Metadata Project: What we want to do!

Götz Bachmann and Yuk Hui, 1.4.08

Metadata is commonly defined as data about data. It has gained a new role in the rapidly growing and changing world wide web, which might indeed morph into something that resembles, at least partly, Tim Berners Lee's vision of a Semantic Web (2001), based on more effective forms of generating, organising, aggregating, computing, exchanging, and using metadata. In computer and information science, metadata is now discussed as a shift in knowledge representation, and as a cultural phenomenon of the creation as well as the limitation of networks (Shirky 2005, 2008, Goulder 2007). We aim to add to this discussion: On the one hand via theories that come out of cultural studies, comparative literature, media theory and philosophy; on the other hand through ethnographic research, interviews, research events, experiments and experimental software.

A. Research questions – our five streams of enquiry

We divide our research into 5 main streams of enquiry, each of which is based on the idea of a shift.

A.1. Control and Power

The first stream of research looks into the role of metadata in the decrease and increase of control. Folksonomies based on tags by users are often heralded as a form of decrease of control and decentralisation of power. Whereas metadata was formerly thought of as something that had to be pre-organised in a structured and often hierarchically organised ontology, which furthermore was organised by a central agency

1

(think here of a library or an archive in the classical sense), a folksonomy is a loose and 'flat' structure that evolves out of the input of tagging users. But not only metadata becomes looser (in part). Additionally to this, metadata (be that in the form of ontology or folksonomy) takes over the role of other forms of control: Think of the shift from time-based programming on television (which is, as we all know, one of the key points of media power) to, what we might call, 'user-generated programming' on Youtube. To take another example: The Japanese social networking site Mixi includes the application Mixi Station, which automatically extracts usage metadata of Itunes and the Window Media Player, aggregates this metadata, and feeds it back to the users in the form of collective daily charts, which are based on the amount of times a piece of music is played – and not how often it is bought. Users decide about the charts not anymore by buying, but by listening.

However, these forms of wider and different distribution of power and loosening of control are at the same time matched with new forms of control and new forms of accumulation of power. Particularly striking examples for new forms of control can be found in some of the new versions of Digital Rights Management, which are deeply embedded in the single piece of content itself (e.g. via MPEG 7 or MPEG 21). Such metadata can decide about who can use the content when and how. It even enables the documentation, thus surveillance, of the usage. A second example: New forms of accumulated power might be found in the small and open communities of "metadata geeks" (quote from an early interviews). Such communities are at the moment at the forefront of the creation of the new ontologies (e.g. the "Music ontology") that develop the semantic web, other communities use such ontolgies to accumulate metadata. This might not be the power that Aristotle had, when he came up with his ontology, shaping in this way Western thinking up till today,

but this (dodgy) comparison might still point to the fact that there is power in the creation of ontologies.1 Nowadays, such communities of metadata geeks operate usually with an ethos of open source and free access, but they nevertheless order segments of the world not only for themselves, but for everyone who will follow on their footsteps.

Our initial hypothesis is that we might witness a shift to a new regime of media control and media power – of course only to a certain degree. The centre of control is not anymore only the in the access to the database to it. Now the access to the single piece of data itself is controlled, and new forms of power might be found in the participation in the collective creation of ontologies that drive the semantic web. Such an initial description of this shift is yet to be elaborated.

A.2. Metadata and Content

Our second enquiry looks at the relationship of metadata and content. Traditionally, this relationship was rather clearly defined: As data about data, metadata is data about content. However, increasingly metadata becomes content itself. To take an example that is already very familiar to all of us: On video platforms such as Youtube, the representation of the number of users, who have formerly watched a video (e.g.: "Views: 604.233"), becomes part of the user experience. The Japanese video-

1 "Ontology" means in computer sciences something slightly different than in philosophy. Both Computer Science and philosophy (before Kant) see ontology as the study of being qua being. However in Computer Sciences, an ontology is simply a *relatively* low level (that is: basic) form of taxonomy. Normally such computer ontologies are organised in hierarchical knowledge trees. There can be many different forms of ontolog*ies* for different purposes – to put it in the language of computer sciences: an ontology is always one form of "specification of conceptualisation" (Gruber 1993) -, and they often sit on top of each other. In philosophy, the idea of ontology is situated more deeply - some philosophers would argue: in existence itself. Aristotle's idea of ontology, for example consists of 10 categories of ways something can "be" (as substance, as quality ... etc.).

