of members and others sharing a common professional interest, and can very well exist just for a limited time performing a specific task like developing tutorial materials, or recommendations and standards in a certain sub-field of documentation.

In principle, the role of the CIDOC Board in this picture is to facilitate the activities of the working groups, initiate the organisation of the annual conferences and establish the necessary infrastructure to disseminate the resulting information. In practice, the Board acts as a communication service working group as well. This is not an ideal situation, not because the Board members get a lot of extra work, it is fun, but because the work is limited to a small group and this can easily create an impression of work monopolisation. Therefore, I invite members to take part in building the communications infrastructure of CIDOC.

The printed newsletters used to be the main source of information about the activities of CIDOC. The advent of the Internet made the printed newsletter in many ways superfluous. However, the Board decided in 2005 that CIDOC should continue to produce an edited newsletter of selected papers and other information. The newsletter is mainly distributed electronically. A limited number of paper copies are printed and sent to libraries and to members without working email addresses.

However, the main channel of information is the CIDOC web site. The new site was launched two years ago. It is multilingual and has currently English, French, Greek and Swedish versions. We hope to get a Spanish translation in the near future. The web site contains, as most web sites for such organisations do, information about the annual conferences, the working groups and the organisation. In addition, the intention is that the web site should comprise a complete conference archive and digital library with all CIDOC related documents both published and gray literature. This is quite ambitious, but doable.

In April 2007, CIDOC applied for funding for the project *The CIDOC Strategic Resource Repository* under ICOM's *Support for the Network programme*. CIDOC applied for €5,000 and received €3,000. The Board decided to run the project by using €1,500 from its own reserves. The project was planned in the late summer 2008 and implemented in November 2008 by Meta AB in Sweden. It is currently being filled with bibliographic records of electronic and non-electronic publications and conference proceedings. It is connected to the electronic archive of conference web sites from the last 9 years. Volunteers to help with cataloguing, scanning and keying in information are most welcome.

The current web site has the necessary functionality to serve as a place to find information about CIDOC activities and events, and as an archive and library. That is, it serves as a passive documentation and in-

et de matériaux concernant le sujet, et troisièmement, il nous faut rendre tout ceci disponible à la communauté des musées.

CIDOC compte actuellement plus de 450 membres via ICOM et encore plusieurs centaines lui sont associés par le biais des groupes de travail et de par leur participation aux conférences annuelles. C'est ainsi que nous avons accès à des centaines de professionnels qualifiés. Néanmoins, CIDOC n'est rien d'autre que ce que nous, tous les membres et les membres élus du conseil administratif, en faisons. Le travail de CIDOC est traditionnellement produit par des groupes de travail. Un groupe de travaille est simplement un groupe de membres et autres réunis par un intérêt professionnel commun, et peut très bien exister durant une période limitée seulement, pour exécuter une tâche spécifique comme le développement de tutoriels, ou encore de recommandations et de normes dans un certains sous-domaine de la documentation.

En principe, le rôle dans ce tableau du Conseil Administratif de CIDOC est de faciliter les activités du groupe de travail, de lancer l'organisation des conférences annuelles et d'établir les infrastructures nécessaires à la diffusion de l'information qui en résulte. En pratique, le Conseil Administratif sert en plus de groupe de travail pour le service de communication. Ceci n'est pas une situation idéale, non pas parce que les membres du conseil administratif prennent en charge beaucoup de travail supplémentaire, il est amusant, mais parce que le travail est limité à un petit groupe et ceci peut facilement créer une impression de monopolisation du travail. C'est pourquoi j'ai invité des membres à participer à la construction de l'infrastructure de communication de CIDOC.

La newsletter imprimée était autrefois la principale source d'information concernant les activités de CIDOC. L'avènement d'Internet a rendu la newsletter imprimée superflue à bien des égards. Cependant, le Conseil Administratif a décidé en 2005 que CIDOC devrait continuer à produire sous forme éditée une newsletter d'articles sélectionnés et autres informations. La newsletter est principalement distribuée électroniquement. Un nombre limité de copies papier sont imprimées et envoyées aux bibliothèques et aux membres sans adresses email de travail.

Néanmoins, le principal canal d'information est le site web CIDOC. Le nouveau site a été inauguré il y a deux ans. Il est plurilingue et actuellement disponible en Anglais, Français, Grec et Suédois. Nous espérons obtenir une traduction Espagnol dans un futur proche. Le site web contient, comme il est habituel pour la plupart des sites web de ce genre d'organisations, des informations au sujet des conférences annuelles, des groupes de travail et de l'organisation. De plus, le site web devrait comprendre dans l'idée un archivage complet des conférences et une bibliothèque numérique avec tous les documents connexes à CIDOC, qu'ils aient été publiés ou non. Ceci est assez ambitieux, mais faisable.

En Avril 2007, CIDOC a demandé des fonds pour le projet *Le Dépôt des Ressources Stratégiques de CIDOC avec l'Appui pour le programme Réseau* d'ICOM. CIDOC a demandé €5,000 et a reçu €3,000. Le Conseil Administratif a décidé de lancer le projet avec €1,500 tirés de ses propres réserves. Le projet a été planifié à la fin de l'été 2008 et mis en œuvre en novembre 2008 par Meta AB en Suède. Il est actuellement en train d'être rempli avec les références

formation bank. How should CIDOC support active online discussions? Today social forums are the trendy stuff and blogs are ok. However, email lists are considered to be very serious and old fashioned (at least according to my 15 year old son). Technically it is easy to set up services of all the three types. So far we have set up a CIDOC discussion forum (<http://meta.se/cidocforum/>) and there is a CIDOC group at the LinkedIn network. It is, however, never successful to try to impose any information monopoly. Thus, it is a good thing if there are active discussions with links to CIDOC in other web forums. Still, in my opinion, we should prefer the CIDOC discussion forum.

Let me finally return to the activities of CIDOC. In the last year there have been several results among our active members. Connected to the always-active CIDOC CRM SIG there is a joint working group with the library association IFLA on the harmonisation of CIDOC CRM and IFLA's library model FRBR (Functional Requirements for Bibliographic Records), called FRBRoo. This international working group has met regularly since 2003 and in 2008 the work was completed. The definition of FRBRoo is to our knowledge the first joint standard document between IFLA and ICOM-CIDOC.

The Documentation Standards Working Group is developing a Statement of principles of museum documentation. The current draft is available for comment at the CIDOC web site. This document has been discussed and updated at the last three CIDOC meetings (in Zagreb, Gothenburg and Athens), then tidied up by Andrew Roberts. Once the Principles have been finalised, we intend to offer guidance based upon them. Where possible, this guidance will use existing authoritative sources, such as the UK Collections Trust's SPECTRUM framework. CIDOC has formally suggested to ICOM that the Principles should be discussed at the Advisory Committee Meeting in 2009. In CIDOC's opinion these principles and the ethical code of ICOM should be seen as a whole.

In the period 2005-2008, all the CIDOC conferences took place in Europe. As a part of our effort to reach a truly international audience this year's annual conference will be in Santiago de Chile and the theme is *Documentation in the XXI Century: connecting cultural heritage information*. I look forward to seeing you in Santiago in September.

Kind regards,
Christian-Emil Ore
CIDOC Chair
c.e.s.ore@iln.uio.no

bibliographiques de publications et actes de conférences électroniques et non électroniques. Il est connecté aux archives électroniques de sites web sur les conférences des 9 dernières années. Les volontaires pour aider à cataloguer, numériser et saisir des informations sont les bienvenus.

Le site web actuel a les fonctionnalités nécessaires pour servir de centre d'information concernant les activités et les événements CIDOC, et d'archive et de bibliothèque. C'est à dire qu'il sert de banque de documentation et d'information passive. Comment est-ce que CIDOC devrait faire pour soutenir la discussion active en ligne? De nos jours, les forums sociaux c'est le truc en vogue et les blogs, ça passe encore. Cependant, les mailing listes sont vues comme très sérieuses et démodées (du moins d'après mon fils de 15 ans). Sur le plan technique, il est facile d'établir des services de ces trois types. Jusqu'ici nous avons établi un forum de discussion CIDOC (<http://meta.se/cidocforum/>) et il y a un groupe CIDOC sur le réseau LinkedIn. Cependant, il n'est jamais profitable d'essayer d'imposer un monopole de l'information quel qu'il soit. Pour cette raison, c'est une bonne chose s'il y a des discussions actives en lien avec CIDOC sur d'autres forums web. Quoiqu'il en soit, à mon avis, nous devrions préférer le forum de discussion de CIDOC.

Retournons pour finir aux activités de CIDOC. Durant l'année passée il y a eu plusieurs résultats parmi nos membres actifs. En connexion avec le toujours actif CIDOC CRM SIG existe un groupe de travail, joint à l'association de bibliothèques IFLA, sur l'harmonisation des modèles FRBR («Functional Requirements for Bibliographic Records»: Conditions Fonctionnelles pour les Références Bibliographiques) de CIDOC CRM et IFLA, appelé FRBRoo. Ce groupe de travail international s'est rencontré régulièrement depuis 2003 et en 2008 le travail fut achevé. La définition de FRBRoo est, à notre connaissance, le premier document normé commun entre IFLA et ICOM-CIDOC.

Le Groupe de Travail sur les Normes de Documentation est en train de développer une déclaration des principes de la documentation muséale. Le brouillon actuel est livré aux commentaires sur le site web de CIDOC. Ce document a été discuté et actualisé aux trois derniers rendez-vous CIDOC (à Zagreb, Gothenburg et Athènes), puis arrangé par Andrew Roberts. Une fois que les Principes auront été finalisés, nous avons l'intention d'offrir une orientation basée sur eux. Dans la mesure du possible, cette orientation utilisera des sources préexistantes faisant autorité, telles que le cadre SPECTRUM de l'U.K. Collection Trust. CIDOC a officiellement demandé à ICOM que les Principes soient discutés à la réunion du Comité Consultatif en 2009. L'opinion de CIDOC est que ces principes et le code éthique d'ICOM devraient être perçus comme un tout.

Dans la période 2005-2008, toutes les conférences CIDOC ont eu lieu en Europe. Dans l'optique de toucher un public véritablement international, la conférence de cette année sera à Santiago de Chile et le thème est *La Documentation au XXIème siècle: mettre en commun l'information des patrimoines culturels*. Je me réjouis de vous voir à Santiago en septembre.

Cordialement,
Christian-Emil Ore
Président de CIDOC

Documentation is dead. Long live Documentation!

Nick Poole
Collections Trust
nick@collectionstrust.org.uk

Documenting collections costs cultural organisations millions of Euro and thousands of man-hours every year. It is difficult work requiring an unusual combination of repetition, expert knowledge and careful attention to detail.

It also never ends. As long as there are collections, we will need to capture knowledge about them, associate them with information, manage and conserve them. It is an ongoing business process of any organisation managing large bodies of material and using that material to create inspiring and sustainable experiences for the public.

So, given that documentation is difficult, expensive and never ends, why do we do it at all?

The reason is that we believe that the benefits of Documentation – for our organisations, our collections and our users – greatly outweigh the costs. As the international community it is our role to make this argument, and keep making it now and for generations to come.

The Story So Far...

Documentation is one of the great success stories of the past 30 years of cultural practice. Although different countries are at different stages in the journey, the global message is clear. Cultural organisations throughout the world are stronger, more confident, more professional and better equipped to face the future because they have invested in improving the standard of their Information Management.

Standards such as SPECTRUM and the CIDOC CRM have provided both a focus and a catalyst for this process – bringing together best practice and sharing it throughout the community.

