Long-term Preservation of Electronic Literature

Sabine Schrimpf

Deutsche Nationalbibliothek
Adickesallee 1
60322 Frankfurt
s.schrimpf@d-nb.de

Abstract

Authors have always been looking for new and innovative ways of aesthetical expression and they have been quick to take advantage of the new possibilities, offered by the World Wide Web and the Internet. Today, there is a respectable community – in terms of size and quality – of "digital poets", who publish their texts in the internet and whose literature above all shares one mutual feature: a prominent and crucial use of computer technologies.

Being the cultural heritage for future generations, electronic literature is worth preserving, just as any other form of contemporary literature. Yet, due to its use of interactive and dynamic elements and reliance on the latest technology, contemporary electronic literature is extremely vulnerable and difficult to document. While archives, libraries and museums are still trying to develop preservation strategies for electronic literature, many of the early works have already volatilized. In Germany, two institutions and a cooperative network have joined forces in order to address this challenge concertedly: nestor, the network of expertise in long-term storage of digital resources, the German Literature Archive, and the German National Library.

Background

As opposed to the preservation of research data, no comprehensive, international efforts address the preservation of electronic literature. So far, there have only been a few initiatives in the USA which address the need to preserve electronic literature, such as the Preservation, Archiving, and Dissemination (PAD) project of the Electronic Literature Organization or Archive-it, a collaboration of the Library of Congress, the Electronic Literature Organization and the Internet Archive. These projects are still in their early stages of development.

In Germany, the Deutsche Literaturarchiv Marbach (DLA, German Literature Archive Marbach) is responsible for collecting, archiving and making available contemporary German literature. Primary sources and secondary literature are collected as comprehensively as possible. Since 1997, the DLA has expanded its efforts to electronic literature, beginning with the inclusion of the collecting field "e-journals".

The Deutsche Nationalbibliothek (DNB, German National Library) is the legal deposit library for all German and German-language publications since 1913.

As of 2006, native digital publications are also included in the German legal deposit law.

In cooperation with nestor, the German network of expertise in long-term storage of digital resources, DLA and DNB have taken the initiative to develop a preservation strategy for electronic literature in Germany. In March 2008, they managed to bring together the relevant stakeholders in the German National Library in Frankfurt: authors, archivists, librarians, and legal experts met to discuss the challenges of long-term preservation for electronic literature. [1]

The issues discussed in relation to the collection and preservation of electronic literature included (among others): selection, collection, context, intellectual property rights, and technical issues, the most crucial aspects of which will be presented in the following.

What is electronic literature?

"Electronic literature" is a simple-sounding label for a broad and multifaceted literary field that has evolved from as well as together with the internet. According to the U.S. Electronic Literature Organization, the term refers to "works with important literary aspects that take advantage of the capabilities and contexts provided by the stand-alone or networked computer" [2]. This includes two aspects: A technical and a sociological/communicational aspect. Apart from integrating the technological opportunities which the internet provides, electronic literature makes use of the particular interactive communication structure of the internet.

Electronic literature has (in the most cases) emancipated from the static, linear text narration to which print literature is bound. With the capabilities of the internet, innovative literary forms have developed in recent years: e.g. hypertext fiction, interactive and collaborative fiction, digital and audio-visual poetry, poetry that is generated by computers, or readable computer art installations.

The Electronic Literature Directory [3] introduced for browsing purposes a twofold genre classification: *Genre/Length* vs. *Technique/Genre*. While the former lists the relatively traditional categories Poetry, Fiction, Drama, Nonfiction, the latter distinguishes between the internet-specific literary forms Hypertext, Reader

Collaboration, Other Interaction, Recorded Reading/Performance, Animated Text, and Other Audio/Video/Animation, Prominent Graphics, Generated Text.

What to preserve? – Selection

The examples given in the paragraph above illustrate internet's primary forms of literature. There is no doubt that such works – like other contemporary literature, too – are worth preserving. Besides them, new forms of secondary literature have developed, for example literary journals, listservs, blogs, and wikis, which digital poets utilize in order to organize their community, but also to facilitate collaborative writing processes. Such secondary literature provides the philological context in which electronic literature is being created. It must be preserved to enable future literary studies. Both primary and secondary literature is in the focus of libraries as well as literary archives.