sharing platform Nico Nico Douga takes this further. It enables the users to write directly onto the moving image. The comments of the users merge with the video itself, as every following user sees the comments of former users directly on the clip, moving in and out of the frame of the moving image (more about this under 4.1.). We believe that such forms of metadata-as-content are at the forefront of a larger shift in the role of metadata (Weinberger 2004). Social networking sites, for example, often turn data about users automatically into metadata. Again, we can observe here a blurring of the boundaries of data and metadata. In advanced versions of the semantic web each website provides well-structured and machine-understandable metadata about other websites linked or otherwise connected to it – and vice versa. Whilst this in itself is nothing new – in a way links have always created a form of metadata about websites -, it gains a new momentum when computers have access to each website's metadata: Now the metadata of website A becomes part of the description of website B and vice versa. To understand this process of metadata becoming content, and content becoming metadata, we plan to engage with Gerard Genette's theory of 'literature on the second level' (1982). Genette's theory was written to understand trans- and paratextuality in literature. We ask, what happens, if such transtextuality takes the form of metadata, and therefore evolves to a new level through the computation of collective input.

A.3. Tags and Ontologies

The third form of enquiry looks into tags and ontologies. In the whole project we aim to establish a two way street between cultural theory and philosophy on the one hand, and developments in the area of metadata and tagging on the other hand. This is especially the case in this stream of enquiry. We will look into some of the hidden philosophical

assumptions in computer sciences, when they discuss the organisation of metadata. In return we will ask, how philosophy can take up some of the innovations in computer sciences. What kind of “philosophy” is, for example, informs a computer language like XML (which allows to embed complex “trees” of descriptive information in the form of tags)? How does this compare to the Resource Definition Framework (RDF), a common framework for metadata, which aims to express metadata by defining relationships (and not, like XML, though direct description)? What happens, if these two principles merge? When, for example, RDF based relationships are expressed in XML (which is often the case)?

To improve our understanding of the background of such questions we might have to look at the archival impulse, at historical practices of collecting and the collection, ask for early versions of catalogues, or look at the cultural history of descriptions that we take self evident (e.g. via authorship). We cannot develop such historical dimensions into a separate form of research. But we can use them to think about new forms of metadata, ontologies and, especially, new forms of tagging. Some of them might be already established, some exist only in our thought experiments, and some we might aim to realise: How, for example, would a system look like that allows tags to go beyond the simple ascription of “round” to an orange? What would happen, if we instead describe an orange as “75% round”? Do tags always have to be semantic? What would change if we use tags that purely relate, without making a further descriptive statement – e.g. a tag of a melody that consist out of a fragment of this melody, and relates this melody to other melodies without including any other form of description? Could we construct tags that get activated when the next user enters a relationship with a piece of content? And how can we theorise the relationship between the tagging user and his or her object? Is it purely descriptive (Jane tags an image of

an orange as ‘round’), or is it more the object (that is: the image of an orange) that ‘uses the users’ (that is: Jane) to find out more about itself?

A.4. Time of metadata

Our fourth enquiry looks into the relationship of metadata and time. As the systems of generating metadata become more and more sophisticated, metadata is turning into a constantly evolving field. The key strategy to prevent constant change in the structure of metadata systems seems to be to keep them very simple (see for example the Dublin Core initiative, which boiled down its categories to 15). As valid as this strategy is, it cannot be the only one. We will look into models for metadata systems that incorporate constant evolution. We also look into how digital objects themselves evolve inside such systems. When digital content and its description become inseparable, change in metadata can be seen as part of a change of the object itself. In that sense, constantly evolving metadata systems create constantly evolving digital objects. We are particularly interested in this, because these constantly emerging digital objects might produce a new form of media presence. The “now” of traditional media was produced through the synchronicity of broadcast and the imagined community of viewers, who watch the same content at the same time. However, if we encounter a constantly emerging digital object, we encounter always its latest version, on the way to further change through the mere fact of our encounter with it. A new form of media presence might emerge – once again, the Japanese video platform Nico Nico Douga provides striking examples.

