

Over het archiveren van elektronische berichten

On the Archiving of Electronic messages

Voorstel van Tjebbe van Tijen geschreven in het voorjaar van 1994 voor het blijvend archiveren van elektronische berichten met relevantie voor het bestuderen van de geschiedenis van moderne sociale bewegingen door het Internationaal Instituut voor Sociale Geschiedenis (IISG). In eerste instantie was de naam van het project WAAL (Wide Area Archive & Library), mede geïnspireerd door de vestigingsplaats van de voorgestelde projectpartner de stichting Antenna, Nijmegen, dat aan de rivier de Waal ligt. Later werd de naam gewijzigd in 'OCCASIO, Digital Social History Archive'. Het duurde enige jaren voordat het project met behulp van een subsidie van het NWO van start kon gaan. Inmiddels heeft het Occasio project een vaste plaats binnen het IISG gekregen.

*But now thou of a Mercury art sire Of thine own name, a post with whom the wind,
Should it contend, would be left far behind. Whole message, as thy metal, strikes
the gold Quite through a wedge of silver uncontrol'd; And in a moment's space
doth pass as far As from the artic to th' antarctic star So proving what is said of
influence, They neither of them have such a quality As a relation to locality; No
places distance hindering their commerce, Who freely traffic through the universe;
And in a minute can voyage make Over the ocean's universal lake.*

Een citaat uit de inleiding van Sir Francis Kinaston voor het boek van John Wilkins "Mercury, the secret and swift messenger", London 1641. Een werk waarin naast cryptografie een uitgebreide verhandeling gegeven wordt van methoden om een bericht snel over te brengen. De meest uiteenlopende voorstellen worden gedaan: 'arrows', 'bullets', 'beasts', 'birds', 'sound', 'tunes and musical notes', 'species of sight' (Pythagoras could write anything in the body of the moon, so as it might be legible to another at a great distance), 'fire', 'smoke', en natuurlijk de snelste methode van allen 'het direct overbrengen van gedachten'. Met veelvuldige verwijzing naar de klassieken, zoals gebruikelijk in de Renaissance, vinden we in dit boek een groot deel van de principes van de moderne telecommunicatiemiddelen beschreven van melodieuze piepende faxen en modems tot encoding en standaardisatie van teksten, wat de basis van alle elektronische berichtenverkeer is. Het heeft een paar eeuwen geduurd voordat deze principes hun optimale gebruik kregen, maar als een technologie eenmaal breed toegankelijk is kan de groei zo razendsnel gaan dat nu met recht gesproken kan worden van een 'telematische revolutie'.

De invloed van berichtenverkeer op het verloop van maatschappelijke gebeurtenissen is groot. John Wilkins vatte het in 1641 al samen: "...the ignorance of secret and swift conveyances, hath often proved fatal, not only to the ruin of particular persons, but also of whole armies and kingdoms". In roerige tijden

droegen Romeinse consuls een duif in hun gewaad om direct het 'thuisfront' te hulp te kunnen roepen. Het gaat niet enkel op voor de mate van geïnformeerd zijn van machthebbers, maar ook voor dat wat we nu de 'publieke opinie' noemen. Zagen we tot voor kort nog een gestaag proces van monopolisering van de media die vorm gaven aan die 'opinie', de laatste jaren is een tegenovergestelde ontwikkeling zichtbaar, waarbij er media ontstaan, die niet langer, of minder, onder controle van de vertegenwoordigers van de heersende maatschappelijke orde staan.

In haar inleiding op het recent verschenen boek 'Global networks, computers and international communication' vat Linda M. Harasim dit samen: "McLuhan foresaw global connectivity decades ago. But whereas the broadcast media of McLuhan's time and vision implied populations of passive consumers, today's computer communication networks enable communities of active participants." Het zijn de immateriële 'geschriften' van deze nieuwe actieve gemeenschappen die naar mijn mening even zo zeer een plaats in de verzameling van het Internationaal Instituut voor Sociale Geschiedenis verdienen als nu verzamelde informatiedragers. Ooit waren er mensen die met verachting spraken over met losse letters gedrukte boeken, die het handgeschreven en geschilderd boek het hoogst achtten en hun verzamelbeleid niet verder wilden laten strekken dan incunabelen.

We zullen door de huidige technologische ontwikkelingen steeds meer te maken krijgen met het fenomeen dat de informatiedragers steeds minder van elkaar te onderscheiden zijn. Door digitale technieken vervagen de grenzen tussen wat nog film, video, foto is. Door 'desktop publishing' vervaagt de grens tussen manuscript en boek. Door elektronische berichtenverkeer is het 'ruwe materiaal' voor kranten, zoals vroeger de telex berichten, direct voor een ieder beschikbaar. Door de verbazende groei van onafhankelijke 'Electronische Bulletin Board Systemen' ontstaan nieuwe communicatie middelen en maatschappelijke platforms als 'elektronische conferenties'.