In spite of the great developments, Documentation is facing a new generation of challenges, and it is up to our community to decide how we face them.

Thinking about Documentation Strategically

The first challenge is to learn to think strategically about Documentation itself. This is a profound challenge both to CIDOC and to organisations such as the Collections Trust, the Getty, Canadian Heritage Information Network and others.



La Documentation est morte. Vive la Documentation!

Nick Poole

La documentation des collections coûte aux organisations culturelles des millions d'Euro et des milliers d'heures de travail chaque année. Il s'agit d'une entreprise difficile qui nécessite la combinaison inhabituelle de répétition, savoir expert et attention particulière au détail.

C'est aussi un travail sans fin. Aussi longtemps qu'existeront des collections, nous aurons besoin d'acquérir du savoir à leur propos, de les associer aux informations, de les gérer et les conserver. C'est un travail continu pour toute organisation que la gestion d'importants fonds de matériel, ainsi que l'utilisation de ce matériel, en vue de la création d'expériences inspirantes et durables pour le public.

Ainsi, puisque la documentation est un travail difficile, dispendieux et sans fin, pourquoi nous y consacrons-nous?

La raison est que nous sommes convaincus que les bienfaits de la Documentation, que ce soit pour nos organisations, nos collections ou nos clients, surpassent de beaucoup les coûts engendrés. En tant que communauté internationale, il est de notre devoir de faire valoir cette idée, de continuer à la défendre maintenant et pour les générations futures.

L'histoire jusqu'ici...

La Documentation est une des grandes réussites de ces quelques 30 dernières années dans le secteur culturel. Bien que différents pays en soient à des étapes différentes du périple, le message global est clair. Les organisations culturelles de par le monde sont plus fortes, plus assurées, plus professionnelles et mieux équipées pour faire face au future, parce qu'elles ont investi dans l'amélioration de la qualité de leur Gestion d'Information.

Des normes telles que SPECTRUM et le CIDOC CRM ont fourni à la fois un point de convergence et un catalyseur pour ce processus: rassembler les meilleures pratiques et les propager à travers la communauté.

En dépit de l'important développement, une nouvelle génération de défis attend la Documentation, et il revient à notre communauté de décider comment les aborder.

Penser la Documentation en termes stratégiques

Le premier défi est d'apprendre à penser la Documentation en termes stratégiques. Il s'agit d'un défi central pour CIDOC aussi bien que pour des organisations telles que la Collections Trust, la Getty, le Réseau canadien d'information sur le patrimoine (RCIP) et d'autres.

En tant que communauté, nous avons tendance à nous focaliser sur les aspects pratiques, pragmatiques et techniques. Notre impulsion est d'établir des normes et de penser à des structures et des systèmes. Au cœur même de la Documentation se trouve une appétence d'abstraction et d'organisation du monde en éléments maniables.

Pourtant, notre travail existe au sein d'un plus large contexte politique, financier et stratégique. Nous dé-



As a community, we tend to focus on the practical, the operational and the technical. Our impulse is to generate standards and to think about structures and systems. At the very heart of Documentation is an urge to abstract and to sort the world into manageable elements.

Yet our work exists within a broader political, financial and strategic context. We depend on the goodwill and support of politicians who have little interest in *how* we manage information, but who care instead about *what* benefits this information can bring to our users.

Our community has still to face two battles. The first is to convince people of the *business case* for managing and recording information. The second is to convince them of the *public benefit* which this process creates.

From Documentation to Information Management

The second challenge is to revisit the word itself – Documentation. When we first started referring to Documentation, it was a tool for unification. It con-

pendons du bon vouloir et du support des politiciens qui ne s'intéressent que peu à *comment* nous gérons l'information, mais qui se soucient plutôt de *quel* bénéfice cet information peut amener à nos usager.

Notre communauté a encore à affronter deux batailles. La première est de convaincre les gens de l'*intérêt commercial* qu'il y a dans la gestion et l'enregistrement d'information. La seconde est de les convaincre du *bénéfice public* que ce processus fournit.

De la Documentation à la Gestion des Informations

Le deuxième défi est de revisiter le monde en question lui-même: la Documentation. Lorsque nous avons commencé à parler de Documentation, c'était un moyen d'unification qui liait les gens au delà des disciplines et des langues, créant un sentiment de cohérence à propos de pratiques qui étaient auparavant sans lien.

Récemment, néanmoins, le mot Documentation est devenu problématique. Il a un sens différent et mieux défini dans des industries bien plus grandes telles que le domaine des logiciels et l'ingénierie. Il ne décrit pas pleinement à lui seul la gamme de pro-

nected people across disciplines and languages and created a sense of coherence around what were previously disconnected practices.

Recently, however, the word Documentation has become problematic. It has a different, better-defined meaning in much larger industries such as software and engineering. It does not in itself fully describe the range of processes involved, and is increasingly serving to isolate museum practice from related disciplines such as librarianship and archiving.

Not only this, but it is also helping to isolate people within their own institutions. In many organisations, Documentation is regarded as something technical, behind-the-scenes, disconnected from other parts of the organisation. For this reason, many organisations are rejecting the word 'Documentation' in favour of 'Information Management'.

As an integral part of Collections Management, Information Management is about connecting together the many different internal and external systems of the organisation to avoid duplication, remove the risk of information 'silos' and maximise the overall return on investment.

From Stock Control to Entertainment

A further profound challenge is the shift in expectation about the content which we are able to deliver. The drive to put collections online has created an expectation that we can deliver deep, rich narrative information linked to copyright-cleared photographs with a fully classified metadata record.

Suppliers of Collections Management Systems have responded very rapidly to this requirement, acquiring the functionality of Digital Asset Management, Digital Rights Management and Content Management in the space of less than a decade.

What has responded less quickly is the *quality* of this content. There is a pressing need both to sustain the integrity of management information and to support narrative information and User-Generated Content which can be used to develop entertaining and engaging services.

From the Server to the Cloud

In future, it is likely that instead of installing software on a hard drive, we will be going online to access applications. Instead of buying a server to host information, we will buy access to space on servers distributed throughout the world.

This vision (sometimes referred to as 'utility' or 'Cloud' computing) holds huge potential for cultural services, as well as huge challenges for Information Management.

Cultural organisations tend to localise things. We like to manage our own systems, maintain our own terminologies and deliver our collections through our own websites. In future, however, these things are likely to become distributed throughout the Web.

Instead of maintaining your own in-house terminology, you might contribute to and use a terminology shared with hundreds of other organisations. Users may come to your website to find your collections, but they are equally likely to find them in Google, or a tourism website over which you have little control.

cessus impliqués et sert de plus en plus à isoler l'activité des musées des disciplines qui y sont apparues, telles que la bibliothéconomie et l'archivage.

Mais ce n'est pas tout; il contribue aussi à isoler les gens au sein de leurs propres institutions. Dans de nombreuses organisations, la Documentation est considérée comme une activité technique qui se déroule en coulisse, détachée du reste de l'organisation. Pour cette raison, de nombreuses organisations rejettent le mot »Documentation« au profit de »Gestion des Informations«.

En tant que partie intégrante de la Gestion des Collections, la Gestion des Informations a pour but de mettre en connexion les nombreux différents systèmes internes et externes de l'organisation, d'éviter la duplication, d'éliminer le risque de cloisonnement des informations et de maximiser le retour global sur l'investissement.

De l'Inventaire des Stocks au Divertissement

Un autre défi de taille est posé par le changement dans l'attente publique quant au contenu que nous sommes à même de produire. Avec la ruée vers la diffusion internet des collections s'est créée une attente à ce que nous produisions de l'information narrative et riche concernant des photographies libres de droit, avec des enregistrements de métadonnées entièrement classifiés

Les fournisseurs des Systèmes de Gestion des Collections ont répondu très rapidement à ce besoin, acquérant la fonctionnalité de Gestion de Ressources Numériques, Gestion des Droits Numériques et Gestion de Contenu dans l'espace de moins d'une décennie.

Ce qui a été moins prompt à répondre, c'est la *qualité* de ce contenu. Il existe une nécessité pressante et de maintenir l'intégrité de l'information de gestion, et de soutenir l'information narrative et le Contenu généré par les utilisateurs qui peut être utilisé pour développer des services divertissants et attirants.

Du Serveur au Nuage

Dans le future, il est probable qu'au lieu d'installer des logiciels sur un disque dur, nous irons sur le net pour accéder aux applications. Au lieu d'acheter un serveur pour héberger l'information, nous achèterons l'accès à l'espace sur des serveurs distribués de par le monde.

Cette vision (parfois appelée informatique »du Nuage« ou »de services«) détient un énorme potentiel pour les services culturels, aussi bien qu'un énorme défi pour la Gestion de l'Information.

Les organisations culturelles ont tendance à localiser les choses. Nous aimons gérer notre propre système, maintenir nos propres terminologies et produire nos collections par le biais de nos propres sites internet. Dans le futur, cependant, ces choses seront probablement distribuées à travers le Web.

Au lieu de maintenir ses propres terminologies maison, il est possible que l'on utilisera et que l'on contribuera à une terminologie partagée avec des centaines d'autres organisations. Il se pourrait que les utilisateurs se rendent sur vos sites pour y trouver vos collections, mais il est tout aussi probable qu'ils les trouvent sur Google, ou sur un site de tourisme sur lequel vous n'auriez que peu de contrôle.

Conclusion

Many other challenges lie ahead. At risk is the loss of momentum and going back on the progress made in the past decade, of becoming marginalised within cultural organisations and of losing touch with our audiences.

At stake is the opportunity to define a new vision of information at the heart of everything which cultural organisations do, enabling our industry to become better-integrated, more efficient and to create more inspiring and responsive services for users.

This then is the challenge to CIDOC and the community it represents!

About the author:

Nick Poole is the Chief Executive of the Collections Trust. A leading figure in the field of standards and Collections Management, he is regularly invited to advise UK and international organisations on issues relating to collections, knowledge, new technology and the law. Nick is a Councillor of the UK Museums Association, a member of the Board of the Museums Copyright Group and a Trustee of ICOM UK. He represents the UK within the Member States Expert Group at the European Commission. Nick also lectures and publishes widely on the subject of new developments in cultural services.

Conclusion

De nombreux autres défis nous attendent. Nous risquons la perte de dynamique, le retour sur les progrès faits au cours de la dernière décennie, la marginalisation au sein des organisations culturelles et la perte de contact avec notre public.

Ce qui est en jeu, c'est l'opportunité de définir une nouvelle vision de l'information, au cœur de tout ce que les organisations culturelles font, permettant à notre industrie de s'intégrer mieux, d'être plus efficaces et de créer plus de services inspirants et ouverts pour nos utilisateurs.

C'est là le défi pour CIDOC et la communauté qu'il représente.

A propos de l'auteur:

Nick Poole est le PDG de la Collections Trust. Figure importante dans le secteur des normes et de la Gestion de l'Information, il est régulièrement invité à conseiller des organisations du Royaume-Uni, et au niveau international, sur des questions relatives aux collections, au savoir, aux nouvelles technologies et à la loi. Nick est un Conseiller de l'Association des Musées du Royaume-Uni, un membre du conseil administratif du Groupe Copyright des Musées, et un administrateur d'ICOM UK. Il représente le Royaume-Uni au sein du Groupe Expert des Etats Membres à la Commission Européenne. Nick enseigne également et publie beaucoup au sujet des nouveaux développements des services culturels.