Additionally to these two forms of ‘metadata-time’ – emergence and presence – we are also interested in a third form of time: Technological progress. The idea of such progress is often said to be out of date. And indeed, it smells a bit dusty: Of 50ties, of Eisenhower and of Stalin.

However, in the discussions around metadata and especially the semantic web we can observe strong and sometimes almost naïve beliefs in the potential of technological progress (the initiative to introduce "Agricultural Information Management Standards" by the United Nations, for example, sets out for no less than to "fight hunger in the world"). We are curious about this, because out of the perspective of most Cultural Studies scholars such beliefs have now been deconstructed for a long time. However, we are also interested, whether we can learn something. The absence of a strong notion of technical progress can be seen as a weakness of our own tools for thought – and many of the critical theories that we are used to engage with, have gained a lot of their strength through such beliefs. The idea of (technological) progress is an underling curiosity of this project.

A.5. Metadata topologies

Our last stream of enquiry looks into metadata topology and into the relationship of metadata to the digital space (of the internet) and the 'real' space of our physical body movements. The topology of metadata might be found in its underlying ontologies. More precise, it might be found in the basic operations that form these ontologies: trees, lists, and so. What kinds of operations are possible to form an ontology? Is it possible to come up with an 'ontology of ontologies'? On a second level, metadata also determines the topology of the web. Classically, metadata is merely seen as a form of description. However, metadata seems to now become more and more a form of relationship. As the topology of the internet moves beyond the classical hyperlink, new forms of connections such as RSS feeds or Mashups emerge. In such connections, a metadata scheme enables the flow of data, while, at the same time, metadata also starts to flows itself. How does it flow and at the same time regulate flows? What happens, when different ontologies collide? Which one is

folded into the other, and what are the criteria for that? The result might be a digital space, where increasingly not only users look for data, but also data starts to look for its users.

All this is further radicalised, when we look at how these trends merge with the space of our (physical) body movements. Location-based internet services on mobile phones, as well as ubiquitous interfaces and data sources in the public sphere (from RFID chips to surveillance cameras) make the management of metadata even more important, but at the same also even more problematic. Questions of power and control gain a new dimensions, when we remember that we already constantly produce metadata about our own body movements – RFID chips are just the most spectacular example for this. With the use of wifi, bluetooth, RFID and GPS, these data networks can connect with each other to form a totality of control.

Unquestionably, this stream of enquiry is the largest of all, and we have to see how far we get here. However, it will provide a background for all other questions. All in all, we think that it is not saying too much, if we predict that the relationships between digital things and other digital things, digital things and non-digital things, digital things and users, and users and other users might change once again profoundly – and that metadata and its management has a key role in this process. The five streams of enquiry are all directly related to each other. We expect to find answers to one of the stream by understanding the other streams of enquiry better.

B. Technologies in focus

Most of the empirical research will be done in the area of the moving image. Video, as a time-based and at the same time visual medium, has a very complex and thus interesting relationship to its own metadata. Here we are especially interested in the role of video-sharing portals. They have to handle extreme amounts of audiovisual material, which is often user-generated and originally nearly bare of metadata. They have to find ways to generate this metadata, and at the same time they have to find ways to represent parts of this metadata back to their users. A third reason why we are interested in video sharing portals: Control over them is one of the most sensitive areas of modern media. We therefore conduct most of our ethnographic research as well as most of our more experimental form of research in this area.

A second and additional possible area of the enquiry is audio data. As another time-based and non-textual medium, it again poses complex challenges to the generation of metadata. But in opposition to audio-visual data, systems for the organisation of audio data are already well advanced (there are, for example, already several sophisticated versions of music ontologies, whereas there has been made less progress in the area of audiovisual data). Social networking platforms could possibly become a third area of research, especially if they integrate music and video.

Around these central areas of research we have identified a set of developments of the web architecture that we need to understand (at least to some degree J).

- Existing ontology vocabularies like OWL (Web Ontology Language), Dublin Core (DC) or Friend of A Friend (FoaF) and the standardization of other ontologies.