In principe verzamelt en bewaart het IISG het materiaal om haar inhoud, niet om de specifieke materiële vorm waarin het zich presenteert (omdat het pamflet of brochure is). De in dit rapport weergegeven selectie van elektronische berichten rechtvaardigt naar mijn een gelijke behandeling.

In het navolgende rapport wordt voorgesteld om dat deel van het wereldwijde elektronische berichtenverkeer te gaan archiveren dat relevant is vanuit het verzamelbeleid van het IISG. Het rapport bestaat uit vijf onderdelen:

- 1 oorspronkelijke tekst van het voorstel "The creation of a Wide Area Archive & Library (WAAL)" [p.4];
- 2 kort overzicht van het soort berichten dat verzameld en gearhiveerd zou moeten worden [p.8];
- 3 volledige teksten van deze berichten [p.13-104];

- 4 een tekst over het gebruik van nieuwe media in sociale bewegingen [p.105];
- 5 een keuze uit een aantal interviews met 'elektronische netwerkers', met name over het probleem van archiveren (ter beschikking gesteld door Peter van der Pouw Kraan, die binnenkort bij de uitgeverij Ravijn een boek over dit onderwerp publiceert) [p.112].

Ik heb er voor gekozen de berichten die onderdeel uitmaken van de 'elektronische bloemlezing' in hun geheel op te nemen om een goed inzicht te geven in de aard van deze documenten zonder dat de potentieel lezer eerst gedwongen is de techniek van doorzoeken van elektronische netwerken onder de knie te krijgen. Voor degenen die zich verder in deze materie willen verdiepen kan ik de volgende titels aanraden:

- GILSTER, Paul "The Internet navigator"; John Wiley & Sons; New York/..., 1993.
- HAHN, Harley/STOUT, Rick "The Internet yellow pages"; Osborne McGraw-Hill, 1994.
- HARASIM, Linda M. "Global networks"; The MIT Press; Cambridge Mas/London, 1993.
- KROL, Ed "The Whole Internet"; O'Reilly & Associates; Sebastopol CA, 1994.
- RITTNER, Don "The Whole Earth online almanac"; Brady; New York/..., 1993.

Rest nog Michael Polman van de Stichting Antenna en Peter van der Pouw Kraan te bedanken voor hun bijdragen.

Tjebbe van Tijen

17 Mei 1994

The creation of a Wide Area Archive & Library (WAAL)

Proposal for the International Institute of Social History Tjebbe van Tijen, 7 April 1994

The archive & library consists of digital documents representing all kinds of information from text (in the first stages of the project) to images and sound (in the future). The reason for constituting a WAAL is that, although the production and proliferation of electronic documents has been astronomic, there has very little been done for long term preservation of this kind of information. It is clear that there is a strong impact on society by the new information technologies, especially through the diffusion of information by telecommunication. This phenomenon has been compared on several occasions with the 'revolution of the printed word' as it developed from the 15th. century onwards. The 'digital revolution' will be a popular subject for historical study soon. To make such studies possible we have to act now to rescue what will otherwise be lost forever.

There are distinct differences between printed and digital information. The first is tangible and readable without any devices (except spectacles in some cases), the second is disembodied and can only be perceived with the help of special appliances. Papyrus, parchment and paper have carried information from generation to generation for more than 4000 years. It is not likely that this 'paper memory system' will be fully replaced by digital documents as some over enthusiastic computer lovers propagate. Nevertheless we should start to take care also of the 'digital memory system', if we do not want to leave our predecessors with a historical void. The new form of information circulation over electronic networks has a very ephemeral quality. Text is often written and read directly on and from the computer terminal screen. Not much thought is given to long term preservation of such texts and if so the necessary facilities, finances and expertise are not available.

There are a few characteristics of this new media that will force us to rethink the concepts we use to determine the selective criteria for building historical collections of information items. The notions of 'small' and 'big' publishers, limited and wide circulation, are less applicable. The ease with which documents can be duplicated, adapted and re-circulated, placed from one electronic bulletin board to a whole network, from one network to other networks, does away with the earlier distinction between mass media and its implicit counter part 'non mass media'. The ease with which one can now circulate information from local to global scale has also consequences for another concept of the 'paper world' collection building, that is the importance given to the 'place of origin' of an information item. Collections are often build up geographically. Consequently, work tasks are also divided over different geographical areas. With the implosion of physical space in the world wide electronic network collection structuring and task division should be revised. Collecting digital information can be done from any point in the interconnected global network. The traditional division in document types like correspondence, manuscript, book, periodical, press release, pamphlet, hand out, leaflet, is getting less distinctive. A whole chain of activities of the publisher, printer, distributor, bookshop, has suddenly been united in one process: computer networking.