Digital curation from the perspective of museum documentation

Regine Scheffel

Hochschule für Technik, Wirtschaft und Kultur,
Leipzig
Fakultät Medien
scheffel@fbm.htwk-leipzig.de

Museum documentation has seen several fundamental changes of focus. Cataloguing all museum objects was the first step in IT supported museum documentation. Then processes of collection management got into the focus. Rather soon it got evident that museums need not only to catalogue their collections but also their information resources in order to use staff and capacities efficiently and to react to the challenge of giving users access to museum information.

In 2001, SPECTRUM Knowledge¹ formulated »Standards for cultural information management« with the aim a) of higher efficiency in museum information management and b) to create new information services for new user needs. SPECTRUM Knowledge recommends a systematic approach: Information management in museums is to be based on policies on collection, retention – and disposal(!) of information. »Information Service Evaluation« and audits check if

¹ SPECTRUM Knowledge : Standards for cultural information management / compiled and edited by Helen Ashby, Gordon McKenna and Matthew Stiff. Cambridge : MDA, 2001, ISBN 1 900642 08 5

La conservation numérique dans la perspective de la documentation muséographique

Regine Scheffel

La documentation muséographique a connu plusieurs changements d'orientation fondamentaux: Le catalogue de tous les objets de musées a été le premier pas dans la documentation de musée basée sur IT. Ensuite, les processus de gestion de collections sont entrés dans le tableau. Bien vite, il est devenu évident que les musées ont besoin non seulement de cataloguer leurs collections mais aussi leurs ressources d'information afin d'utiliser efficacement personnel et capacités et de réagir au défis de donner aux usagers l'accès aux informations des musées.

En 2001, SPECTRUM Knowledge¹ a formulé des »Normes pour la gestion des informations culturelles« avec pour but a) une meilleure efficacité dans la gestion des informations des musées et b) la création de nouveaux services d'information pour répondre à de nouveaux besoins des usagers. SPECTRUM Knowledge recommande une approche systématique: la gestion des informations dans les musées se doit d'être basée sur les politiques pour la collecte, la rétention – et l'élimination(!) de l'information. »L'Évaluation des Services d'Information« et les audits con-

¹ SPECTRUM Knowledge : Standards for cultural information management / compiled and edited by Helen Ashby, Gordon McKenna and Matthew Stiff. Cambridge : MDA, 2001, ISBN 1 900642 08 5

user needs were met and identify opportunities for improvement. Transparency of historic and contemporary information resources and their use for defined target groups in the museum and within the broader audience is gained by »Information Resource Cataloguing«. The underlying concept is sustainability of information by applying standards and by interoperability.

In many aspects information management according to SPECTRUM Knowledge can be seen as the »predecessor« of digital curation: The analysis of deficiencies in information practice in museums² leads to the focus on reuse of data. Methods of the archival work, like the criteria based selection and the decision about disposal or archiving of information, are applied in digital curation – and in SPECTRUM Knowledge. But digital curation goes beyond that: it bridges the gap between information management and long-term preservation of digital objects. The term »digital curation« evokes the necessity to curate digital information objects like museum objects and to raise this duty to the level of curatorial work in museums.³

In the United Kingdom the Digital Curation Centre (DCC) has been founded to support these tasks. Its definition of digital curation raises focus on the selectiveness of curated information, on its organisation in a »life-cycle« and on the necessity of a proactive attitude towards this task: »Digital curation is maintaining and adding value to a trusted body of digital information for current and future use; specifically, we mean the active management and appraisal of data over the life-cycle of scholarly and scientific materials.«⁴

While SPECTRUM Knowledge argues from the position of a museum towards its information management, DCC argues from the perspective of a digital object. The DCC digital lifecycle model is »OAIS compliant«. This means that it conforms to the elements and processes of the Open Archival Information Systems (OAIS) Reference Model⁵ (ISO 14721) which is widely accepted for modelling digital preservation processes and planning archival systems for long-term preservation. The DCC lifecycle is applicable to all sorts of digital information objects in museums: digital museum objects (digital art, etc.), research data and digital information about museum objects (digital documentation, photographs, databases with metadata about objects, (un-)published articles, etc.).

»Full lifecycle actions« in the centre concern the maintenance of digital data at any time and throughout all sequences of the lifecycle (see figure⁶). They contain the attribution of descriptive, administrative, technical, structural, rights and preservation metadata in order to facilitate retrieval, curation and preservation of digital objects. Preservation planning and

trôlent si les besoins des usagers ont été comblés et identifient les opportunités d'amélioration. La transparence à propos des ressources d'information historiques et contemporaines et leur utilité pour des groupes ciblés définis dans le musée et dans le grand public est atteinte grâce au »Catalogage des Ressources d'Information«. Le concept sous-jacent est la permanence de l'information permise par l'application de normes et par l'interopérabilité.

A de nombreux égards, la gestion des informations en fonction de SPECTRUM Knowledge peut être considéré comme le »prédécesseur« de la conservation numérique: l'analyse de lacunes dans la pratique de l'information dans les musées² conduit à la mise au point de la réutilisation des données. Les méthodes du travail d'archivage, comme la sélection par critères et la décision quant à l'élimination ou l'archivage d'informations, sont appliquées dans la conservation numérique – et dans SPECTRUM Knowledge. Mais la conservation numérique va au delà de cela: elle comble l'espace entre la gestion de l'information et la préservation à long terme d'objets numériques. Le terme »conservation numérique« évoque la nécessité de conserver les objets d'information numérique comme des objets de musées et d'élever cette tâche au niveau du travail de conservation dans les musées.³

Dans le Royaume Uni, le Centre de Conservation Numérique (DCC) a été fondé pour soutenir ces tâches. La définition qu'il donne de la conservation numérique attire l'attention sur la sélectivité de l'information conservée, sur son organisation en un »cycle de vie« et sur la nécessité d'une attitude proactive envers cette tâche. »La conservation numérique maintient et ajoute de la valeur à un corpus d'information numériques de confiance pour un usage actuel et future; plus spécifiquement, il s'agit de la gestion active et de l'évaluation des données sur la durée du cycle de vie des matériaux académiques et scientifiques.«⁴

Alors que SPECTRUM Knowledge adopte la position d'un musée envers sa gestion de l'information, DCC adopte la perspective de l'objet numérique. Le modèle de cycle de vie numérique du DCC est »conforme à OAIS«. C'est à dire qu'il correspond aux éléments et aux processus du Modèle de Référence⁵ (ISO 14721) du Système Ouvert d'Archivage d'Information (OAIS) qui est généralement accepté pour la modélisation des processus de préservation numérique et la planification de systèmes d'archivage pour la préservation à long terme. Le cycle de vie du DCC est applicable à toutes sortes d'objets d'information numérique dans les musées tels que des objets numériques de musées (art digital etc.), des données de recherche et de l'information numérique au sujet d'objets de musées (photographies numérique de

² For example, the last version of information in a printed catalogue often is not being retransferred into the database and therefore lost for further scientific or custodial work.

³ This is important because it demands competent staff with custodial responsibility and status.

⁴ <http://www.dcc.ac.uk/about/> [All URLs have been accessed May 16, 2009]

⁵ Reference Model for an Open Archival Information System (OAIS) / Consultative Committee for Space Data Systems (CCSDS). CCSDS 650.0-B-1. BLUE BOOK. January 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>, p. 4-1

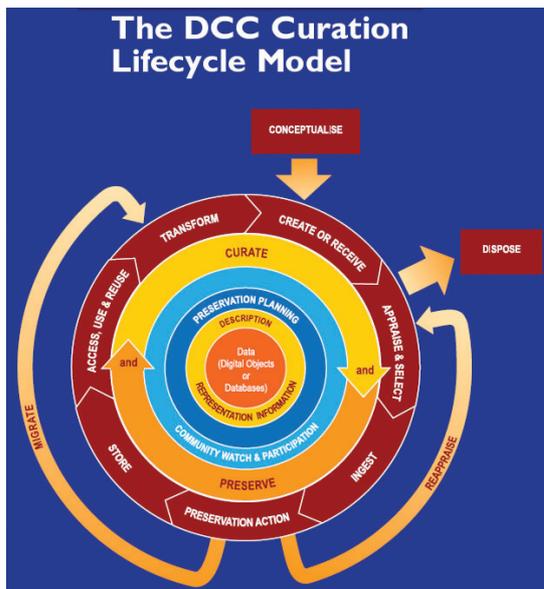
⁶ <http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf>

² Par exemple la dernière version d'information imprimée dans un catalogue n'est souvent pas retransférée dans la base de données et ainsi perdue pour un travail future, scientifique ou de conservation.

³ Ceci est important parce que le personnel doit dès lors être à même d'assumer des responsabilités et un statut de conservation.

⁴ <http://www.dcc.ac.uk/about/>

⁵ Reference Model for an Open Archival Information System (OAIS) / Consultative Committee for Space Data Systems (CCSDS). CCSDS 650.0-B-1. BLUE BOOK. January 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>, p. 4-1



evaluation of processes and their results are ongoing activities concerning all phases of the digital lifecycle. Successful migration or other preservation activities require technology watch (formats, technology, tools, etc.) and community watch and participation in order to build up an efficient infrastructure for digital preservation.

At this point another important resource should be mentioned: The »Digital Preservation for Museums : Recommendations«, published by the Canadian Heritage Information Network (CHIN) in 2004.⁷ Shortly after the publication of the fundamental »Charter on the Preservation of the Digital Heritage« by the UNESCO in 2003,⁸ CHIN gave pragmatic recommendations to museums how to deal with the new challenge of digital preservation. In the context of the digital curation lifecycle it is especially interesting to have a nearer look at the »Action Recommendations«, concerning the elaboration of policies, workflows and the definition of responsibilities within museums, and the »Broader Recommendations« dealing with collaboration and participation with the aim of a digital preservation infrastructure for museums within the cultural heritage sector.

Most interesting in the DCC lifecycle are the »Sequential Actions«, focusing the phase of conceptualisation before the creation (or collection) of digital objects: purpose, target groups (the »designated community« in terms of digital preservation), capture method, format decisions and storage options define the basic decisions concerning the digital object to create or collect. All later decisions have to reflect these initial ones and evaluate if the conditions, under which they have been taken, are still the same, or if they need revision. In the »Create or Receive« phase the addition of metadata is essential. If they are incomplete (e.g. the rights metadata) further usage or preservation actions may perhaps not be possible.

documents, bases de données avec métadonnées à propos d'objets, articles (non-)publiés etc.).

Les »actions de tout le cycle de vie« au centre concernent le maintien de données numériques à tout moment et durant toutes les séquences du cycle de vie.⁶ A elles revient l'attribution de métadonnées de préservation, de droit, descriptives, administratives, techniques et structurelles, afin de faciliter le retrait, la conservation et la préservation d'objets numériques. La planification de la préservation et l'évaluation des processus et de leurs résultats sont des activités toujours en cours qui concernent toutes les phases du cycle de vie numérique. La réussite d'une migration de données ou toute autre activité de préservation a besoin de veille technologique (formats, technologie, outils etc.) ainsi que de veille communautaire et de participation afin de construire une infrastructure efficace pour la préservation numérique.

