Creation Driven Marketing
A vision paper

Joost KIRCZ

Institute for Media and Information Management
Hogeschool van Amsterdam, Nederland
j.g.kircz@hva.nl

Abstract
Coding
The unique force of electronic products is in the coding. On the one hand we have the electronic object as such. It can be a plain text document, a photo, a full colour flyer, a video, a software program, a game, or even a PDF version of an old-fashioned book. The electronic object contains lots of coding. At present this coding mainly describes the lay-out and structure of the electronic file.
Metadata
On the other hand, just as in traditional the marketing & sales, every product has a name and a product description. These descriptions are codes, mostly clearly separated from the item itself, describing the item. So-called metadata.
When the electronic object has to find its way to a consumer and becomes a product in a commercial chain, normally, and very traditionally, only the metadata are used. The coding, added in the creation of the electronic product, is hardly used.
Vision
The vision defended in this paper is that the creator of an electronic product has an essential capability in positioning the product in the market by clever integration of metadata in the creation process itself. The paper starts with a description of the landscape, followed by an analysis of the notion metadata. Subsequently the production of metadata and the role of the creator is discussed. The paper concludes that by cleverly coding the semantic information during the creation process, the creator will be able to play a much larger role in targeting the ultimate consumer market now and in the future.

Curriculum vitae
Joost Kircz started studying chemistry and finished in molecular physics. His quest for more knowledge drove him into science publishing, which allowed him to peek into even a wider range of subjects. For more than a decade he was publisher for the internationally renowned North-Holland Physics programme of Elsevier Science. Unable to clean his desk, he became interested in electronic storage and publishing of information. As from 1987 he is engaged in research in that field and as from 1992 he is visiting scientist at the University of Amsterdam. Curiosity drove him out of international management tasks in order to start his own publishing research company in 1998. As from 2006 he is lector electronic publishing at MIM. For more information please visit www.kra.nl

Introduction
What is a publisher?
  Before we start talking about electronic publishing and marketing we must have a clear view on what we call publishing. Certainly in our present day world where everybody claims to be a publisher, a librarian, a database provider and a host or portal facilitator at the same time, it is important to give clear meaning to seemingly obvious terms.
  Publishing, in my definition, is the united action of a number of functions that together enable the creation, production, marketing and dissemination of a product.

The product range to information objects is wide and consists of text, sound, pictures and film. In other words: magazines, journals, books, videos, films, etc. To make things even more simple, these information objects can be novel creations but also parts of or combinations of existing items that belong to an already existing collection such as a sculpture collection of an art gallery. It is important to stress that we do not deal with the carrier of the information, be it stone, clay, punch cards, parchment, paper or a blue ray DVD.

The basic raw materials are the knowledge and emotions of (or in, we don't know) the human brain. They are explicated in language, sounds, pictures or gestures and create information. Information can be described or better denoted. Information can be handled, packaged, counted, piled and stored away into a carrier.

The publisher's function

The publisher's role is to find, identify and collect interesting knowledge and emotions and subsequently have them expressed into information in writing, depicting or performing. Here and in the following I use the term publisher in a generic way. It can also be a film producer. The publisher is the organising pivot around which all players -authors, performers, technicians, editors as well as production and marketing & sales staff- circulate. Firstly the publisher has to define the final result, the product, then it has to be certified as being original or unique for the creator as well as validated by endowing it with a quality stamp of some level. The information can be a treatise on lunar research in a top ranked scientific journal or on loony tunes on the moon for pre-adult entertaining. The traditional publisher's role entails subsequently the organisation, distribution, marketing and sales of the product. Please be aware that within this definition we still can talk about a great variety of products, from poetry, via stock market graphics to games. In its drive to find a consumers' market, efforts are made to describe the product in such away that it dovetails with desires, wants or needs from perceived customers. Often this is called product driven, but subsequently it will become clear that I don't adhere to that prerogative.