Of course the International Institute of Social History should make a selection of the hundreds of thousands digital documents that are (still) available now. One major network with an information content closest to the collection profile of the Institute is the Association for Progressive Communications (APC). APC started in 1984 in the San Francisco Bay Area as an initiative of the Ark Communications Institute, the Center for Innovative Diplomacy, Community Data Processing and the Foundation for the Arts of Peace (at that time called PeaceNet). In 1987 PeaceNet was managed by the newly formed Institute for Global Communications (IGC), set up by the Tides Foundation. Other networks were created, such as EcoNet and ConflictNet. Among the financial supporters of these initiatives was Apple Computer. Later the network made connections with similar initiatives in other countries such as GreenNet in England. In 1987 Peter Gabriels directed financial support to the project from a fund raising rock concert in Tokyo (the year before). The transatlantic link with GreenNet proved so successful that other funds for furthering the net could be found from foundations like MacArthur, Ford,

General Service and the United Nations Development Program. In 1990 the Association for Progressive Communications was formed to coordinate the by now global networking activities. There were more than 15.000 subscribers in 90 countries in 1993, mostly Non-Governmental Organisations (NGO) (see map).

The outline of the proposal that I discussed last week with Michael Polman and Alfred Heitink from the Antenna Foundation in Nijmegen reads as follows:

First step: gather all archive material of the APC network, in as far as it has been preserved somewhere in the world. A rough estimate is that it will be between 2 and 3 Gigabyte since 1984. The daily feed of material is now around 1 Mb per day. This estimate is mainly material in English, but also includes text in Spanish, German and Portuguese. The proposal is to make a contract with the representative of the APC network in the Netherlands, the Antenna Foundation. The Antenna Foundation will make a separate agreement with another partner, GreenNet in London, to assure long term continuity. In principle all APC materials are free on the network. Participating host organisations in different countries have an agreement that they only will charge for the transport costs of the information, not for the information itself. There are some exceptions, as with the materials from International Press Service (IPS). In such cases separate deals need to be made with these information providers.

At the moment the most cost effective and safe method of preservation is writing the digital archive material to CD-ROM. Each CD-ROM has a capacity of a bit more than 600 Mb. The whole APC archive could be written on 5 to 6 of such CD-ROMs. With the lowering of the prices of hardware and software it is feasible now to buy a CD-Recordable writing device with a dedicated computer and appliances for a price around fl. 15.000,-. Blank CD-Recordable discs cost now between fl. 50,- and fl. 75,- a piece. The writing of the CD's can thus be done 'in house'. The great advantage is that once the material has been prepared for storing on a CD-ROM, other copies can be made easily and cheaply, either by 'burning' another CD-Recordable, or duplicate them in a small copy range through a duplicating company. Also the same digital material can be formatted on CD-ROM for usage on different platforms (PC, MAC, UNIX). Also duplicates of archives can be exchanged with other institutions or made into a publication. Of course there need to be permissions by copyright holders before such a publication can be made.

.The main steps for the APC project will be:

- archiving/preservation;
- classifying/normalisation;
- making the material public available.

Each of these tasks can be divided in separate steps:

Archiving/preservation

- through direct Internet connections;
- archive materials on DAT cartridges;

- the original structure of bulletin boards and networks with news groups, subject lists, conferences, electronic journals and file sections will be preserved as much as possible;
- deselection by automatic filtering, for instance all messages of less than 5 Kb, or messages that consist mainly of quotations of other messages;
- detection of double items on the basis of unique 'message ID' (only within a news group);
- registration of verification by using ... of original text.

Classifying/normalisation

- automatic description on the basis of formal elements in the headers of messages (from - date - subject line);
- registration of conference(s) or list(s) where the message has been posted (also multiple appearance);
- automatic classification of specific names derived from full text (person, corporations, geographical names) on the basis of expertise dictionaries;
- semi automatic classification with descriptors/keywords on the basis of expertise dictionaries, in such way that sets of message descriptions can easily be selected or deselected by the classifier;
- normalisation of text that has been non correctly formatted;
- reformatting for CD-ROM of view copy of texts that use national/language specific routines for non lower ASCII characters.

Making the material public available

- Bringing the indexes that refer to the full text on line (through an existing bulletin board system, direct dialling, on Internet, distributing the index to other bulletin boards);
- bringing the whole text on line (so-called FTP site), either based at a computer at the Institute or for instance on the GreenNet computer in London;
- establishing a service whereby on the basis of the descriptions (indexes) selections of text can be made 'on line' or by buying a floppy disc for use at home; the requested material can than be delivered on floppy, in an email box or on a CD-Recordable (with an automatic billing and payment registration program);
- and of course consultation directly at the Institute.