A ce stade, une autre ressource importante mériterait d'être mentionnée: Les »Préservation numérique à l'intention des musées : Recommandations« publiées par le Réseau canadien d'information sur le patrimoine (RCIP) en 2004.⁷ Peu après la publication de la fondamentale »Charte sur la conservation du patrimoine numérique« par l'UNESCO en 2003,⁸ RCIP a donné des recommandations pragmatiques aux musées pour faire face aux nouveaux défis de la préservation numérique. Dans le contexte du cycle de vie de la conservation numérique, il est particulièrement intéressant de se pencher de plus près sur les »Recommandations d'Actions«, concernant l'élaboration de politiques et de flux de travail et la définition des responsabilités au sein du musée, et sur les »Recommandations Générales«, qui traitent de la collaboration et de la participation avec pour objectif une infrastructure de préservation numérique pour les musées dans le secteur du patrimoine culturel.

Dans le cycle de vie du DCC, les »Actions Séquentielles« sont particulièrement intéressantes, mettant en lumière la phase de la conceptualisation avant la création (ou la collection) d'objets numériques: objectif, groupes cibles (la »communauté désignée« en termes de préservation numérique), méthode de capture, choix de format et options de stockage définissent les décisions de base concernant l'objet numérique à créer ou collectionner. Toutes les décisions prises postérieurement doivent refléter celles prises initialement et évaluer si les conditions sous lesquelles elles ont été prises sont toujours d'actualité ou si une révision s'impose. Dans la phase »Créer ou Recevoir« l'addition de métadonnées est essentielle. Si elles sont incomplètes (par exemple les métadonnées de droit) il n'est pas exclu qu'il soit impossible de continuer à les utiliser ou les préserver et l'objet devra être rejeté. »Évaluer et Sélectionner« implique la sélection de ressources d'information qui valent la peine d'être préservées à long terme. Les autres ressources devront être envoyées à d'autres

⁷ Digital Preservation for Museums : Recommendations / by Tim Au Yeung; commissioned by the Canadian Heritage Information Network, 2004. http://www.chin.gc.ca/English/Digital_Content/Preservation_Recommendations/index.html

⁸ http://portal.unesco.org/ci/en/files/13367/10700115911Charter_en.pdf/Charter_en.pdf

⁶ <http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf>

⁷ Préservation numérique à l'intention des musées : Recommandations / Tim Au Yeung, Université de Calgary, Commandée par le Réseau canadien d'information sur le patrimoine, 2004. http://www.rcip.gc.ca/Francais/Contenu_Numerique/Preservation_Recommandations/index.html

⁸ http://portal.unesco.org/fr/ev.php-URL_ID=17721&URL_DO=D_O_TOPIC&URL_SECTION=201.html

sible and the digital object will have to be rejected. The »Appraise and Select« phase means selection of information resources being worth long-term preservation. Other resources may be sent to other archives, or be destroyed (if necessary in a documented manner like in archives).

The »Ingest« phase means the transfer of the information object into a digital archival system, or – this is different from digital preservation – to another »repository, data centre or other custodian«. ⁹ At this point, it gets evident, that digital curation finds application to both long-term preservation archives and safe repositories for continued use.

At the »Ingest«, the digital object and its metadata (the so called Submission Information Package, SIP) have to be checked according to standards and policies.

The Archival Information Package (AIP) – the »container« with the digital object and its »representation information« (metadata that will allow to represent the digital object adequately for later use, or to construct a suitable usage environment in distant future) – will be the subject of preservation actions (bit-stream preservation, migration, emulation).

Grid technology and the method of multiplied copies, realised by LOCKSS¹⁰ may be interesting storage options for repositories. It is not evident whether they are options for long-term preservation, too, because the high criteria for trusted repositories¹¹ seem to be easier to fulfill by centralised data archives.¹²

By now, there are experiences with the »Access, Use and Reuse« concerning repositories. Careful evaluation will be necessary to find out if »Access and (Re-)Use« are working well for »real« digital archival systems because the main focus of research by now laid more on the »Ingest« than on the »Access«. The »Sequential Action of Access« may make it necessary to transform the digital object, e.g. to migrate it into a publication format.

The credo of digital curation: only sustainable data and information management will allow reuse at any time. The curation of digital data or information is necessary independently of the chosen time scale: for continuous use as well as for long-term preservation.

But there are more reasons why digital curation is interesting for museums, museum documentation and museum informatics: the DCC has been founded because the creation of research data is extremely expensive and requires collaborative effort for maintaining these data usable. »The cost of producing these data is very high: satellites, particle accelerators, genome sequencing, and large scale digitisation and electronic publishing collectively represent a cumulative investment of billions of pounds [...]. Long-term curation and preservation of digital re-

archives ou détruites (si nécessaire de façon documentée, comme aux archives).

»Importer« signifie transférer l'objet d'information à un système d'archivage numérique ou – ce qui est différent de la préservation numérique – à un autre »référentiel, centre de données ou autre dépositaire«⁹. A ce stade, il devient évident que la conservation numérique s'applique à la fois aux archives de préservation à long terme et aux référentiels pour l'usage courant.

Au moment de l'importation, l'objet numérique et ses métadonnées (pour ainsi dire le paquet d'Information Subordonnées, SIP) doivent être vérifiés selon les normes et les politiques.

Le Paquet d'Information Archivé (AIP) – le »contenant« de l'objet numérique et de son »information de représentation« (les métadonnées qui permettront de représenter l'objet numérique de façon adéquate à une utilisation future ou à construire un environnement adapté à son usage dans un future distant) – sera l'objet d'actions de préservation (préservation de flux binaires, migration, émulation).

La technologie de Grille et la méthode des copies multiples, réalisée par LOCKSS¹⁰ pourraient être des options de stockage intéressantes pour les référentiels. Il n'est pas évident à dire si elles sont également des options pour la préservation à long terme, parce que les exigences élevées pour les référentiels¹¹ de confiance semblent être plus facile à remplir par des archives de données centralisées.¹²

A l'heure actuelle, il existe des expériences avec »Accès, Utilisation et Réutilisation« concernant les référentiels. Une évaluation attentive sera nécessaire pour déterminer si l'Accès et la (Ré-)Utilisation fonctionnent bien pour les »vrais« systèmes d'archivage numériques, parce que le principal sujet de recherche à présent a porté plus sur l'Importation que sur l'Accès. L'Action Séquentielle d'Accès pourrait rendre nécessaire la transformation de l'objet numérique, par exemple de le migrer à un format de publication.

Le credo de la conservation numérique: seule la gestion durable des données et informations permettra la réutilisation à tout moment. La conservation de données ou d'informations numériques est nécessaire, indépendamment de l'échelle de temps, pour une utilisation continue aussi bien que pour la préservation à long terme.

Mais il y a encore d'autres raisons pourquoi la conservation numérique est intéressante pour les musées, la documentation et l'informatique muséale: le DCC a été fondé parce que la création de données de recherche est extrêmement coûteuse et nécessite un effort de collaboration pour maintenir ces données en état d'usage. »Le coût de production de ces données est très élevé: satellites, accélérateurs de particules, séquençage de génome, numérisation à grande échelle et publications électroniques repré-

⁹ <http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf>

¹⁰ Lots of Copies Keep Stuff Save. <http://www.lockss.org/lockss/Home>

¹¹ See, for example, the criteria of the German NESTOR community. http://www.langzeitarchivierung.de/downloads/mat/nestor_mat_08-eng.pdf

¹² Like, for example, in the KOPAL system used by the German National Library. <http://kopal.langzeitarchivierung.de/index.php.de>

⁹ <http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf>

¹⁰ Lots of Copies Keep Stuff Save. <http://www.lockss.org/lockss/Home>

¹¹ Voir par exemple les critères de la communauté NESTOR Allemande. http://www.langzeitarchivierung.de/downloads/mat/nestor_mat_08-eng.pdf

¹² Comme par exemple dans le système KOPAL. <http://kopal.langzeitarchivierung.de/index.php.de>

sources is seen as a challenge which is difficult if not impossible for individual institutions to resolve on their own due to the complexity and scale of the challenges involved.¹³

This plea for collaborative solutions is also interesting regarding further integration of information from cultural heritage institutions but also regarding the integration of research and documentation data within museums (e.g. data on biodiversity research in natural history museums).

It is the digital information object that is at the focus of digital curation. But in view of »virtual archives«, it will be interesting to observe the use of the Records Continuum Model,¹⁴ published in 1996 by Frank Upward and widely discussed in the (Anglo-American) archival community.

The final conclusion may be allowed: Digital curation as a means to make museum information »fit for future« throughout continued use to long-term preservation should be adopted as an aim of museum documentation. Consequently, the qualification profiles of registrars and information specialists will have to be revised regarding the profile of a digital curator in museums. Further research will have to be done in the application of digital curation to museum documentation.

sentent collectivement un investissement cumulatif de milliards de livres [...]. La conservation à long terme des ressources numériques est perçue comme un défi difficile sinon impossible à résoudre seules pour des institutions individuelles, en raison de la complexité et de l'échelle des défis impliqués.¹³

Ce plaidoyer pour des solutions de collaboration est non seulement intéressant en ce qui concerne la poursuite de l'intégration des informations provenant d'institutions du patrimoine culturel, mais aussi en ce qui concerne l'intégration des données de recherche et de documentation dans les musées (par exemples des données sur la recherche sur la biodiversité dans les musées d'histoire naturelle).

C'est l'objet d'information numérique qui est au centre de la conservation numérique. Mais compte tenu des »archives virtuelles«, il sera intéressant d'observer l'utilisation au Records Continuum Model,¹⁴ publié en 1996 par Frank Upward et largement débattu dans la communauté d'archivage (anglo-américaine).

La conclusion finale est permise: La conservation numérique comme moyen rendre l'information des musées »adaptée au future«, de l'utilisation continue à la préservation à long terme, devrait être adoptée comme objectif de la documentation muséographique. Par conséquent, les profils de qualification des spécialistes de l'information et de l'enregistrement devra être révisée en tenant compte du profil d'un conservateur numérique pour musées. D'autres recherches devront être menées sur l'application de la conservation numérique pour la documentation muséographique.

The spell of ubiquitous knowledge: *Europeana*, a portal to European cultural and scientific knowledge

Jörn Sieglerschmidt¹⁵

Library Service Centre Baden-Württemberg
joern.sieglerschmidt@uni-konstanz.de

Qui scit ubi scientia habenti est proximus. [Who knows, where the knowledge is, is closest to the knowing.]¹⁶

Par ces instruments d'ubiquité, d'universalité et d'éternité, l'homme se sera rapproché de l'état de divinité,...¹⁷

Content has been and will always be made available on various Internet platforms over the years, increasing exponentially: text, sound, (still and moving) im-

Le charme de l'ubiquité des connaissances: *Europeana*, un portail au savoir scientifique et culturel européen

Jörn Sieglerschmidt¹⁵

Qui scit ubi scientia habenti est proximus. [Qui sait où le savoir se trouve, est plus proche de le posséder.]¹⁶

Par ces instruments d'ubiquité, d'universalité et d'éternité, l'homme se sera rapproché de l'état de divinité,...¹⁷

Du matériel a été et sera toujours rendu disponible sur diverses plateformes internet au fil des années, s'accroissant de façon exponentielle: texte, son, image (fixe et animée). Grâce à des fonds privés ou publiques, des projets de numérisation, parfois gigantesques, ont été accomplis pour quitter la galaxie

¹³ <http://www.dcc.ac.uk/about/what/>

¹⁴ Upward, Frank. Structuring the Records Continuum – Part One: Postcustodial principles and properties. <http://www.infotech.monash.edu.au/research/groups/rcrg/publications/recordscontinuum-fupp1.html>

¹⁵ Thanks to Susanne Schuster and Werner Schweibenz for their comments.