- The properties of XML and XHTML: computer languages, which are often (but not always) used to carry and exchange metadata.
- The development of RDF - a framework, which can incorporate different ontologies (as well as tagging) into a single machine-readable document. It was proposed by W3C, and is thought to be the key to the semantic web.
- The current usage of open source programming languages, which enable the increase of metadata's significance: PHP (e.g. Facebook), Python (e.g. Youtube), Asynchronous JavaScript and XML (AJAX), Database (mysql) and microformat.
- Video formats/protocols like MPEG-7, MPEG-21 and the widely used interface progamming language Adobe Flash Action Script (the latter is used on Youtube and Tudou). Such formats provide extended possibilities for embedded metadata.
- Browsers and browser plug-ins. They are key to the aggregation of metadata, metadata extraction, RSS feed subscription, and to mashups of data (one example: the Operator plugin for Firefox extracts metadata of an event from a webpage and links them to Google or Yahoo calendars).

C. Research

We have devided our empirical approach into 5 method streams: Each will combine a certain method with one or several case studies. The five method streams do not match directly the five enquiries mapped out under A. The opposite is the case: All method streams will not only address one streams of enquiry, but several at the same time.

C.1. Longer time ethnographic participation

We aim to research in our London ethnographic fieldwork the 'birth' of ontologies. Ontologies order metadata, stimulate its production, facilitate its exchange and ease its usage. In a way they can be described as the 'genetic code' of the new semantic web. Ontologies are normally born in projects that involve worldwide collaboration. They are thus not really confined to one particular place. However, London and the South of England seem to us as a particular hub for this (Tim Berners Lee, to take the most obvious example, is institutionally based at the University of Southampton). Thus we have decided to research this side of our project – the early days of new ontologies – largely in Britain. The second reason to pursue this side of the project in the UK is that we aim to follow the early days of such ontologies over a longer period of time – 18 months, to be more precise. As we are based in London, we can use here a form of media ethnography that is it looser and more long-term.

However, the process of the early development of such ontologies of course extends the 18 months. We have therefore decided to conduct our long-term ethnographic fieldwork in four settings. Each represents a certain moment of time in the process of the come-to-being of ontologies: (1) We aim to observe one of the working groups of W3C, the World Wide Web Consortium. W3C promotes the semantic web, and sets up recommendations. We are interested in W3C, because this organisation sets the frame for the development of ontologies – for now, but also, of course, for future ontologies. (2.) Transmission.cc is an international network of citizen journalists, programmers and video makers, who try to set up a metadata standard for independent audiovisual material, thus improving the search and the accessibility of such material. This group aims to create a new ontology, and it is still pretty early days. (3.) The Music Ontology (MO), developed by Yves

Raimond et al., is a new ontology for music-related metadata that aims to include all descriptions related to music. Currently this ontology has come to a relatively stable version, and will be used by the large music metadata sharing website MusicBrainz in the near future. We can therefore observe the (relative) stabilisation of an ontology. (4.) Omras2 (Online Music Recognition And Searching). This research project develops a new form of metadata management for music files. It will store more complex information (e.g. on the preferences of other listeners) and will thus enable, for example, smart playlists for music, comparable to a desktop-based form of last.fm. In doing so, it will implement the Music Ontology that was developed by Yves Raimond (see above). We will take part in the process of its implementation.

All these projects are rather loose networks of academics, activists, metadata geeks and stakeholders in the industry form. Some of them have never met in person. We will follow all four case studies for 18 months. We will conduct interviews with some of the England-based developers; follow the mailing lists; attend meetings, workshops, and events; and, possibly, and to some extend, also participating in their production process.

This research will be conducted in close collaboration of Goetz and Yuk.

C.2. Intensive ethnographic case studies in East Asia

Two intensive ethnographic case studies around video-sharing platforms will form the core of the project.