Once the information is preserved on CD-ROM an 'on line resource center' will be constituted at the International Institute of Social History.

Costs

A rough estimation of costs that can be divided in one time investments and annual exploitation costs. Although the dynamic hardware and software market

will make it necessary to renew the hardware and software on a regular basis.

Starting options:

- Hardware and software for archiving materials on CD-ROM 15.000,-
- Multiple CD-ROM player to put in local and external network 5.000,-
- Software development, training and support 10.000,-
- Transport costs of data 5.000,-
- Peripheral equipment (high speed modems, cabling, network facilities) 5.000,-

How to proceed

I propose that the project will be developed in stages whereby at the first stage the project will be set up by an external company on the basis of a contract with a fixed price. The Antenna Foundation will be the most suitable candidate. For the project there will be formed a steering committee with 2 representatives of the Institute and two of the company. The project will include training of personnel of the Institute. The dedicated software that has to be developed should, as much as possible, be made up of combinations of existing widely accepted software modules and its construction should be modular and be open to adaptations by the Institute. The software should be able to handle a wide variety of text and database material formats and platforms.

Extras

Of course when the Institute decides to do the writing of CD-ROMs 'in house', the equipment can be used at the same time for other projects as:

- a compilation of existing text format inventories of the Institute and affiliates (with one general index);
- publication of new inventories on CD-ROM (ID Archiv);
- Archives de Bakunin;
- publication of the general catalogue (OPC) on CD-ROM;
- back up safety copies of images files of the iconographic department.

Kort overzicht van berichten

Om een representatief overzicht te geven van de vele tienduizenden berichten die de afgelopen 6 jaar op het APC Netwerk gecirculeerd hebben is vrijwel onbegonnen werk. De hierna volgende bloemlezing is binnen een kort tijdsbestek samengesteld met als doel buitenstaanders een eerste indruk te geven. Het aantal onderwerpen dat op dit netwerk aan de orde komt is veel ruimer dan in dit overzicht. Ik hoop dat het desondanks enig inzicht geeft.

Eerst volgt een kort overzicht dat gegroepeerd is naar door mij aan ieder bericht

toegekende sorteersleutels: (een enkel) trefwoord, werelddeel, land(en), jaar, maand en dag. De originele berichten zijn daarna in hun geheel indezelfde volgorde opgenomen. Achter ieder korte samenvatting staat het paginanummer waar het betreffend bericht te vinden is. Indien nog aanwezig heb ik de zogenaamde 'headers' van de berichten in takt gelaten. Het ziet er vaak wat ingewikkeld uit, maar voor kenners valt uit zo'n header veel uit af te leiden: herkomst, bestemmingen, op welke wijze de verzending heeft plaats gevonden.

Het feit dat de hier opgenomen berichten voornamelijk in het Engels zijn betekent niet dat dit voor alle berichten op het APC netwerk geldt. Er zijn omvangrijke berichtenstromen in een groot aantal andere talen op dit netwerk te vinden.

(...)

We cannot depend on producers of these electronic forums to maintain full archives of the title, any more than we have been able to count on print publishers to maintain a historical archive of their works.

[Billy Baron in his description of the history of the CIC Electronic Journal Archive, August 1993]

INTRODUCTION

The WAAL Project is not a one off initiative but a first step in a series of literary far reaching changes. With the advent of new digital technologies all kind of different media can be stored, copied, accessed and distributed over a more and more uniformed communication system. Different kind of (physical) information carriers, like books, photo's, sound recordings, movie film or video can all be transferred to digital formats and thus be processed as described before.

The traditional form of indirect access to information through the search in catalogues and other sources of bibliographical information, sometimes expanded by the use of abstracts and followed by the ordering and delivery of the requested document in a manual way, will be superseded by an access form whereby information about information and the information itself will merge into one 'search & delivery' system. The fact that this information can be formatted in a standardised way (Z39.50-1992, SGML, HTML) and thus be accessed from any point on the fast growing global network of computer systems, will have a profound effect on the future functioning of institutions like the International Institute of Social History. For some this may seem developments that will happen in a far away future, but for several years a project is realized in the IISG itself that potentially is exactly such a merged form of information retrieval and delivery.

The digital iconographical materials catalogue (Beeldprojekt) of the IISG (integrated in the General Computer Catalogue) is a project whereby descriptions of different documents, on different kind of information carriers, ranging from stickers and photographs to posters and postcards, are directly linked to a

digitized picture of the document itself. On a few special computer terminals with two monitors the user can search in the catalogue, using a traditional text screen. When a document with the indicator (AV = Audio Visual materials) is found the image can be retrieved on the adjacent graphical monitor. The use of this facility is still restricted to the IISG itself and copyright restrictions might frustrate remote access to it, but potentially it is a step on the way to the new electronic archive & library.