The library function

On the other end of the line we have the library function. Here, I also mean the function and not the organisation or the building. The library function fulfills clear roles in a local or domain dependant field. It is the finding, selecting and collecting of information from a great variety of sources, fit for a well-defined user group. This user group may be small, but in the case of a public lending library also very large. User groups are defined by both what might interest them, as well as by what goes beyond the span of attention. Collecting is not only a positive act but also the negative act of rejection. In order to cater for the user group the librarian must classify its holdings according to an index or classification scheme that is explicitly tailored for (the context) of the user-group. We can then use the term library also in a generic way, describing the place where consumers find pre-selected products.

In a very blunt way we can equate the publisher with an industrial manufacturer exploiting the brain power instead of the physical power of contracted people. The librarian can then be depicted as a shopkeeper with a well-defined supply. In the product chain the producer-publisher adds product information and subsequently the librarian-shopkeeper adds specified consumers/customers information targeted for the own local market. In the most primitive way a shopkeeper is doing so already, by putting the same type of products such as, games, thrillers, or socks together on the same shelf.

The issue now is, in an electronic environment, to what extent this whole process from creator to consumer is fully determined by the publisher and librarian. In the following we will deal with the question of how the creator him/herself can exercise influence on this process and to what extent the new technologies enable as well as force changing patterns of long term usage.

Metadata, the new grail?

Metadata can be defined as data about data. Instead of a product description such "*An ideal peaceful, family game for under the Xmas tree*", which suggests that the game is solely targeted for a special occasion, we can think of more general descriptions. The idea of metadata is to create more-or-less logical systems that describe the product as well as that it enables manipulation on the meta level without having to "taste" or "smell" all individual underlying items. In the world of metadata we have to make a clear distinction between the different contexts in which metadata are relevant and useful. Let's make a first list of possible different metadata schemas to understand better where the various players in the value and product chains play a role.

Types of metadata
1- *Backbone.* We start with the data describing the technical and physical characteristics of the original. This entails issues like e.g., WordPerfect or MSWord text file, PDF, etc., length in bytes, ASCII or Unicode, handwritten manuscript on parchment, using video standard X, PNG or Tiff, etc. These data are important to transcribe the original in a production environment that allows storing into, as well as rendering the information into certain media. It is obvious that if we want cross media capabilities, these data are crucial. A Unicode file is nor yet accepted by all environments, though if the creator wants longevity it is better to create in Unicode rather than ASCII. A hand written text can be scanned. It is the quality of the scanning that defines to what extent the digital file is only representing the text or also the texture of the paper. The digital preservation discussions and arguments pertain to this category of metadata.
2- *Structure.* We need metadata that describe the structure of the information. This can be lay-out structures in the case of text, or for video: structural information on spatio-temporal components of the content such as scene cuts, segmentation in regions, region motion tracking, etc.
3- *Content.* We have the content descriptors or traditional index terms, so-called keywords. Despite the pipedreams of superior success by single word manipulation in the present first generation search engines, more and more the need for context dependent thesauri is understood. A well-developed thesaurus with a clear formal structure between the terms is now called an ontology. As an aside, it is important to note that in the ongoing ontology research it remains very difficult to deal with non-hierarchical relations, such as the traditional "*see also*" relation.
4- *Administrative.* We have the data that describe the administrative, legal and personalized items. Here we have to think of the name, address, etc. of the creator and maybe also the editor, the performer in case of music, the distributor, the rights owner and as with film, the credit titles. Also in this box we have the data that describe the actual legal rights for reproduction and use.
5- *Post-production.* The metadata that are added after the creation. This can be the number of downloads, sales, but also -and very important- data that link the product at issue to other products. Note that here we also deal with references to new versions and updates of the same product. These products can be older or newer. This last type of metadata introduces a whole new field of indexing. In some on-line shops, like Amazon, we already see this type of metadata, created on the fly, if we select a book. Immediate look-alike books are suggested, in some case also linked to the books you selected on a previous visit. The important open question is to what extent we have to keep these data and make it part of the product. In the same way as on the back flap of a book reprint comments from reviewers are printed. This is typically an editorial and marketing question. By the usage of the product, new information is being produced which is essential for the understanding and the decision for new use by a new consumer. Think about the instructions leaflet for drugs. They have a latest date of change, but in an electronic environment, updating pertinent useful data and metadata becomes possible and will become compulsory in this case.
6- *Super-metadata*. Finally, after having invented all those almost independent but interrelated metadata schemas, we become badly in need for a metadata schema that enables the manipulation of all that product information.