¹⁶ Otlet, Paul. *Traité de documentation*. Bruxelles : Mundaneum, 1934, p. 2

¹⁷ Otlet, *ibid.*, p. 431

¹³ <http://www.dcc.ac.uk/about/what/>

¹⁴ Upward, Frank. Structuring the Records Continuum – Part One: Postcustodial principles and properties. <http://www.infotech.monash.edu.au/research/groups/rcrg/publications/recordscontinuum-fupp1.html>

¹⁵ Merci à Susanne Schuster et Werner Schweibenz pour leurs commentaires.

¹⁶ Otlet, Paul. *Traité de documentation*. Bruxelles : Mundaneum, 1934, p. 2

¹⁷ Otlet, *ibid.*, p. 431

age. Owing to private or public funding sometimes gigantic digitisation projects were carried out, in order to leave the Gutenberg galaxy. One of the best known projects is *Google Books*. But there are many more initiatives in European countries – and world-wide. This contribution means to give a short survey of some of them, their intentions and their future.

The target of *Europeana* is »to make Europe's cultural and scientific resources accessible for all.« In detail the aims are:

- »Providing access to Europe's cultural and scientific heritage through a cross-domain portal,
- co-operating in the delivery and sustainability of the joint portal,
- stimulating initiatives to bring together existing digital content,
- supporting digitisation of Europe's cultural and scientific heritage.«¹⁸

These objectives are (and must be) rather general and vague. But at the same time it makes us aware of the fact that we are at the very beginning of a long process. Although *Europeana* has gone online in December 2008, to this day the access to cultural and scientific resources is fairly fragmentary. *Europeana* presents the material that is delivered on a voluntary basis by European cultural heritage institutions. Procedures of data delivery and import are far from regular and easy to handle. But the structure evolves and will be developed by aggregators of data in European and national projects. For example, ATHENA will be a domain specific aggregator of European museums.¹⁹ The *German Digital Library* will begin its work next year and will be a national aggregator for *Europeana*. In many European countries similar endeavours have been and will be undertaken:²⁰ Gallica,²¹ Culturaitalia,²² the German BAM portal,²³ and many others. So we may expect that in ten or twenty years *Europeana* will be the reliable platform of European digitised material for scholars, students, and the most unknown user: the general public.

The technical approach of *Europeana* is simple. *Europeana* itself administrates the metadata and images in a defined resolution forming the so-called surrogate for the presentation on the Web. The full text e.g. of books, the high resolution image, the moving image are accessible only through the source information of the institution that originally delivered the data. This approach has already been realised by the BAM portal. *Europeana* therefore presents material that is online in a more or less fully digitised form.

When in former times digitised content was offered to the public on different Internet platforms, sometimes it was rather concealed in the width of the World Wide Web than published in the traditional sense,

de Gutenberg. L'un des projets les plus connus est *Google Books*. Mais il y a encore beaucoup d'autres initiatives dans les pays européens – et à l'échelle mondiale. Cette contribution a pour but de donner un court aperçu de certaines d'entre elles, leurs intentions et leur future.

Le but de *Europeana* est »de rendre les ressources scientifiques et culturelles de l'Europe accessibles pour tous«. En détail, les objectifs sont:

- »Produire un accès au patrimoine culturel et scientifique de l'Europe à travers un portail inter-domaine,
- coopérer à la livraison et à la durabilité d'un portail commun,
- stimuler les initiatives pour réunir le matériel numérique existant,
- soutenir la numérisation du patrimoine culturel et scientifique.«¹⁸

Ces objectifs sont (et doivent être) assez généraux et vagues. Mais en même temps, ceci nous rend conscient du fait que nous n'en sommes qu'au début d'un très long processus. Bien qu'*Europeana* soit en ligne depuis décembre 2008, à ce jour l'accès aux ressources scientifiques et culturelles est plutôt fragmentaire. *Europeana* présente le matériel qui est livré volontairement par les institutions du patrimoine culturel européen. Les procédures de livraison et d'importation de donnée sont loin d'être faciles et habituelles à traiter. Mais la structure évolue et sera développée par des agrégateurs de données en des projets européens et nationaux. ATHENA par exemple, sera un agrégateur pour un domaine spécifique des musées Européens.¹⁹ La Bibliothèque Numérique Allemande commencera son travail l'année prochaine et sera un agrégateur national pour *Europeana*. Dans plusieurs pays européens de semblables projets ont été et seront entrepris:²⁰ Gallica,²¹ Culturaitalia,²² le portail BAM allemand,²³ et bien d'autres. Ainsi, nous nous attendons à ce que, dans dix ou vingt ans, *Europeana* soit la plateforme fiable de matériel numérisé pour universitaires, étudiants et la majorité des usagers inconnus: le grand public.

L'approche technique d'*Europeana* est simple: *Europeana* même administre les métadonnées et images dans une résolution définie, formant pour ainsi dire le substitut pour la présentation sur le Web. Le texte complet, par exemple de livres, des images haute définition, des images animées, sont accessibles seulement depuis l'information de source fournie par l'institution ayant livré les données à la base. Cette approche a déjà été adoptée par le portail BAM. *Europeana* présent ainsi du matériel qui est en ligne sous une forme plus ou moins numérisée.

Lorsqu'autrefois, le contenu numérique était offert au public sur des plateformes internet différentes, il était parfois plus dissimulé par l'ampleur du World

¹⁸ <http://www.europeana.eu/> [All URLs have been accessed April 5, 2009]

¹⁹ <http://www.athenaeurope.org/>

²⁰ cf. Culture&Recherche 118-119 presenting the papers of the conference Numérisation du patrimoine culturel in November, 2008. <http://www.culture.gouv.fr/culture/editions/r-cr/cr118-119.pdf>

²¹ <http://www.gallica2.bnf.fr>

²² <http://www.culturaitalia.it/pico/>

²³ <http://www.bam-portal.de>

¹⁸ <http://www.europeana.eu/>

¹⁹ <http://www.athenaeurope.org/>

²⁰ cf. Culture&Recherche 118-119. <http://www.culture.gouv.fr/culture/editions/r-cr/cr118-119.pdf>

²¹ <http://www.gallica2.bnf.fr>

²² <http://www.culturaitalia.it/pico/>

²³ <http://www.bam-portal.de>

and sometimes only known to a small number of selected users. In contrast, now the user has to access only one platform in order to get the information collected on it. Tools that are and will be developed by *Europeana* will enhance the usability of the presentation and provide an easy navigation through the mass of results. At the moment users can browse the results by facets like media type, language, country, date, and provider.

The promoters of digitisation – in the beginning mostly from the field of libraries – soon realised, that the world not only consists of manuscripts and books, in particular because other institutions such as archives and museums began to digitise their collections. Thus not only books were digitised, but all kinds of media and objects were transformed from an analogue to a digital representation. Media archives (print media, radio and television) produced (and are still producing) more texts and images than all traditional agents of cultural heritage together. Sciences and applied sciences, especially medicine, have been producing exabytes of digitised images and texts through the past decades. It is and will continue to be a challenge to find technical and semantical paths through this gigantic mass of information, particularly since projects on language recognition like e.g. the German *Theseus* programme²⁴ will multiply this mass. Retrieval tools are developed, too, by the French *Quaero* Programme.²⁵

In many European countries the projects were lacking a coherent strategy; others like UK, France, Norway, Sweden adopted a centralised approach. Although these efforts had effects on the consistency of results, the technical and professional variety of approaches was enormous. What technical measures are appropriate? High resolution images? Full text digitisation or image? What metadata schemata should be followed? Which metadata schemata are most appropriate to the various branches of documentation? Should ontologies like CIDOC CRM be used in order to harmonise different metadata schemata? How should the objects be presented on the screen? In an identical fashion without regard to special presentation models e.g. of the archives? Should the data have a defined quality, in order to meet the requirements of user needs and presentation? Should the full texts be also accessible by search engines of *Europeana* or *Google*? To date many of these questions have only been answered partially. The above mentioned projects will only offer some of the answers. But all endeavours are directed at the main aim: the networked knowledge and the vision that the linking of so different forms of human knowledge might lead to new insights, to new and better forms of knowledge. Whether the hopes of the American Society of Learned Societies in a paper on cyberinfrastructure published in 2005 – the betterment of all – can be realised, seems open to dispute, because the enlargement of technical possibilities too often has enlarged human foolishness – and the Internet is the best example for that. Nevertheless *Europeana* could be a reliable source of authentic material examined and offered by hundreds of thousands of European

Wide Web que publié dans le sens traditionnel, et parfois seulement connu d'un petit nombre d'utilisateurs sélectionnés. Pour contraster avec cela, l'utilisateur doit maintenant accéder à une seule plateforme pour obtenir l'information qui y est rassemblée. Les outils qui sont et seront développés par *Europeana* rendront ces prestations plus faciles à utiliser, produiront une navigation aisée à travers la masse de résultats. À l'heure qu'il est, les utilisateurs peuvent afficher les résultats par facettes, comme type de média, langue, pays, date et producteur.

Les promoteurs de la numérisation – initialement en grande partie du domaine des bibliothèques – réalisèrent bientôt que le monde ne consiste pas seulement en livres et manuscrits, en particulier parce que d'autres institutions telles que des archives et des musées commencèrent à numériser leurs collections. Aussi les livres n'étaient-ils pas les seuls à être numérisés, mais toutes sortes de médias et d'objets étaient transformés d'une représentation analogique à numérique. Les archives de médias (médias imprimés, radio et télévision) produisaient (et produisent toujours) plus de texte et d'images que tous les agents traditionnels de patrimoine culturel mis ensemble. Les sciences et les sciences appliquées, particulièrement la médecine, ont produit des exaoctets d'images et de textes numérisés durant la décennie passée. Trouver des chemins techniques et sémantiques à travers la gigantesque masse d'information est et continuera d'être un défi, particulièrement en vue de la multiplication de cette masse par des projets sur la reconnaissance de la langue, comme par exemple le programme allemand *Theseus*.²⁴ Des outils de récupération sont aussi développés par le programme français *Quaero*.²⁵

Dans beaucoup de pays européens il manquait aux projets une stratégie cohérente; D'autres, comme le Royaume Uni, la France, la Norvège, la Suède ont adopté une approche centralisée. Bien que ces efforts aient eu des effets sur la cohérence des résultats, la variété technique et professionnelle des approches était énorme. Quelles mesures techniques sont appropriées? Images haute résolution? Numérisation intégrale du texte ou de l'image? Quels schémas de métadonnées devraient être suivis? Quels schémas de métadonnées sont plus appropriés aux diverses branches de la documentation? Devrait-on utiliser des ontologies comme CIDOC-CRM afin d'harmoniser les différents schémas de métadonnées? Comment les objets devaient-ils être représentés à l'écran? De façon identique, sans tenir compte des modèles de présentation spéciaux, par exemple les archives? Les données devraient-elles avoir une qualité définie afin de répondre aux exigences des utilisateurs et de la présentation? Est-ce que les textes complets devraient également être accessibles par des moteurs de recherche de *Europeana* et *Google*? À ce jour, bon nombre de ces questions n'ont obtenu que des réponses partielles. Les projets mentionnés ci-dessus de proposent que quelques-unes des réponses. Mais tous les efforts sont dirigés vers l'objectif principal: la mise en réseau des connaissances et l'idée que l'établissement de liens entre

²⁴ <http://theseus-programm.de/en-US/home/default.aspx>

²⁵ http://www.quaero.org/modules/movie/scenes/home/index.php?FUSEBOX_LANG=2

²⁴ <http://theseus-programm.de/en-US/home/default.aspx>

²⁵ http://www.quaero.org/modules/movie/scenes/home/index.php?FUSEBOX_LANG=2



Paul Otlet in his office, 1934 / Paul Otlet dans son bureau, copyright Mundaneum, Bruxelles

institutions from all cultural and scientific heritage sectors.