At this moment 'remote access' to the computerized catalogue of IISG is only possible within the restricted network of the local Amsterdam network of OPACs of university libraries and academic institutions: ADAMNET. For the rest the potential user can inform him/herself through the traditional consultation of published guides and inventories to the archives and other collections of the IISG, in other libraries or at home, or has to visit the institute personally. This policy of 'when someone is interested they should come to us', was the prevailing attitude only a few years ago. Now the institute has 'to get out to them'. One could paraphrase a title of a promotional booklet published on the occasion of the recent relocation of the IISG to the new premises at the Cruquiusweg "Moving Marx": 'the whole world will be moved'. The ongoing information revolution is not so much an abrupt rupture with the past, but "a prolonged, irreversible cumulative process with effects that become ever more pronounced the longer it goes on".

A CONTEXT FOR THE ELECTRONIC FUTURE

Access to the IISG catalogue through national and international computer networks (InterNet) is planned already. A series of other possible future developments will be sketched first to give the long term context in which the WAAL project should be seen. The short list looks as follows:

- 1) Online Access to IISG & Associated Catalogues (OPACs)
- 2) Virtual Book- & Archive Shelves
- 3) Online Electronic Archives
- 4) Special Database Projects Online
- 5) Electronic Document Distribution (FTP Site)
- 6) LISTSERVER for Electronic Journals & Conferences
- 7) Electronic Catalogue CATALOGORUM (Gopher Social History)
- 8) Electronic Publishing & Bookshop IISG
- 9) Electronic Tours and Special Exhibitions
- 10) Electronic Document Delivery on Demand

The given examples in the more detailed list hereafter are not exhaustive, but just there to give an impression:

- 1) ONLINE ACCESS TO IISG & ASSOCIATED CATALOGUES -OPAC IISG - Catalogue IIAV -Electronic catalogues of small IALHI members and the like who do not have the finances to bring them online.
- 2) VIRTUAL BOOK- & ARCHIVE SHELVES -The covers of brochures/pamphlets, periodicals and certain books can be digitized as an image, these 'digital

convolutes' can have a general description as a group, thus the laborious cataloguing of each small pamphlet can be overcome. The user can browse through these digital facsimile. By organizing these relative small graphic files in a few basic categories, the material can be rearranged according to the selection by the user. By making a fast graphic browser this method will bring a large part of the collection of small pamphlets and periodicals to life again. It is as if you make the settings for the instant creation of a multitude of lovely jumble second hand bookshops, where you can peruse creatively through history. (The older project proposal for the microfilming of the covers of the collection of 40.000 brochures (grouped in geographic and subject areas) of CSD is at the base of this idea).

3) ELECTRONIC ARCHIVES -WAAL (see detailed description elsewhere) - International Press Service (IPS) -As there is a dramatic change from paper to electronic information in the near future may conventional archives and even publications (especially specialized journals) will only exist in electronic form. What seems to be exceptionally now in electronic archiving will become the rule.

4) SPECIAL DATABASE PROJECTS ONLINE -CATOE (Cataloguing and Archiving The Other Europe) -Local to Local Film/Video & TV Network (Next 5 minutes & beyond), a visual database of videos and films -ID Archiv Directories (Verzeichnisse Zeitschriften, Broschuren, Archiven) -Amsterdamse Pamfletten database (Frans Panholzer) These databases will have a double function: on the one hand they support existing persons, groups, institutions with the distribution of information on their (publishing) activities, on the other hand will they form the basis for the acquisition policy of the IISG (this is the way in which ID Archiv is functioning already). The online access to this information has enormous advantages. Changes of addresses and the like can be directly done (often by the persons/groups themselves), the costs for printing and mailing can be over the coming years slowly be cut and budget can be transferred to this more direct way of information exchange.

5) ELECTRONIC DOCUMENT DISTRIBUTION (FTP SITE) -IISG, NEHA, Permuseum catalogues and inventories -Local to Local Network -CATOE database -ID directories -Annual Report and other promotional IISG publications

6) LISTSERVER FOR ELECTRONIC JOURNALS & CONFERENCES - International Review of Social History -IALHI Newsletter -NEHA Bulletin -Bulletin Nederlandse Arbeidersbeweging -Other electronic journals of individuals and groups associated with the IISG -Special electronic conferences related to workgroups, projects, and as preparation and continuation of regular conferences

7) ELECTRONIC CATALOGUE CATALOGORUM (Special GOPHER Social History) -Description of online catalogues and information systems of institutions with related subject areas, whereby the user will automatically be connected through to a chosen electronic catalogue elsewhere

8) ELECTRONIC PUBLISHING & BOOKSHOP IISG -Browsing & buying of paper and electronic books via World Wide Web (including online payment by credit card) -Publishing specialized books and articles in electronic form

9) ELECTRONIC TOURS AND SPECIAL EXHIBITIONS -General and specialized tours via World Wide Web of the IISG and associated institutions -Special visual tour describing the different archives in a visual way. -Exhibits directly related to the work of the IISG will also be represented permanently in electronic form.