Transparency
From the above it will be clear that we need a transparent metadata language to allow talking with and about metadata. At present the XML family serves that goal excellently. This presentation it is not the place nor the moment to express my obvious love for XML, but be aware, as soon as you are dealing with structured marketing using metadata, XML is immediately around the corner.
With all these descriptors, indicators and measures we are confronted with an interesting phenomenon, unique for an electronic environment. The length in bits of the collection of all the metadata easily exceeds the size of the original work, in particular if we are dealing with text. In the case of the fully XML tagged scientific journal articles of Elsevier the average overhead is around 100%, with peaks of 150%.

The beauty of this new minted coin is on both sides. With all metadata neatly in place and contained in XML structured files we will be able to route a product via various ways to distinctly different platforms. In other words a product will end up on different shelves depending on how the consumer rates the importance of certain hallmarks or features. This can be copyright free copying, certain quality levels in relation to the data (in the case of medical products), a typical combination of content descriptors (non violence, soft colours) or just all products from brand (or creator) X.

A new type of language
In conclusion one can say that just as humankind created and continuously develops language to exchange information on feelings, objects and desires, in the electronic era humankind creates a new

type of language called metadata which is a kind of well-organised shorthand for all that what can be said about the presentation of what one feels and thinks. Thinking in regular language is a human enterprise. Thinking in metadata will become mainly a machine operation, provided creator and consumer know how to instruct the machine.

Who creates or adds metadata?

Why all the above? After all we are not in a meeting devoted to data structures, or traditional classification schemas and indexing methodologies. The very name Bobcatsss is already a metadata indicating that. The easy answer to this serious question is that in an electronic environment we are dealing with a different dynamics than in the past. An electronic product is not anymore a single item, despite the monumental attempts by digital rights aficionados.

Stacking metadata & rights

Electronic products are stacks of information and related rights which on each level and between each component know fixed as well as permanently changing relations. For the sake of clarification let us just look at a power point presentation on this conference. First of all we have the ideas and intellectual property rights of the creator. The speaker wants to advance an idea and therewith uses lots of ideas of others, which all in their own right might have their own intellectual rights protection. By typing the slides, the speaker uses a PC on which he/she has a licence for the operating system, a licence for Power Point, a subscription for on-line internet use, and so on and so forth. On the other end of the line the consumer needs the same series of rights and adjustments in order to consume the message. It is all much more complicated underneath than sending a handwritten letter. Electronic communication demands an avalanche of legal, technical and conceptual adjustments, tuning, and transparent codification. No object is anymore single, all objects as well as their parts or combinations are interweaved and knotted together.

Even in colloquial inter-human electronic communications we have left the solid and safe analog shores of print on paper. One can say that every electronic message, product or not, is completely dependent on continuously changing sets of metadata, that describe the communication and need interpretation before consumption. For the regular citizen this has already the effect that people sheepishly buy their upgrades, new hard- and software and start believing that you cannot live anymore without them.

Luckily in the tug of war between technologies, standards, hypes and must-have gadgets, there is a space where the creator is still able to influence the process. It goes without saying that this place is primarily defined by the content of the product, by the very words, pictures or sounds carefully produced by the originator. But in the estranged world of electronic infrastructures survival is based on metadata. As argued above, they are the descriptors that funnel, like a swarm of lubricating servants, the product from creator to consumer. As in all power games the question now is who rules that swarm of helpers, goblins and angels alike? The bottom line has now been reached. After all we are still trying to sell a product!

Consumption is unwrapping metadata

A product is something a consumer only experiences in the usage, be it reading, eating or demolishing. The knowledge and emotions, mentioned above, that we have ably wrapped into metadata will be unwrapped by the consumer. In other words, the metadata wrap or shell is essential for the ultimate capability to consume the product. The beauty of an electronic environment is that in contrast to the old days the creator can exercise power on the meta level of metadata.