Such a vision had Paul Otlet in 1934, when he wrote in his *Traité de documentation* about the future of the book: «...la table de travail ne serait plus chargée d'aucun livre. À leur place se dresse un écran et à portée un téléphone. Là-bas au loin, dans un édifice immense, sont tous les livres et tous les renseignements... De là, on fait apparaître sur l'écran la page à lire pour connaître la réponse aux questions posées par téléphone, avec ou sans fil. Un écran serait double, quadruple ou décuple s'il s'agissait de multiplier les textes et les documents à confronter simultanément; il y aurait un haut parleur si la vue devait être aidée par une donnée ouïe, si la vision devait être complétée par une audition. Utopie aujourd'hui, parce qu'elle n'existe encore nulle part, mais elle pourrait bien devenir la réalité pourvu que se perfectionnent encore nos méthodes et notre instrumentation. Et ce perfectionnement pourrait aller jusqu'à rendre automatique l'appel des documents sur l'écran, automatique aussi la projection consécutive...»²⁶

The cultural object and its image: some thoughts on frequent confusions

Regine Stein

German Documentation Centre for Art History
Bildarchiv Foto Marburg
r.stein@fotomarburg.de

The increasing Internet presence of cultural collections and especially emerging web portals that aim at resource discovery in a cross-collection and cross-sectoral (archives, libraries, museums,...) manner like *Europeana* reveal a well-known problem to documentation professionals: What is the subject of documentation, here, the subject of the metadata published? Is the cultural object itself described – which might be a visual art work, also a photograph – or the (digital) image as a visual representation of the object?

Metadata formats used in such portals are often based on the Dublin Core metadata standard which

²⁶ Otlet, *ibid.*, p. 428

de si différentes formes de la connaissance humaine pourraient conduire à de nouvelles visions, à de nouvelles et meilleures formes de connaissance. Que l'espoir de la Société Américaine des Sociétés Savantes, dans un article sur les cyber infrastructures publié en 2005 – l'amélioration de tout – soit ou non réalisable est ouverte au débat. En effet, l'élargissement des horizons techniques a trop souvent élargi la bêtise humaine, ce dont Internet est peut-être le meilleur exemple. Néanmoins, *Europeana* pourrait être une source fiable de matériel authentique examiné et offert par des centaines de milliers d'institution européennes de tous les secteurs du patrimoine culturel et scientifique.

Telle était la vision de Paul Otlet en 1934, lorsqu'il écrivit dans son *Traité de documentation* à propos du livre du futur: «...la table de travail ne serait plus chargée d'aucun livre. À leur place se dresse un écran et à portée un téléphone. Là-bas au loin, dans un édifice immense, sont tous les livres et tous les renseignements... De là, on fait apparaître sur l'écran la page à lire pour connaître la réponse aux questions posées par téléphone, avec ou sans fil. Un écran serait double, quadruple ou décuple s'il s'agissait de multiplier les textes et les documents à confronter simultanément; il y aurait un haut parleur si la vue devait être aidée par une donnée ouïe, si la vision devait être complétée par une audition. Utopie aujourd'hui, parce qu'elle n'existe encore nulle part, mais elle pourrait bien devenir la réalité pourvu que se perfectionnent encore nos méthodes et notre instrumentation. Et ce perfectionnement pourrait aller jusqu'à rendre automatique l'appel des documents sur l'écran, automatique aussi la projection consécutive...»²⁶

L'objet culturel et son image: quelques réflexions sur des confusions fréquentes

Regine Stein

La présence se fait croissante sur Internet de collections culturelles, et en particulier des nouveaux portails web, comme *Europeana*, visant la découverte de ressources tirées de différentes collections et secteurs (archives, bibliothèques, musées,...). Ce phénomène révèle un problème bien connu aux professionnels de la documentation: Quel est ici le sujet de la documentation, c'est à dire le sujet des métadonnées publiées? S'agit-il de la description de l'objet culturel même – qui pourrait être une œuvre d'art visuelle aussi bien qu'une photographie – ou est-ce l'image (numérique) comme représentation visuelle de l'objet qui est décrite?

Les formats de métadonnées utilisés dans de tels portails sont souvent basés sur les normes de métadonnées Dublin Core qui ont été conçues pour la

²⁶ Otlet, *ibid.*, p. 428

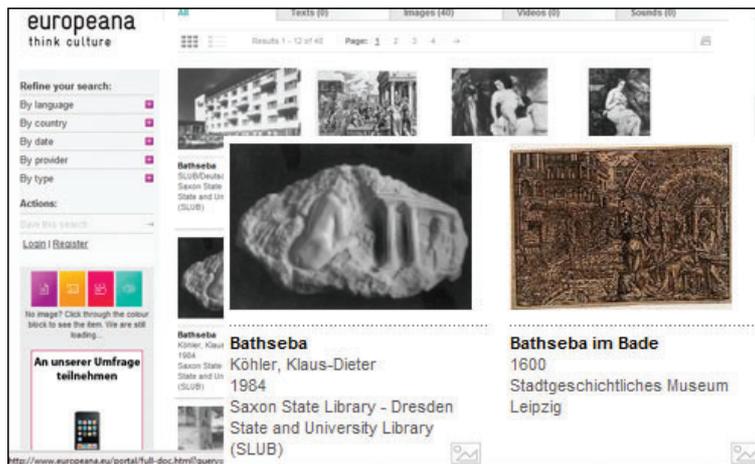


Figure 1. Extract of *Europeana* search result for »Bathseba« / Extrait de la recherche *Europeana* pour »Bathseba«

was designed for web resource discovery, and, for data from heritage collections, they lack a clear definition what kind of resource is described. For the *Europeana* prototype metadata format ESE, the issue is even more obscure: the proper *europeana:type* element focuses on presentation, not on retrieval, and provides only four types classifying digital resources available via Internet: text, image, video and sound. Thus most museum objects as well as architecture e.g. appear in the category 'image' as they are represented online by an image – but what is the metadata about?

An example from *Europeana*²⁷ might demonstrate the semantic confusion resulting from this approach (Figure 1): Searching for »Bathseba« returns an item from the Stadtgeschichtliches Museum Leipzig with *dc:type* 'Holzschnitt' (woodcarving), *dc:subject* 'Bathseba', *dc:date* '1600', etc. – the metadata describes the original museum object. There appear also a number of hits from the Saxon State Library, its metadata show mixed information: an example item has *dc:type* 'image', *dc:subject* 'Skulptur' (sculpture) – this refers to the visual representation, but *dc:date* '1984' is the production date of the cultural object, and *dc:creator* 'Köhler, Klaus-Dieter; SLUB/Deutsche Fotothek' provides creation information on both the cultural object itself and its visual representation. E.g. a *Europeana* »Advanced search« in the subject field for »Bathseba« returns only the woodcarving but not the sculpture as a result, even though in the original context the sculpture has the iconographic keyword »Bathseba«. The different notions of mapping object and image data hence produce unsatisfying retrieval results for data that could be merged very well, seen from the original context.

We have to consider another aspect: photographs, especially historical ones, may themselves be museum collection items. Originally representing the depicted cultural object only, they may become a proper source for research as the following example illustrates.

The oldest house at the market place in Trier, the »Steipe«, built in 1430, is subject of an image created by the »Königliche Preußische Messbildanstalt Berlin« in 1901. A positive, retouched for publication in

découverte de ressources web, et, lorsqu'utilisées pour les données provenant de collections du patrimoine, il leur manque une définition claire du genre de ressources décrites. Concernant ESE, le format pour les métadonnées, prototype d'*Europeana*, la question est encore plus obscure: l'élément *europeana:type* même se concentre sur la présentation, non sur la découverte, et ne fournit que quatre types de classification des ressources disponibles via Internet: texte, image, vidéo et son. Ainsi, la majorité des objets des musées, aussi bien que l'architecture, par exemple, apparaissent dans la catégorie »image« car ils sont représentés en ligne par une image – mais à quoi la métadonnée se rapporte-t-elle?

Un exemple tiré d'*Europeana*²⁷ peut démontrer la confusion sémantique résultant de cette approche (Figure 1): à la recherche du mot »Bathseba« on obtient un objet du Stadtgeschichtliches Museum Leipzig avec *dc:type* »Holzschnitt« (sculpture sur bois), *dc:sujet* »Bathseba«, *dc:date* »1600«, etc. – les métadonnées décrivent l'exemplaire original de l'objet du musée. Plusieurs résultats de la Saxon State Library apparaissent également, dont les métadonnées donnent des informations confuses: un élément utilisé comme exemple possède les données suivantes: *dc:type* »image«, *dc:sujet* »Skulptur« (sculpture) – ceci se réfère à la représentation visuelle – mais *dc:date* »1984« est la date de production de l'objet culturel, et *dc:créateur* »Köhler, Klaus-Dieter; SLUB/Deutsche Fotothek« fournit des informations de création à la fois sur l'objet même et sur sa représentation visuelle. Par exemple une »recherche avancée« avec *Europeana* sur le sujet »Bathseba« ne donne comme résultat que la sculpture sur bois mais pas la sculpture comme résultat, bien que dans le contexte original, la sculpture possède le mot clef iconographique »Bathseba«. Ainsi, les interprétations conflictuelles de la sémantique entre image et objet donnent des résultats insatisfaisants pour de l'information qui pourrait très bien être combinée, compte tenu du contexte original.

Il nous faut considérer un autre aspect: les photos, notamment historiques, sont peut-être en elles mêmes des articles de collection de musée. Initialement prévues pour la seule représentation de l'objet culturel

²⁷ Search <http://www.europeana.eu>

²⁷ <http://www.europeana.eu>

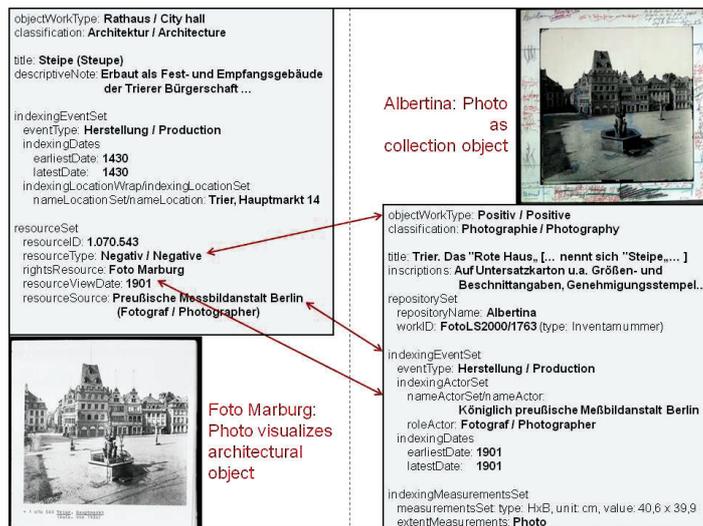


Figure 2. museumdat representation of the Foto Marburg and the Albertina object records / Représentation museumdat de la Foto Marburg et de l'enregistrement de l'objet Albertina

1915, is now part of the Albertina photo collection.²⁸ Its metadata record shows the same categories as for any other work of art: work type, creator, creation date, measurements, technique, and the visual content of the photograph appears in the title only. In contrast, the image archive »Foto Marburg«²⁹ describes the architectural object, the building itself, in the object record of its database, with information on a negative of this same image attached as a visual resource (Figure 2).