Another category of material is mainly of interest for literary archives and museums: writers' correspondences, manuscripts, personal diaries and calendars are collected and stored for future literary studies. Here, the digital world poses some particular challenges: Instead of or in parallel to a journal, the internet poet may write a blog on his homepage. His correspondence is stored, for example, in his Outlook folder on his hard disk, and so is his calendar. His manuscript may consist of a number of Word files, the newest version always overwriting the previous ones. Consequently,tracking down changes may be impossible.

So, while there is a plentifulness of chat and notes on webpages, and archives can hardly manage to capture all of it, the long-term preservation of relevant resources is seriously endangered.

Libraries, archives and museums need strategies how to cope with this dilemma. They are faced with the question of whether they have the capability to manage and the means to afford the preservation of all of those literature related resources that can be found online. The German Literature Archive and the German National Library are still refining their selection profiles with regard to electronic online literature. While each institution will need to define its own specific selection profile, it may also be feasible for the community as a whole to agree on certain common selection criteria.

Apart from selection criteria for electronic literature, literary archives need a policy how to deal with the new kind of "literary estates" that comes in on hard disks, CD-ROMs and USB-Sticks instead of boxes full of inscribed paper.

Harvesting the literary web – Collection

In order to guarantee long term availability of electronic literature resources, libraries and archives store local copies on their servers, respectively repositories. To this end, the selected resources can be harvested from the internet once or in periodic intervals. Archives and libraries can already choose among a range of existing software solutions. According to a predefined harvesting policy, a list of predefined URLs is downloaded to the library's or archive's server automatically.

In contrast to printed works, where copies of published editions are collected, the collection of electronic literature requires new agreements. The definition of "edition" is challenged by the technological progress. The conventions of the print era can most likely be applied to electronic journals, which are usually issued periodically. Every new, completed issue can be collected.

It becomes more complicated with independent literary works, when borders between different editions are blurred because a version is constantly changed, refreshed, complemented etc. Where does one draw a line between different versions with regard to form and content, where does a new edition begin?

The same applies to authors' homepages, which are regularly updated, and to ever active blogs. Shall every change be reported and each revised version be transformed into an archival object? Shall objects be harvested weekly? Monthly? Daily? Shall old versions be overwritten or preserved together with the latest version? With regard to such questions, archivists and librarians, if necessary together with authors, need to reach sensible agreements.

At the Frankfurt workshop, the authors agreed with the archivists and librarians on a (rather traditional) approach to collect authors' homepages and primary works, literary magazines and dialogue forms like blogs.

The German Literature Archive already collects, indexes, and stores such material in a pilot operation. To this end, it cooperates with the state library service centre of Baden-Württemberg (BSZ) and shares the BSZ Online Archive. During the pilot phase, the German Literature Archive contacted the rights owners of selected literary resources and asked for their permission for harvesting their sites.

Only those sites for which permission was received were then downloaded to the BSZ Online Archive, indexed, and made accessible via the Online Catalogue of the German Literature Archive.

Medienart	[Zeitschrift Internetquelle]
Titel	Electronic journal Literatur primär / Hrsg.: Franz Krahberger
Weitere Titel	e.journal Literatur primär
Person	Krahberger, Franz [Herausgeber]
Impressum	Wien
Ersch.Verlauf	1996 -
Fußnoten	"Das Electronic Journal bietet Essays und theoretische Beiträge zum Thema Neue Mediale Kommunikation an, ist jedoch auch offen gegenüber allgemeinen künstlerischen, literarisc Äußerungen, Absichten, Einsichten und Konzepten." [Information des Anbieters] Die seit 19 Zeitschrift enthält Beiträge vorwiegend österreichischer Literatur, Besprechungen und umfa weiterführende Linklisten
Sprache	Deutsch
Land	Österreich
ISSN	1026-0293
Kette	S10.10.1.3.2 / Deutsche Literatur 1990- / Einzelne literarische Zeitschriften
URL	http://www.ejournal.at/ http://ejournal.thing.at/ http://naxos.bsz-bw.de/boaDla/downloads/frei/78/0/index.html
Bemerkung zur URL	Verlag Verlag Langzeitarchivierung

Shapshot of a catalogue entry of an electronic journal at the German Literature Archive with reference to the original source on the internet and to the archival copy at BSZ Online Archive.