The first one will look into the practices around the Japanese platform Nico Nico Douga (nicovideo.jp). This platform was originally created as a

mashup of Youtube on the one hand, and 2-channel on the other hand – the latter is Japan's most popular BBS (bulletin board system). As a mashup, it is itself (as a platform) the result of more effective metadata management. Nico Nico Douga exists since little more then one year, but is by now the 6th most visited website in Japan. Even though Nico Nico Douga will probably remain a specifically East Asian phenomena - BBS are much more popular in East Asia then in the West -, Nico Nico Douga has on top of this several features that make it interesting for our study of metadata. As already mentioned above, users of Nico Nico Douga can write directly on the moving image while they watch videos. All following users can then see the remarks of former users as part of the video (the remarks scroll into frame of the moving image from right to left, and remain visible for about 4 seconds). They comment the video as well as each other's comments. This feature is the key to the success of Nico Nico Douga. The new re-launched version of Nico Nico Douga (Nicoscript, launched in March 2008) ads further features: The possibility to insert interactive buttons in the video, from which users can jump to any other video at any moment in time (that is: from video x at minute 1.45 to video Y at minute 1.23). It also offers the possibility to integrate little polls or games directly inside the video, and the opportunity to create multi-linear narratives. Additionally to this, the interface of Nico Nico Douga also displays a high complexity of other forms of metadata. The content itself consists largely of re-sampled fragments of TV series and music pieces. Content is uploaded by the users just as on Youtube. However, initial analysis seemed to indicate that content is often directly aimed to provoke the generation of metadata. It almost seems as if on Nico Nico Douga it is not metadata, which serves data, but data, which serves metadata.

Our second case study will look at the Chinese video-sharing platform Toodou (Tudou.com). Based in Shanghai, Todou was launched on April 15, 2005. By September 2007 it uses the world's largest CDN (Content delivery network) to serve over 55 million videos each day. It competes with Youtube to be the largest video-sharing website in the world (it streams less clips, but more video minutes than Youtube). Most of the features of Toodou resemble more or less the features of Youtube. Our interest in Todou is thus more inspired by its specific role in the Chinese mediascape.

Both ethnographic case studies will aim to collide as much data as possible in short but intensive periods of fieldwork in Tokyo and Shanghai, which will each last 4-6 weeks. A locally based assistant will help the research associate to organise the interviews. He or she will translate where necessary, and provide additional technical and cultural expertise. His or her task will also be a first analysis of the portal and the content, and a comparison of the portal with other national websites as well as with international competitors.

At the centre of the fieldwork itself will be interviews and focus groups with users. Some of these interviews will include ethnographic elements such as joint watching, uploading, video producing, and writing on the video (in the case of Nico Nico Douga). We aim to organise Nico-Nico-Douga- and Todou-parties – fun-events, where several users show each other their favourite clips. We also aim to interview some of the 'stars' that these platforms have produced. We might watch content with people, who do this for the first time. We will produce and upload some content ourselves – ideally this would take the form of targeted mini-experiments. On the producer side we aim to interview interface designers, programmers, content and channel managers, marketeers and

the founders of the platforms. If possible, we will conduct short ethnographic studies of one or two weeks inside the companies. Ideally we will be able to model these studies similar to the classical 'gate-keeper-studies' in media ethnography. We aim to do so in both platforms, but particularly so in the case of Todou (however, access might be here especially difficult due to Chinese sensitivities). We also aim to talk to other commercial content producers (how do they react to the fact that fragments of their content are used?) and media lawyers, and we will discuss the platforms with film critics, anime critics, commentators such as bloggers, academic and non-academic internet theorists, media artists and other artists, who combine text and image.

Goetz will conduct this part of the research, possibly with some cooperation with Yuk (in the research in China). He will be assisted with local assistance in all three countries (3 months in Japan and China, 6 weeks in South Korea)

C.3. Experiments

The ethnographic research outlined above already includes some smaller experimental elements. Additional to this we plan to conduct one larger Experiment: the MTM-experiment (Marathon-Thoughtbyte-Metadata -experiment). The basic idea is to create several hundred short videos with intellectual content, put them on different video sharing platforms, and harvest the metadata 6 months later – this way we will be able to observe not only which videos died, and which have acquired a life on their own, but also how they have re-assembled with other video data and user feedback.