10) ELECTRONIC DOCUMENT DELIVERY ON DEMAND -Special service whereby materials that used to be photocopied and send will be delivered in electronic form. By using a digital 'photocopy' machine all copying (including copies from microfilms) can be stored in a special electronic copy archive (a deselection process can be included). By barcoding the original sources before copying a link with the OPAC can be made and later orders of the same document can be automatically supplied (viewed online as digital facsimile) . These digital copies can be delivered in different formats: as data, fax, paper or digital to film. The sending can be realized over InterNet (also for the international routing of faxes). This system could work through the inter-library loan systems (that are more and more moving in this direction). Over the years a fundament for a full electronic archive and library can thus be build.

NON-COMMERCIAL CONNECTIONS

Growth in connectivity means also a growth of collaborations. The IISG is not the only institute facing these coming changes and it must seek collaboration with other institutions on all kind of levels, from the big academic network to the amateur bulletin board system. An interesting phenomena is that the differences between the extremes are less big, the dichotomic distinction, from before, between 'mass' and 'low' media (big newspapers, radio & television) no longer holds. The means of production have moved from the studio/work floor to the home/desktop. Also here the drive towards monopolisation and commercialisation persists. The means of distribution which are partly still non-commercial (connectivity through the InterNet system) are under great pressure and could lose their actual cooperative structure. The position of the IISG should be to strengthen the non-commercial approach and look for the right partners for its electronic projects. It will first of all safe money and it will necessitate a more flexible approach: one must continuously follow the dynamic development 'on the net' to be able to document it.

EXAMPLES TO LEARN FROM

As in the book and paper world of regular archives and libraries, the terrain of 'social history' on itself is already too big to be covered by just one institution. Other partners need to be found, pooling and resource sharing in the electronic world are much more feasible as in the paper world with its slow logistics. Examples might be the former Radio Free Europe Archives now moved to Prague who publish already for a few years most of their material in electronic form, or the Etext political underground archive, a personal initiative of Paul Southworth supported by the University of Michigan. The fast growing group of 'electronic journals' have found already their 'cybrarians' (a.o. Ed Vielmetti, Billy Baron) who

developed, from 1992 on, a joint project: CICNet Electronic Journal Archive. It has been set up by the Committee of Institutional Cooperation (a group of academic institutions in the Mid West of the United States) to collect, preserve and distribute electronic serials (540 electronic journals by June 1993), with a mirror of the archive running from a British academic computer site.

POLY-CENTRAL SYSTEM

The organisational models and experiences used in these projects have many facets that could be used by the IISG for the set up of its network. Without going too much in detail some general guidelines can be drawn already. The concept of a 'central electronic archive' is not on the agenda, such an institution that needs to be created by diplomatic machinery through endless meetings, formulations of principles and definition of vocabulary can only end up in a costly rigid structure that fails to respond to the dynamics of electronic nets. A camouflaged version of this approach by 'decentralizing' the process will still have the same hang ups, as this model tends to give too much importance to the centre and leaves the 'de-centres' with not enough means to fulfil their task. A more logical model seems to be a 'poly-central' set up whereby different, in principle equal, centres collaborate. These centres can be of dissimilar size, each of them will be independently functioning and their association is voluntary. Dedicated and specialized centres need not to be too big, they will be created by direct needs and disappear when the attention shifts to other subject areas. The IISG should associate itself with such centres, or even initiate or support the creation of them on the condition that temporary archives will in time be transferred to a site where long term preservation of electronic data can be secured.

FREWARE TO SET STANDARDS

The proposed development of software for the WAAL project is an embodiment of this idea. It is thought of as 'freeware' that will be distributed for free and were the payment of a small fee once someone decides to use it regularly is voluntary (with the benefits of getting regular updates). The processed of the WAAL software will thus dynamically be tested by a wide group of users that will give feed back and enables the development of a flexible and viable standard.

COUNTERING MAINSTREAM INFORMATION

The amount of information in electronic form seems overwhelming, often the suggestion is made as if with the making of a database, the pressing of a CDROM title the ultimate product in a certain information area has been made. Nothing is further from the truth. When we look more closely and with some scrutiny at for instance the recent wave of CDROM publications the same wide variety of quality differences as on the traditional book market can be observed. There is not one (and only) CDROM on Shakespeare, there are already several on the market and they differ in quality. A lot of the actual electronic information on offer is a recirculation of materials already available in bookform. Most of the electronic products are filled with mainstream information. The main part of the collection tradition of the IISG has been the realm of what can be labelled as

'independent', 'revolutionary', 'dissident', 'radical', or 'alternative'. This tradition will not be changed by the electronic information epoch. The proposed WAAL project is a logical continuation of that policy and also the rough sketch of other possible developments can be seen in this perspective. The IISG find itself (again) in a pioneer role: saving materials for the study of social history before it is too late, not being a vampire institution that sucks historical artifacts out of society only for its academic needs, but an institute that can build up a fantastic collection because it also gives something back to the individuals and movements it studies.