The collection procedure of the German National Library is not limited to electronic literature, but includes it. The library's collection field of online publications covers all text, image and sound-based works which are made available in public networks.

Suitable procedures for a large-scale collection, cataloguing and preservation of online publications are gradually developed and continually tested. For the time being, the German National Library uses an automatic collection procedure, which involves submission via registration form or OAI harvesting. While OAI harvesting is mostly used by large publishers, submission via registration is feasible for small publishers or independent authors, who release a manageable number of individual works online. It is probably the preferred method for electronic literature. When delivering via registration form, the author or publisher registers his online publication via a web form at the German National Library and manually delivers his publication to the library's deposit server.

Online publications are then catalogued by library staff, documented in the German National Bibliography and made accessible via the Online Catalogue of the German National Library. Persistent identifiers (URNs) ensure permanent addressing and long-term citability of online publications.

KATALOG DER DEUTSCHEN NATIONALBIBLIOTHEK

← Zurück zur Trefferliste

Treffer 10 von 154

Link zu dieser Seite	http://d-nb.info/975467271
Titel	Hypertextual fiction on the Internet: a structural and narratological analysis [Elektronische Ressource] / vorgelegt von Roman Zenner
Verfasser	Zenner, Roman
Erscheinungsjahr	2005
Umfang/Format	Online-Ressource
Anmerkungen	Langzeitarchivierung gewährleistet
Hochschulschrift	Aachen, Techn. Hochsch., Diss., 2005
Persistent Identifier	urn:nbn:de:hbz:82-opus-11023
URL	Archivserver der Deutschen Nationalbibliothek http://sylvester.bth.rwth-aachen.de/diss... http://sylvester.bth.rwth-aachen.de/diss... kostenfrei
Schlagwörter	Internetliteratur ; Erzähltheorie ; Online-Publikation
DDC-Notation	802.85 [DDC22ger]
Sachgruppe	800 Literatur, Rhetorik, Literaturwissenschaft

Shapshot of a catalogue entry of an electronic resource at the German National Library with reference to the original source on the internet and to the archival copy at the library's archival server.

Context

The creation of electronic literature involves hard- and software, a number of applications, and a set of technologies. The preservation of electronic literature implies the preservation of all of these technical components, so to speak of the technical "environment" under which it was created – or the replacement of this technical "environment", for example per emulation.

A similar challenge is to preserve the sociological environment of an electronic document; this means the context in which an electronic text is embedded. Electronic documents may for example refer to other documents via hyperlinks, to images embedded in the text, to comments added to it etc. When collecting electronic literature, a decision has to be made, how much of a text's context ought to be captured.

Experiences with harvesting tools reveal that it is not easy to capture the context, in which an electronic document was originally published. The most obvious problem: It is in the nature of selective harvesting that only the desired pages are harvested. Consequently, external links are deactivated. So the original "environment" of a web page is lost. Another problem consists of externally generated elements, like images, a calendar function or even advertisement.

To preserve dynamic elements is another specific challenge. Of all things, dynamic is a constitutive element of electronic literature. Beat Suter introduced a model of the "development of electronic writing" [4], in which he distinguishes four phases of electronic writing, marked by increasing use of dynamic and interactive elements:

1. flexible text
2. hypertext
3. networked writing
4. "pending writing"

Flexible text means linear text that is simply generated on a computer (e.g. an electronic manuscript that is used for generating the printer's copy). The flexibility is characterised by the opportunity to "cut and paste" text blocks, to move lengthy passages within one document or among different documents. The publishing format for flexible text is typically PDF.