The mapping into museumdat³⁰ – a metadata format for encoding core data from museum holdings in a CIDOC CRM compliant way, largely based upon CDWA Lite³¹ and designed for contributing museum information, and specifically data and images relating to the object, to union catalogues and portals – shows the correspondent information in a semantically correct way: The Albertina collection object is classified as »photograph«, the documented object at Foto Marburg as »architecture«, the object/work type »positive« corresponds to the resource type »negative«, creator and creation date of the Albertina object record correspond to the resource's view date and resource's source of the Foto Marburg record. Both objects – the photograph and the architecture – could be related in the museumdat records via a related work set, containing identifier, document URL and the relationship type »is about / is subject of«.

In terms of the CIDOC CRM, the building »Steipe« is subject of an image which represents both information carriers, the positive of the Albertina collection and the negative held as a visual representation of the architectural object described at Foto Marburg (Figure 3).

Getting back to the initial question what information is available – and retrievable – in web portals, for the delivery of cultural object data into portals it is crucial to choose an appropriate metadata schema: it should

décrire, elles peuvent devenir une véritable source pour la recherche, comme le montrent les exemples suivants.

La plus vieille maison de la place du marché de Trier, la »Steipe«, construite en 1430, est le sujet d'une image créée par la »Königliche Preußische Messbildanstalt Berlin« en 1901. Un tirage, retouché pour la publication en 1915, fait maintenant partie de la collection photo Albertina.²⁸ Ses métadonnées montrent les mêmes catégories que n'importe quel autre œuvre d'art: type d'œuvre, créateur, date de création, dimensions, technique, le contenu visuel de la photo n'apparaît que dans le titre. En revanche, l'archive de l'image »Foto Marburg«²⁹ décrit l'objet architectural, le bâtiment même, dans l'enregistrement de l'objet dans sa base de donnée, avec des informations sur un négatif de la même image attaché comme ressource visuelle (Figure 2).

Museumdat³⁰ est un format de métadonnées pour l'encodage des données de base sur les collections d'un musée de façon conforme à CIDOC CRM. Ce format est en grande partie basé sur CDWA Lite³¹ et conçu pour que les musées puissent contribuer des informations, et spécifiquement des données et des images se rapportant à l'objet, à des catalogues mutualisés et des portails. La représentation dans ce format montre l'information correspondante de façon sémantiquement correcte: L'objet de la collection Albertina est classifié comme »photographie«, l'objet documenté à Foto Marburg comme »architecture«, l'objet/type d'œuvre »positive« (tirage) correspond au type de ressource »négative«, le créateur et la date de création de l'objet Albertina correspondent à la date de consultation de la ressource et à la source des enregistrements Foto Marburg. Les deux objets – la photo et le monument – pourraient être reliés dans l'enregistrement museumdat via un ensemble de travail partagé, contenant un identifiant, un document URL et le type de relation »est à propos de / est le sujet de«.

²⁸ See <http://gallery.albertina.at/eMuseum>

²⁹ See <http://www.bildindex.de/obj20243976.html>

³⁰ <http://www.museumdat.org/index.php?ln=en&t=home>

³¹ http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite.html

²⁸ <http://gallery.albertina.at/eMuseum>

²⁹ <http://www.bildindex.de/obj20243976.html>

³⁰ <http://www.museumdat.org/index.php?ln=en&t=home>

³¹ http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite.html

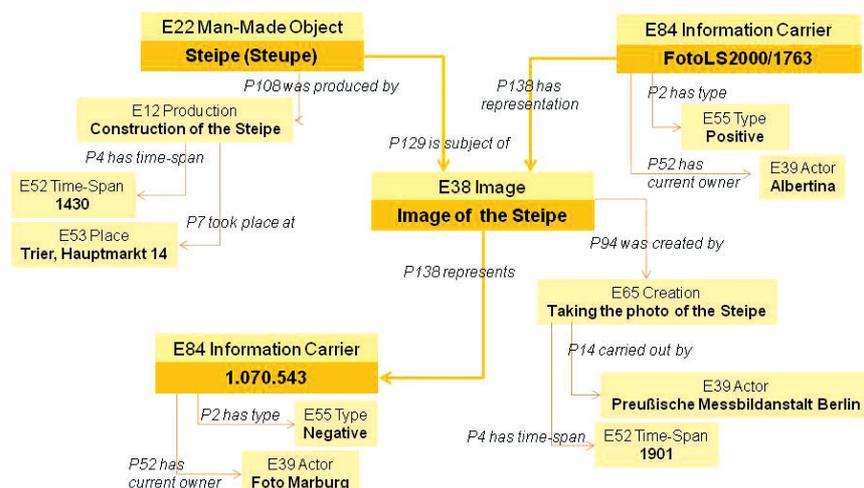


Figure 3. CIDOC-CRM graph for the »Steipe« and its image / Diagramme CIDOC-CRM pour le »Steipe« et son image

be possible to express the richness of the original data if the provider wishes to deliver more than basic information. Current developments to harmonise CDWA Lite and museumdat into one single, internationalised schema reflect this need and will be continued within the CIDOC.

En termes de CIDOC CRM, le bâtiment »Steipe« est le sujet d'une image représentée dans les deux porteurs d'information, le tirage de la collection Albertina et le négatif considéré en tant que représentation visuelle de l'objet architectural décrit à Foto Marburg (Figure 3).

Pour en revenir à la question initiale, celle de savoir quelle information est disponible – et accessible – sur les portails web, remarquons qu'il est crucial de choisir un schéma de métadonnées approprié lors de la fourniture de données d'objets culturels aux portails: il devrait être possible d'exprimer la richesse des données d'origine si le fournisseur désire donner plus que les informations basiques. Ce besoin est reflété par des développements actuels pour harmoniser CDWA Lite et museumdat en un seul schéma international, qui seront continués au sein de CIDOC.

Challenges in museum documentation for transdisciplinary information integration: From field names to events

Karl-Heinz Lampe and Siegfried Krause
 Chair and co-chair of CIDOC Working Group on Transdisciplinary Approaches in Documentation
 k.lampe.zfmk@uni-bonn.de
 s.krause@gnm.de

Museums are memory institutions. They provide information on objects for visitors, scholars and scientists. In particular researchers expect easy access to detailed information on museum objects by using modern information technology. Information should be directly available to answer questions in various research scenarios. Today a major challenge in museum documentation is to connect cultural heritage information from different domains ranging from archaeology to zoology.

The traditional practice of documentation in museums shares with analogous practices in archives and libraries a focus on classes such as locality, collector, or author. A trend is now underway across scholarly and scientific domains in which the focus of documentation is expanding to include processes and

Défis dans la documentation des musées pour l'intégration d'informations transdisciplinaires: Des noms de champs aux événements

Karl-Heinz Lampe et Siegfried Krause

Les musées sont des institutions de mémoire. Ils fournissent de l'information sur des objets pour les visiteurs, les universitaires et les scientifiques. Les chercheurs s'attendent en particulier à un accès facile aux informations détaillées sur les objets des musées en utilisant les technologies modernes de l'information. L'information devrait être directement disponible pour répondre aux questions de divers scénarios de recherche. Aujourd'hui, un des plus importants défis dans la documentation des musées est de mettre en lien l'information des patrimoines culturels de différents domaines allant de l'archéologie à la zoologie.

La pratique traditionnelle de la documentation dans les musées partage avec des pratiques analogues, dans l'archivage ou les bibliothèques, une attention portée à des catégories telles que la localité, le collectionneur, l'auteur. Une nouvelle tendance se dessine aujourd'hui dans les domaines universitaires et scientifiques, où l'attention de la documentation

events. Thus, the entities what, who, where, and when are now being related to each other through processes and events.

Rather than documenting information just on the basis of classes entitled with simple field names, that are used in traditional card indexes, information is modelled here event-centric with an ontology based on the CIDOC Conceptual Reference Model (CIDOC CRM; ISO 21127). In informatics, ontology is a formal specification of semantic concepts. The CIDOC CRM is a compact object-oriented model consisting of 86 named classes or entities (E1, E2, etc.), which are interlinked by 137 named properties (P1, P2, etc.). The properties can be understood as verbalised links pointing to other defined entities of the CIDOC CRM, thus allowing the creation of simple »Subject Predicate Object« relations. Therefore these classes and properties provide a formalised language – a lingua franca – to describe common high-level semantics. One example should elucidate the principle of this event-centric kind of information integration: An interlinked and semantically enriched documentation of a bedbug which is firstly referred to by an engraving in cultural history museums and is secondly part of a biological collection object in natural history museums (Figure 1).

In this example each activity (event) is carried out by a particular person at a given place within a certain time-span. The example of a biological collecting event with the resulting biological collection object is shown in green while the production of a collection object in form of a copper engraving is depicted in red. For clarification both production events are documented in a simplified manner, because the real production process consists of several production steps. For biological collection objects collecting-, preparation-, determination- & documentation events are carried out by a collector, preparator/taxidermist, taxonomist and a curator. For engravings drawing-, engraving-, printing- & documentation events are carried out by an artist, an engraver, a printer and a publisher/curator. A semantically enriched documentation is the key to connect various information sources from natural and cultural history and other

s'étend jusqu'à inclure des processus et des événements. Ainsi, les entités quoi, qui, où et quand sont en relation les unes avec les autres au travers des processus et des événements.

Plutôt que de documenter l'information sur la seule base de catégories réduites à de simples noms de champs, utilisées dans les indexes à cartes traditionnels, l'information est modelée ici, au centre même des événements, avec une ontologie basée sur le CIDOC-Conceptual Reference Model (CIDOC-CRM; ISO 21127). En informatique, l'ontologie est une spécification formelle de concepts sémantiques. Le CIDOC-CRM est un modèle compacte, orienté vers les objets, qui consiste de 86 classes ou entités nommées (E1, E2,...), qui sont liées entre elles par 137 propriétés nommées (P1, P2,...). Ces propriétés peuvent être comprises comme des liens verbalisés pointant vers d'autres entités définies du CIDOC-CRM, permettant ainsi la création de simples relations »Sujet Prédicat Objet«. C'est pourquoi ces classes et propriétés produisent un langage formalisé – une Lingua franca – pour décrire une sémantique commune de haut niveau. Un exemple devrait éclairer le principe de cet intégration de l'information centrée sur des événements: la documentation d'une punaise qui figure d'abord dans des gravures dans les musées d'histoire culturelle, et qui fait partie d'une collection biologique dans un musée d'histoire naturelle (Figure 1).