COPY RIGHTS & WRONGS

A major problem to be solved will be the limitations set by the diffuse ownership status of many of the electronic documents. First the relativity of the problem should be addressed: archives and libraries have not been stopped in collecting books or whatever paper documents because they were copyrighted. Photocopying, or other forms of duplicating, of books and articles, for study or personal reference, has now been formally limited (in Dutch law) and some small levy needs to be put on top of each photocopy made in public institutions. An extra charge has been fixed on magnetic tape for the consumer market. This money goes to a semi-government (copy) fund that redistributes it in the cultural arena. The copyright discussion and legislation is mostly approached by the authorship lobby, but there are signs that the other side of the story, free or at least equal access to information, will be put forward as well by the users that still have to find their voice. Instead of totalitarian digital control systems on the reuse of information we need a new sensible approach for the 'cultural collage' epoch we living in whereby almost all products are eclectic reconfigurations from the existing cultural reservoir.

GOOD SERVICES FOR SMALL FEES

When we make a local telephone call we feel not too much cheated by the price we are paying and the information hungry spend already a reasonable part of their income on photocopying that what is not (any more) available on the info marketplace. But we feel barred and frustrated when it comes to speaking over the phone with a far away relative, a free flow of information will soon be followed by the telephone company cutting off your line because we could not pay the bill. The free flow of information, as it takes now place on the net, is felt as an emancipation, liberating us from restrictions that were part of the paper dominated information world. It is clear that the ease and abundance of nowadays copying equipment will make copyright in its actual forms obsolete. As everything seen on screens or heard from speakers can be captured the only effective way of control will be not to show it all. Good service for small fees will be the most realistic approach for managing this coming 'mirror world'.

LEGAL FORMS

A practical approach for the WAAL project could be the forming of an association of WAAL users with a small membership fee and a acceptable price for individual reference copies for non-members. Special deals could be made for big

institutional users on the basis of real usage (downloading of files). Very important is the calculating of the costs of such 'income' generators as they easily can become 'cost breeders'. Ideally, by automating these processes, some income could be generated that should for a big part be redirected to the information providers. An important distinction to be made is the difference between individual use of electronic information and the recirculation in electronic or other form of the same or altered information. A simple standard legal disclaimer procedure can be implemented in the electronic document access or delivery system which will specify the terms. As a starting model the 'disclaimer' of the electronic Gutenberg project can be used.

INFORMATION OR INFOTAINMENT

Attempts should be made to safeguard free access to information, staying in the tradition of the 'biblioteca publica' started in the 15th century (libraries open for the interested laymen), the 18th century enlightenment cumulating in the French Revolution (access to books was not any more the privilege of scientists clergy and nobility, state censorship on books was fading), and the 'public library' system in Great Britain, as an outcome of the early 19th century Reform Movement, inspirational for most of the European public libraries. It is distressing that at the end of the 20th century its precisely this 'public library' system that is under attack. Some are calling for its abolishment: 'commodity society does not need any more this paternalistic up lifting of the common man'. The availability of information seems to be bigger as ever, however the terminology used to describe its faculties points to the continuing relevance of 'public information' facilities: 'information flood', 'information overflow', 'worldwide tangle of highways an byways'. The renewed public library can give guidance in navigation the info-oceans, can show the way around the massive amount of information fused with entertainment ('infotainment') that the mass culture industry is preparing in its fight against obsolescence. In that vision we will get information in a package deal with being exposed to advertisement and directly linked interactive shopping facilities (for a good description of this phenomena see the chapter 'Disinformocracy' in Howard Rheingold's "The virtual community, homesteading on the electronic frontier).

COUNTERBALANCING INEQUALITIES

A minimum and basic standard for the electronic activities of the IISG should be that at least all information about information (catalogues, databases, directories and the like) will always be provided for free, much the same as anybody can now enter the building and use its public facilities for free. The idea of up lifting the disadvantaged has still a topicality, but on a different scale. Not so much up lifting of the 'lower classes' but compensating for the global inequalities. Using the simplified dichotomy of the 'North South divide', it must be realized that much of the hype talk on 'global networks' is utterly false. It are networks from the post-industrialized countries that have a 'valve connection' (that some time leaks a bit of information in the 'wrong' direction) with the other parts of the world. Over here we can be well informed about what happens over there, but this does not mean that these positions are reversible yet. Unequal development, neo-colonialism all

that terminology is still quite relevant, but the chances to do something to change these relations are greater than ever. When we are conscious of the discrepancies we will be able to do something about in a simple way, changing the 'trickle down' to at least a healthy shower of information exchange. The WAAL project has included such reversal techniques by including a CD-ROM exchange with local Bulletin Board Systems in countries with limited access to the InterNet and other networks. In this way people over there can assemble according to their needs information (off line WAAL reader) have it transferred to a CDROM and link these CDROMS to their local network. This should be done for minimal fees and could well be subsidized.