Creator controlled metadata

In the analog world, in almost all cases, the publisher decides how the final product will be presented, bound or cut. The marketing is a publishers' pursuit and closely related to perceptions of the present market. A Xmas game is published in the autumn and not in January. An author of a book is chased or put on ice depending of the seasons and the weather forecasts. A product must get immediate attention, television and radio interviews with its creator and preferably shelf space on eye level, or otherwise be piled on the floor at a place where everybody has to pass by. This place next to the cash register is equivalent to the place on a website or in a sponsored link in a search engine. Eye level becomes a metaphor. The creator in this case hardly plays a role, other then being interrogated about many aspects of his/her life, mostly with only a flimsy relationship to the work at stake, in interviews or at signing and presentation sessions.

General descriptors

How different is the world in an electronic environment. Keyword fields are cheap and every publisher wants to play safe. Even in cases where the publisher and the creator deeply disagree about the way a product must be positioned in the market, the middle ground is found by allowing or even insisting that the creator adds relevant keywords to the work. But also in cases where publisher and creator are not in disagreement but just are both unsure about the possible outlets, the enveloping metadata structure allows various strategies. Even if the already mentioned game is invented for Xmas, it is completely reasonable to add to this single target product descriptors that allow different interpretations.

For example the traditional sentence "*An ideal family game for under the Xmas tree*" can be taken apart into the following metadata: Kind=game, Difficulty=low to average, Number of simultaneous consumers= 3-5, X rating= 0, Violence level= low, etc., etc. This metadata taken as input for a marketing effort can also be easily interpreted as "*The solution for a rainy class outing on a camping site*". The sales possibilities then can be both; a cosy family at the fireplace and a bunch of bored youngsters in a tent. In other words: by a proper generic description, manipulating metadata can suggest a series of different potential consumer groups.

Multiple descriptors

Even more important then generic keywords as augmentation to more targeted descriptors is the possibility to have more different description lines. A historical novel or film always has various different dramatic lines. In an electronic environment it is very easy to work out these different lines in separated sets of metadata. One set can describe the historical context and can even relate to works of history or products that deal with the same period. Another set of metadata can deal with the clothing or the food presented, whilst yet another set is dealing with the psycho-analytical aspects of the characters.

Never out of print

In an electronic environment, a product has, in principle, an infinite life and never goes out of print. Infinite life can become a heavy burden as we all know from zombie and vampire films. The only way out is to re-energise periodically, not by sucking new blood from other products but by redefining the work into new and fashionable metadata. Such a new cycle can only be started if the creator, but also the publisher, is aware of the many different ways a work can be interpreted or viewed.

Out of print products now have to be searched for and most of them remain unknown, in oblivion. Simply because nobody has time to re-consume everything again in the hope that some old works turn out to be timeless. Here metadata come to rescue the product life. Searching on the level of well-defined metadata schemes remain possible, because they deal with generic descriptors and not only with time dependent notions or words. Take as a counter example the usage of a search engine on digitalised books of the 19th century. Most terms and words are different from our own language, and this seriously hampers our understanding of what they are talking about. The clever creator however adds metadata to describe his/her work in a variety of ways, therewith multiplying the chance that a work is not put on ice as soon as the production costs are covered and publishers or producers happily shelve the product. In the electronic era, out-of-print becomes an out-of-print term. In principle all works, literary, gossip, music, newsreels, scientific, etc. all remain available. Not necessarily on the original publisher's website or electronic warehouse, but maybe on a consumer's website or on-line archive. Old and new works are standing side by side waiting for consumption. The consumption of this infinite life is determined by the understanding of its contents. An important difference is the language in which the old and the new are presented. The meaning of words changes over time and new words are constantly popping up. This again is an argument against so-called free text searching on words only. The content needs some stratified coordination. We need an understanding that is expressed in notions about the content; in other words in metadata.

Conclusion

Creators in an analog world are forced to follow the marketing strategies of their publishers. These strategies are determined by local and temporal financial and cultural concerns. But a work can become interesting anew or again for the same or other audiences than originally thought of. The best way to enable this is that the creator, the one who translated knowledge and emotions into information racks his/her brain and tries to define multiple descriptors, following more general as well as just more specific aspects and add these notions to the product. The multiple and infinite life of a product in the electronic era is highly dependent on how the creator is able to explicate its knowledge and emotions into a variety of metadata.

27 November 2006