The experiment includes 4 steps: (1.) Acquiring the video data. (2.) We slice up the video material into “thought-bytes” of 1-5 minutes length. All in all we aim for about 300-500 of such thought-byte-videos. We stick these on Youtube. We could do the same on Todou and on Nico Nico Douga. In this case we might have to subtitle it, or even get additional Japanese and Chinese material. (3.) After 6 months, we harvest the metadata (this will be done through writing computer programs, using Google API – Yuk can do that). (4.) When we look at the data, we ask questions like: Which thought-byte died, which flourished? How did it flourish? What are its tags, comments, relationships and connections to other videos, new videos and video comments, subgenres, user numbers, channels? How do these thought-bytes come alive inside the metadata environment of these platforms? We will treat video as living and changing entities inside these platforms, and metadata as a generating environment (instead of a regulating one).

This research will be developed as a collaboration between Scott, Goetz and Yuk. We might use some of our additional resources to hire assistance for the editing, subtitling and uploading of the video clips.

Additional to this form of an experiment in real life we conduct constant thought experiments, thinking through theoretical ideas via imaging the construction of applications, and thinking through possible applications via imagining how they would look like, if this or that theory would be applied. We do so not only internally inside the project. Thought experiments are an essential part of our expert interviews (see C.5), the ethnographic studies and the development of applications (see C.1 and 2).

C.4. Application

Part of our research is to build an experimental piece of software. We are still exploring the possibilities. However, we have made two decisions: Firstly, we aim to develop this application not alone, but in cooperation with partners. Secondly we want to build an interface and a metadata mangement system for audiovisual content. It should a.) enable rich and new forms of input of metadata, b.) display the metadata in a way that metadata becomes part of its content, and c.) react to the specific requirements of the moving image, and especially long blocks of filmed public talks.

One idea would be an interface that allows users to comment directly on the moving image – similarly to the interface of Nico Nico Douga. We expect that such a possibility to comment directly onto a film of a talk of, say, Katherine Hayles, could produce interesting effects. Users could comment Katherine Hayles talk, but also interact with each other – again all this against the backdrop of the timeline of Katherine Hayles talking …

A second idea is time-based tagging: A tag for a segment of a video. Such tags are tricky: They work differently to comments. A comment makes sense as something that is consumed after the particular piece of video, which originally inspired the comment. A tag would only make sense, if the tag is placed before the moment in the video: If someone clicks on the tag “orange”, he or she will want to see the start of the sequence of the film of the orange, and not its end. But the retrospective identification of the start of a sequence could make the process of tagging potentially too hazardous for the tagging user. How can we solve this problem? Until now this is still an open question.

Thirdly, we would like to explore here the idea of “fuzzy tags” (the description of a segment of a film as “70% round” or “90% boring”). We also aim to explore ideas about an emotive ontology for video of intellectual content. If we are successful, we might be able to create an interface that would allow a hybrid of surfing and zapping, where users jump between different videos at different moments.

This part of the research will be developed of Zimmer and Yuk, with some participation of Goetz and Scott.

C.5. Expert interviews

In UK, China, Japan and South Korea we will conduct additional expert interviews (in China and in Japan we will do these during the trips outlined in chapter 3. These expert interviews will be designed to help us to gain a deeper understanding of the semantic web in general. We have a particular interest in any area that might help us to understand the ‘dark’ side of the semantic web: its possibilities of control, censorship, surveillance. We also follow up in our interviews some of the trends of the usage of metadata in social networking platforms. Furthermore, we are interested in new ideas about ‘metadata art’ and ‘metadata curating’. We will therefore conduct expert interviews, for example with visionaries and critics of the semantic web, experts on DRM; activists against the ‘Great Firewall’, media artists and curators.

Scott and Goetz will conduct this part of the research, possibly with some involvement of Yuk in the interviews in China. We aim to synchronise the research in China with the fieldwork conducted by Chris Berry in Shanghai.

D. Outcomes

Academic papers will summarise the results of our ethnographic case studies (intensive as well as long term) and our experiments. A website and a blog will document the ongoing research process and allow the publication of early drafts of our papers. Interested people can gain access to most areas of our internal collaborative workspace on Google Documents (however, they would have to subscribe via an email). Events form a further outcome – the November-workshop in Goldsmiths and, possibly, research events in Japan, China and South Korea. We will participate in the creation of at least one new application (or an ontology, for that matter). We are also exploring the possibility of the presenting results in other non-textual forms (in conjunction with Nina Wakeford's ESRC fellowship).