Electronic format levels:

- 1 Host/node
- 2 User(s) groups on 1
- 3 Sites on 2
- 4 Areas/Conferences/Catalogues/Databases on 3
- 5 Messages/Electronic Journals/other Information objects on 4
- 6 Parts of 5

Cataloging models can follow MARC and ISO or anticipate such rules by dividing the description into small enough elements. Some ideas can be found in: - "Cataloging nonbook resources -A How-To-Do-It manual for librarians" by Mary Beth Fecko, Neal-Schuman, New York/London, 1993. There is a special section on cataloging 'Electronic resources' (electronic serials, online catalogs and online databases and other electronic services). The context in which these new media are developing is briefly described in the introduction to this chapter: "With shrinking budgets, and rising prices for books and periodicals, many librarians are considering electronic resources as a viable alternative to print and microforms. ()The idea of a library as a physical structure housing resources will become outmoded, with a shift towards the library acting as a clearinghouse for accessing information.." (p.147). Other sources mentioned are: "Dictionary of data elements for online information resources: MARBI discussion paper No.49" and "Internet-accessible library catalogs & databases" by Art. St. George and Ron Larsen.

Automatic and semi-automatic information retrieval with the help of thesauri and dictionaries in relation to the WAAL project is a realistic option. There are many ongoing researches in this field and some working applications in other sectors of information providers (literature (fiction), medical) that could be used as a model. Some recent information sources are: - "Classification research for knowledge representation and organization -proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, June 24-28, 1991" (FID = Fédération Internationale de Documentation). There are concepts like 'syntagmatic relationships in indexing', 'frame-based index languages', 'knowledge-based indexing', 'thesaurus-based information systems', 'intermediary expert systems', 'syndetic information retrieval system', 'sub-thesauri as part of a metathesaurus', 'chain indexing'. - There have so many failed attempts to construct a 'universal classification system' that we should try in no way to

construct one for the WAAL project. The content of WAAL will be a wide variety of subjects differing also in depth (from popular to scientific).

One of a series of different approaches could be that the information is put into 'folders', the folders in 'boxes', the boxes on 'shelves', the shelves in 'storage rooms', the storage rooms are part of specific 'archive buildings' and the archive buildings are part of the urban environment of an information metropole. Such a spatial metaphor is one way of dealing with the classification problem. The folders, boxes and shelves can easily be transformed or morphed into pots, drawers and cupboards or any other metaphor that will facilitate searching.

Another approach would be that groups of subjects or subject areas will be loosely assembled on 'tables' or in 'tableaus' thus creating a kind of broad rubrics. Such congregation of notions could be represented in the simplest form with words in a diagram, or with the use of a graphic interface with words of different sizes (as in concrete poetry) or words combined with all kinds of pictograms and other iconic materials.

There remains always the option of formal elements like 'where', 'when', 'who' and the more arbitrary 'what'. For the average Email message this information can automatically be subtracted from the 'header'. This information is of course just about who did write or post it, when and from where. It would be most valuable to see if the 'where', 'when' and 'who' could also be subtracted from the text of the message/information item itself (what geographical entity is described in the text, which time period is referred to, who are the 'actors' in the text, and ideally also what is it about, what kind of actions, what kind of motives). Here a constantly updated dictionary/thesaurus of names of persons, organisations and geographical names, supplemented with an objectivised translation of terminology (from natural language words to descriptors) could be used.

The visualisation of the terminology used (this could be a mega-thesaurus that will point to a series of specific/specialized sub-thesauri for different knowledge domains) can be a great tool. Thesauri can be transformed from dull lists and databases to 'screen events' that give a playful discovery of how knowledge is generated and communicated through the use of language. A simple form can be a so called 'graphic display program for thesauri' (a heuristic graph displayer; ref. H. Watanabe, International Journal of man-machine studies, 30 (1989) no.3, p.287-302). Here a central word is put in a box with its relation to other words in other boxes around it connected by lines, the terminology can thus dynamically be represented (Broader, Narrower, Related terms). Tree structures could be helpful and new ways can be found like using newly discovered techniques of 'video browsers', 'video mappers' and fish eye representation of spatial data. There are many developments in what is now called 'knowledge representation' that can at a later stage be implemented.