Hypertext is as well first generated on a computer, but afterwards converted into HTML and made available via the internet. Hypertext utilizes hyperlinks, which interconnect a number of text fragments. Hypertext is no longer linear, because the reader can browse through the text fragments on individual pathways. Moreover, the author has the possibility to link from his text to external references, thus establishing intertextual references.

The next stage, networked writing, implies dynamic or interactive features, like commenting and collaborative writing processes. Typically, several authors team up for a networked writing project and share a mutual working space, often in public. So the text is made available to readers well ahead of its completion. The reader can participate in the writing process by commenting or by getting involved in discussions with the authors.

"Pending writing" is the most elaborate form of current electronic literature. It is characterized by a dynamic, interactive writing process and the use of various technological features such as computer-assisted text generation or mechanical, arbitrary organization of text. "Pending" works are always in a state of incompleteness, because authors and readers constantly interfere with the text, modify or manipulate it, add own texts or links to external resources. Thus, the context itself becomes a constituting feature of a pending text.

Technical Issues

The technical complexity of preserving electronic literature increases with each phase described above. Objects of the first two phases – text files and hypertext - can relatively easily be preserved as single objects. The third phase, networked writing, produces multilayered objects. For preservation, such complex objects have to be fragmented into a number of single objects. Emulation appears to be a workable solution, too. The preservation of objects from phase four, pending writing, requires very elaborate emulators.

The fundamental claim of the German Literature Archive and the German National Library is to document and preserve the digital avant-garde literature as comprehensively as possible. This implies that on the one hand neither of the two institutions wants to exclude any data formats from its collections. On the other hand, with an increasing amount of archived formats, the complexity of preservation measures increases manifoldly.

The participants of the Frankfurt workshop agreed that the best way seems to be to involve authors in the preservation process of their more complex works in order to document the entire compilation environment. Especially in order to successfully and adequately preserve the more complex networked works and dynamic forms of the "pending writing" process, the collaboration of authors, archivists, and librarians appears necessary.

Intellectual Property Rights

Authors and archivists as well as librarians are troubled by many unresolved legal questions relating to the preservation of electronic literature: Under what circumstances are archives and libraries allowed to create copies of the archived objects? Must copies be authorised by authors? Is the owner of a literary blog allowed to grant intellectual property rights on all entries in his blog to the archive? What about the rights of third parties like web designers?

The German National Library collects electronic literature under national legal deposit legislation. It does not have to request the right holders' consent before harvesting their websites. The German Literature Archive and other archiving projects are faced with the necessity to ask permission every single time they want to collect a new web resource.

The fact that national legal deposit libraries are allowed to collect web resources does not mean that they are entitled to make them automatically available. Like other libraries, the German National Library has to negotiate access conditions individually with the right holders.

In order to simplify the resolution of such intellectual copyright issues, the participants of the Frankfurt workshop advocated the adoption of public licences for electronic literature, of which Creative Commons might be the most widely known.

Perspective

In a field as multifaceted as electronic literature, the involvement of authors in preservation processes seems strongly advisable. The potential synergy that lies in a common consent with regard to selection criteria, the process of collection building, legal solutions, and the technical framework, to mention the most crucial aspects, increases the likelihood that today's electronic literature will be preserved for future generations. The Frankfurt workshop can be seen as a first step in the right direction of such collaboration for the German-speaking area. The participants are determined not only to continue but even to extend their collaboration: Further steps are envisaged, such as the compilation of a "preservation guide for authors".

References

[1] Documentation of the workshop is available from URL:
http://www.langzeitarchivierung.de/modules.php?op=modload&name=Downloads&file=index&req=viewsdownload&sid=29
[2] Electronic Literature Organization. URL: http://eliterature.org/about/
[3] Electronic Literature Directory. URL: http://directory.eliterature.org/index.php
[4] Beat Suter: Das Neue Schreiben 1.0. (2004) URL: http://www.netzliteratur.net/suter/dasneueschreiben1.html.