E. Time plan

March 08
Organisation of the project – content, case studies, methods, workflow, admin, blog/website.

March 08 – July 09
Fieldwork in Omras, Music ontology, W3C, transmission.org (loosely ongoing for 1.5 years)

April 08
Preparation fieldwork Japan (Osaka and Tokyo), Analysis of the website

May 08 – June 08
Fieldwork in Osaka and Tokyo (Nico Nico Douga, expert interviews)

July 08 – November 08
First phase MTM-experiment (creation and uploading of the thought-bytes); transcription and start of analysis of the Japan material; start work on the Tag-the-Tate Application (negotiations with the Tate etc…)

November 08
Workshop, presentation of initial results (Nico Nico Douga)

December 08 –January 09
Preparation fieldwork China and Korea

February and March 09
Fieldwork in Shanghai (Todou, expert interviews) and Seoul (expert interviews, possibly an event)

April 09 – May 09
Transcription and start of analysis of China and Korea material

June 09
Second phase of the MTM-experiment (harvesting the data)

July 09 – Feb 10
Writing papers, creating experimental art works, finishing of the Tag-the-Tate-application.

F. Changes to the original proposal

In comparison to the initial proposal and the first preliminary report, the project has not changed in any substantial way. However, we have made a series of specifications and decisions since the first report:

(1.) We have developed the original three streams of enquiry to five. The two new streams are 'metadata and time' and 'metadata as content'. They evolved out of the work on the presentation on 'metadata and archive', which we developed in December 2007 together with Rachel Moore. They pose questions explicitly that we have to address in any case. They are therefore not really new. They are just turning implicit questions into explicit ones.

(2.) All five streams of enquiry are closely connected, and they can, so our hunch, only be answered in conjunction – theoretically as well as empirically. In this way we developed the methods and case studies. Each of them addresses several streams of enquiry (whereas in the original proposals we identified each enquiry with a specific research method and case study).

(3.) We added Nico Nico Douga to our research agenda, which requires fieldwork in Tokyo. For reasons see above. This was made possible through the additional funds that we have gained though the employment of Goetz Bachmann as research Associate (his former buy out can now be used for new purposes, e.g. this extension).

(4.) We have elaborated the idea of the application. We aim to collaborate with partners. This will enable us to work on a more ambitious application. It will also provide us with a complex and rich ethnographic

setting, as we will have access to the research agendas and practices of our potential partners.

(5.) We have modified our methodological idea of 'laboratory events'. Originally we aimed to bring people with different stakes together: activists with industry stakeholders, and so on. However, we have been informed that especially in the East Asian research this is not the best way forward: It could cause more silence than interesting discussion. In Japan this is due to general styles of communication, in China due to the sensitive nature of some of the issues raised. We have not abandoned this idea for good – we will still organise events, and we still like to see the whole project as a form of laboratory. But we have decided to put it as method in a less central position, and use additionally a range of more traditional methods (intensive ethnography, interviews, long-term ethnography, small events, focus groups). We also added the idea of the 'experiment" – so we do not loose out on our ambition to develop innovative methods for cultural research.

We are well aware that this is a very ambitious plan. Not all questions might get answered, not all case studies might be followed to the end, and new ideas might constantly spring up. This is still very early days. We will be lucky and happy, if we achieve a significant chunk of what we aim for.

G. Literature

Tim Berners-Lee, James Hendler and Ora Lassila, 2001, The Semantic Web, Scientific American Magazine.

Scott A. Golder and Bernardo A. Huberman, 2007, The Structure of Collaborative Tagging Systems, Information Dynamics Lab, HP Labs.

Thomas R. Gruber, 1993, Toward principles for the design of ontologies used for knowledge sharing, in N. Guarino and R. Poli, (Eds.), International Workshop on Formal Ontology, Padova, Italy.

Gerard Genette, 1982, Palimpsestes: La littérature au second degree, Seuil.

Clay Shirky, 2005, Ontology is Overrated.

Clay Shirky, 2008, Here Comes Everybody: The Power of Organizing Without Organizations, Allan Lane.

David Weinberger, 2004, The end of data. http://www.hyperorg.com/backissues/joho-oct15-04.html#data (retrieved March 1st 08).