Dans cet exemple, chaque activité (événement) est assurée par une personne particulière à un endroit donné, durant un certain laps de temps. L'exemple de l'évènement de faire une collection biologique avec l'objet de collection biologique résultant est montré en vert tandis que la production d'un objet de collection sous la forme d'une inscription sur cuivre est dépeinte en rouge. Pour plus de clarté, les deux événements de production sont documentés de façon simplifiée, parce que le processus de production réel consiste en plusieurs étapes de production. Pour les objets de collections biologiques, la préparation, la détermination et la documentation sont assurés par un collectionneur, préparateur/taxidermiste, taxonomiste et un conservateur. Pour les gravures, le dessin la gravure, l'impression et la documentation sont assurées par un

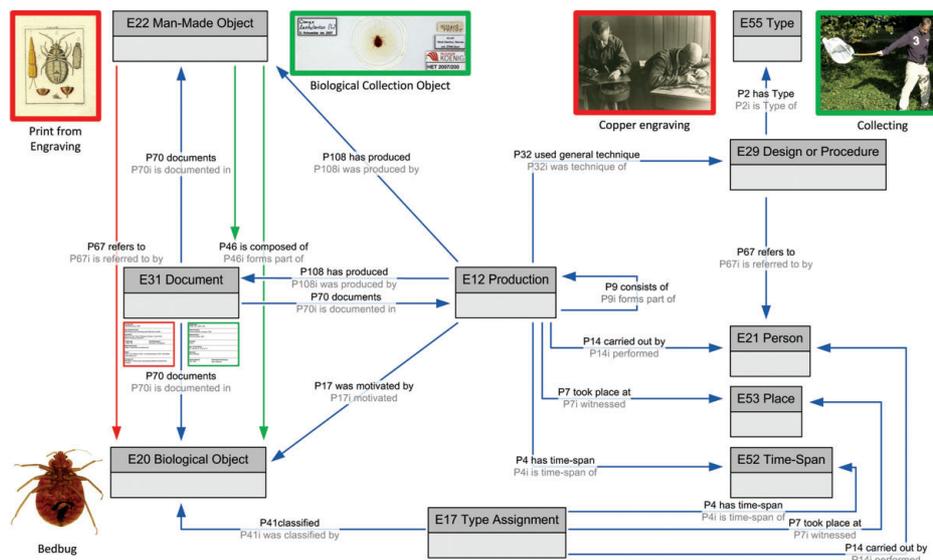


Figure 1. Semantic relationship between the production event of a biological collection object and the creation of a copper engraving in terms of the CIDOC CRM / Relation sémantique entre l'évènement de production d'un objet de collection biologique et la création d'une gravure sur cuivre, exprimé en termes du CIDOC-CRM

domains. It allows overarching queries such as: »Are there any proofs about the occurrence of species at a given place within a certain time-span in various scientific- and scholarly information sources?« Potential information sources in this context range from prehistoric cave paintings, Bronze Age ceramics, medieval drawings, paintings and hand-written documents and even cookery books through to digital images taken today or biological voucher specimens collected yesterday.

Through a formal ontology, information handling and dissemination can be improved by semantic enrichment. The formal specification of semantic concepts makes scientific activities understandable to a wider audience. Finally, in scientific as well as scholarly disciplines a formal ontology can give simple object documentation its full scientific depth beyond the limited purposes of collection management.

To document processes and events seems to be a prerequisite for transdisciplinary information integration, which is needed for developing knowledge networks and knowledge representation tools on the Internet.

artiste, un graveur, un imprimeur et un éditeur/conservateur. Une documentation sémantiquement enrichie est la clef qui permet de connecter différentes sources d'information provenant de l'histoire culturelle et naturelle et d'autres domaines. Elle permet d'englober des questions telles que: »Y a-t-il des preuves de la présence d'espèces à une période donnée, durant un certain laps de temps, dans diverses sources d'informations savantes et scientifiques?« Les sources d'information potentielles dans ce contexte vont des peintures murales préhistoriques, céramiques de l'âge du Bronze, dessins, peintures ou écrits médiévaux et même livres de cuisine jusqu'au images digitales prises aujourd'hui ou spécimen biologiques collectés hier.

Grâce à une ontologie formelle, le traitement et la diffusion de l'information peuvent être améliorés par un enrichissement sémantique. La spécification formelle de concepts sémantiques rend les activités scientifiques compréhensibles à une audience plus large. Finalement, dans les disciplines scientifiques aussi bien qu'académiques, une ontologie formelle peut donner à la simple documentation d'objets toute sa profondeur scientifique, au delà des fins limitées de la gestion de collections.

La documentation de processus et d'événements semble être un pré requis à l'intégration d'informations transdisciplinaires, elle-même nécessaire au développement de réseaux de savoir et d'outils de représentation du savoir sur Internet.

The CIDOC Co-reference Working Group

Øyvind Eide
University of Oslo
oyvind.eide@iln.uio.no

In order to make more usable information systems for the wider cultural heritage field, it is necessary to connect persons, places, and objects described in and owned by museums. Such information should also be connected to references in e.g. digital versions of printed books. The CIDOC Co-reference Working Group was established in 2007. The main focus of the group is to assist in the further development of museum documentation through promoting work in the area of co-reference.

An example of a reference is the string »The table by the window« referring to a physical object. If another string »The beautiful table« is referring to the same physical object, the two strings are said to co-refer.

Tracking down co-references has always been one of the practical tasks performed by researchers, conservators, librarians and others processing information about real world objects described in texts. The results have sometimes been published, e.g. in indices, or in footnotes. But most often they have been saved in the finder's head, maybe also in his notes.

Co-reference management should be part of the overall information strategy in all institutions. The curator, researcher or exhibition professional should be able to add information to their local system about internal co-reference as well as external resources co-referring with internal resources. This should then be available in a format readable for computers as well as humans.

Le Groupe de Travail CIDOC sur le Co-référencement

Øyvind Eide

Afin de rendre plus utilisables les systèmes d'information pour le plus large domaine du patrimoine culturel, il est nécessaire de mettre en lien les personnes, les lieux et les objets décrits par et appartenant aux musées. De telles informations devraient aussi être mises en lien avec des références, par exemple dans des versions numériques de livres imprimés. Le Groupe de Travail Coréférentiel de CIDOC a été créé en 2007. Le principal objectif du groupe est d'aider à poursuivre le développement de la documentation muséale par le biais de la promotion de travaux dans le domaine de la coréférence.

Un exemple de référence est la chaîne "La table auprès de la fenêtre", en référence à un objet physique. Si une autre chaîne, "La belle table", se réfère au même objet physique, les deux chaînes sont dites coréférentielles.

Traquer les coréférences a toujours été l'une des tâches pratiques effectuées par les chercheurs, les conservateurs, les bibliothécaires, et d'autres, traitant l'information à propos d'objets du monde réel décrits dans les textes. Les résultats ont parfois été publiés, par exemple dans les indices ou les notes de bas de page. Mais la plupart du temps, il a été enregistré dans la tête du chercheur, peut-être aussi dans ses notes.

La gestion de co-références devrait être faire partie de la stratégie global d'information dans toutes les institutions. Le conservateur, chercheur ou autre professionnel de l'exposition, devrait être à même d'ajouter de l'information au système local, à propos de co-références internes aussi bien que de ressources externes en

An example of co-reference information: *The catalogue in museum M. states that several artefacts are found by someone called Tom Smith. The curators know that this name refers to two different persons, one of them being the grandfather of the other. As the catalogue is a good one, each artefact record has a connection to the object in the system representing the person who found it. So for some of the objects the finders are co-referred to Tom Smith the elder, and for others to Tom Smith the younger. Further, the grandfather was also an author of several books. In the library catalogue of the museum, the author information for these books is co-referred to Tom Smith the elder. The records in the museum catalogue and the ones in the library catalogue are co-referred to each other as well, so we know which of the artefacts were found by the author and which were found by his grandson. Thus, the system gives us a correct and complete answer to the question: »Which artefacts were found by the author of this specific book?«*

Several institutions are now developing systems for co-reference management in a museum environment, while other groups are working on the theoretical side of the problem. Grant proposals are also written for projects in which co-reference resolution and the use of sets of co-reference facts may be utilised. This will then be a part of a larger network of information curated by memory institutions.

This network of information will depend on the fact that co-references are documented not only within an institution, but also with reference to other institutions, museums as well as other memory institutions.

Tom Smith the younger was an active photographer. His niece, Sarah Smith, has digitised many of his pictures. She also made a database on the web consisting of her uncle's pictures together with old pictures taken by other members of her family. She is eager to make good resources for the users, so she added a link from her system to the system of museum M. This means that the information about her uncle is connected to the information in the museum system. A similar connection exists for Tom Smith the elder, because he is the motif of some of the pictures. She has also convinced museum M. to include references to her database in their public catalogue, so users can now find images taken by and of the two Tom Smith's connected to internal museum information by searching the museum's online catalogue.

Co-reference will not in itself solve all the problems in museum documentation. The relationships between the persons in the examples, e.g. that they were relatives, have to be stored and made accessible using other techniques. But co-reference should be an integrated part of the information systems of museums in the future.

For more information,
contact Øyvind Eide at
oyvind.eide@iln.uio.no
or Mika Nyman at
mika.nyman@synapse-computing.com.

co-référence avec des ressources internes. Ceci devrait ensuite être disponible dans un format lisible tant par les ordinateurs que les humains.

Un exemple d'information coréférentielle: *Le catalogue du musée M. affirme que plusieurs objets ont été trouvés par un certain Tom Smith. Les conservateurs savent que ce nom se réfère à deux personnes différentes, l'un étant le grand-père de l'autre. Comme le catalogue est bien fait, chaque enregistrement d'objet d'art est en lien avec l'objet dans le système représentant la personne qui l'a trouvé. Ainsi, pour certains objets, les découvreurs sont en coréférence à Tom Smith l'aîné, et pour d'autres à Tom Smith le jeune. En outre, le grand-père était également l'auteur de plusieurs livres. Dans le catalogue de la bibliothèque disponible au musée, l'information sur l'auteur de ces livres se co-réfère à Tom Smith l'aîné. Les dossiers dans le catalogue du musée et ceux dans le catalogue de la bibliothèque sont aussi en coréférence l'un avec l'autre, ce qui nous permet de savoir quels objets ont été trouvés par l'auteur et lesquels ont été trouvés par son petit fils. Ainsi, le système nous donne une réponse correcte et exhaustive à la question: "Quels objets ont été trouvés par l'auteur de ce livre-ci?"*

Plusieurs institutions développent à présent des systèmes pour la gestion coréférentielle dans l'environnement des musées, tandis que d'autres groupes travaillent sur l'aspect théorique du problème. Des propositions de bourses sont aussi faites pour des projets dans lesquels la résolution des coréférences et l'utilisation d'ensembles de faits coréférentiels pourraient être utilisés. Ceci fera ensuite partie d'un plus grand réseau d'information conservé par des institutions de mémoire.

Ce réseau d'information dépendra du fait que les coréférences soient documentées non seulement au sein d'une institution, mais aussi en référence à d'autres institutions, des musées aussi bien que d'autres institutions de mémoire.

Tom Smith le jeune était un photographe actif. Sa nièce, Sarah Smith, a numérisé un grand nombre de ses photos. Elle a également confectionné une base de données sur le web constituée des photos de son oncle avec de vieilles photos prises par d'autres membres de sa famille. Elle aimerait beaucoup faire de bonnes ressources pour les usagers. C'est pourquoi elle a ajouté un lien de son système au système du musée M. Ceci signifie que l'information à propos de son oncle est connectée à l'information dans le système du musée. Une connexion similaire existe pour Tom Smith l'aîné, parce qu'il est le sujet de certaines des photos. Elle a également convaincu le musée M. d'inclure des références à sa base de donnée dans leur catalogue public, afin que les usagers puissent maintenant trouver des images par et de chacun des Tom Smith connectées à l'information muséale interne, en parcourant le catalogue en ligne du musée.

La coréférence ne résoudra pas en elle même tous les problèmes de la documentation muséale. Les relations entre personnes, dans le cas par exemple d'un lien de parenté, doivent être stockées et rendues accessibles grâce à d'autres techniques. Mais la coréférence devrait faire partie intégrante des systèmes d'information des musées dans le future.

Pour plus d'information,
contacter Øyvind Eide à
oyvind.eide@iln.uio.no
ou Mika Nyman à
mika.nyman@synapse-computing.com.