Chinese Internet Literature: Preserving Born-Digital Literary Content and Fighting Web Piracy

C. David Hickey
University of Florida

Follow this and additional works at: http://scholarsarchive.byu.edu/jeal
Part of the Chinese Studies Commons, and the Digital Humanities Commons

BYU ScholarsArchive Citation
Available at: http://scholarsarchive.byu.edu/jeal/vol2015/iss161/7

This Article is brought to you for free and open access by the All Journals at BYU ScholarsArchive. It has been accepted for inclusion in Journal of East Asian Libraries by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu.
Chinese Internet Literature: Preserving Born-Digital Literary Content and Fighting Web Piracy

C. David Hickey, University of Florida

Abstract

This paper undertakes to define Chinese internet literature (CIL) and discuss the unique circumstances that make the subject of CIL content preservation so problematic. Copyright infringement and text corruption, in which the text is changed or adapted without author’s permission, are major problems in China. Text corruption mostly results from censorship or web piracy. A recently-developed type of anti-piracy scheme called “literary works fingerprinting” can protect the content of the works themselves, guarding them against plagiarism. If as many of the original CIL documents as possible are consistently archived and digitally fingerprinted, then the born-digital works that remain online will not disappear and cannot be altered, nor can they be easily copied on other sites.

Introduction

What exactly is Chinese Internet Literature (henceforth referred to as CIL)? This paper offers a working definition, as well as a survey of its unique historical evolution. CIL’s lack of preservation guarantees and vulnerability to web piracy may result in corrupted texts that undermine the authors’ original intentions.

After a review of the relevant literature in Chinese and English, a discussion of copyright efforts and plagiarism concerns will lead into a description of the anti-piracy technique known as “literary works fingerprinting,” which is now available to guarantee long-term preservation for works originally created in online editions. These digital literary creations may or may not ever be transferred to print or another physical medium.

Background

The novel Chengdu, Please Forget Me Tonight by Murong Xuecun 慕容 雪村 first appeared sometime in 2002 in serial form on a Guangzhou internet forum, as noted by Edward Wong in a November 7, 2011 New York Times article. After network reader buzz caught the attention of a book publisher, a censored version of the work appeared in print with something like 10,000 characters missing. In protest, the author then posted online his complete uncensored manuscript. This formed the basis later for a complete print edition by another publishing house, ensuring that this digitally-born work would be preserved uncut and true to the author’s original intentions.
Several aspects of the CIL phenomenon known in China as “wang luo wen xue 网络文学” are evident in this example. First, writers online have relative freedom from censorship as opposed to those in print. Second, the immense popularity of network literature in general inspires a market-driven natural selection of particular works for further publication in another medium like print.

CIL has contributed to the transformation of Chinese society at fundamental levels, including that of daily speech. The quantity of web literary creations has led to enormous diversification of literary expression, with genres like romance, martial arts, fantasy and science fiction proliferating.¹

At the end of 2012, 233 million CIL users, a full 41% of net users (“netizens”), were tallied as spending more than 100 million yuan, or 16 million US dollars, in commercial literature websites,² driving home the enormity of the mainstream CIL presence in contemporary People’s Republic of China (PRC). This online market enterprise has become so important that China’s Nobel Laureate Mo Yan accepted the honorary role of principal at “Wang Luo Wen Xue University,” a distance education enterprise set up in 2013 to train upwards of 10,000 aspiring writers hoping to succeed in the competitive world of authorship on major literary sites.³

Readers, authors, website publishers and socio-cultural critics frequently call for an increase in literary works of quality and originality. The reality is that the number of popular hits has become key in determining the lasting meaning and influence of web literary pieces. Under this popularity criterion, such writing is rightly regarded as “popular culture” and worthy of academic library collection development, with its derived components of access, description and long-term preservation.

In terms of online literature being completely absorbed in and by extension influencing academic library collection development policies and procedures, there is the particular handicap of ephemerality, unlike print and audio-visual media platforms. Literary publications only existing on the web may not be passed on to posterity without extreme diligence taken in both personal and institutional preservation efforts. Personal archiving can rely on tools such as the Internet Archive Wayback Machine (http://web.archive.org), which collects snapshots of web pages from around the world for permanent preservation, and on download mechanisms such as citation repositories and software-managed storage libraries via Zotero, Mendeley and other commercial content management programs.⁴

Institutional responses to content preservation and anti-piracy protection of born-digital works are increasingly common in the United States: an example would be the academic library Institutional Repositories that archive, preserve
and protect the works of the institutions’ communities, including works related to the digital humanities. While some campus authors opt not to place original literary compositions in the Institutional Repository, preferring to keep these in embargo until they can be published elsewhere and thus eligible for remuneration, the institutional mechanism for protecting against copyright infringement and preserving the content of such digital creations is in place.

China actually has only a handful of Institutional Repositories at key universities. These preserve mostly “STEM”-oriented (Science, Technology, Engineering and Mathematics) works by faculty and researchers, with some of the items unavailable in full-text. Digital humanities has not yet come into any sort of maturity in China. In fact, Michel Hockx, one of the major authorities on CIL outside China, illustrates this lack of web content archiving and preservation by Chinese institutions by singling out the National Library of China as of March 2014 for not keeping a general archive of websites, unlike other national libraries like the British Library and the Danish National Library which have legal permission to deposit complete contents of UK and Danish websites into their respective web archives.

**Literature Review**

In China there has been a groundswell of journalistic and academic interest in CIL in the last decade. Even the official state newspaper *People’s Daily* began coverage of “wang luo wen xue” as early as 1999 with an announcement about a CIL prize competition, followed by an increasing number of articles through the years depicting mundane as well as serious elements of CIL. The latter is exemplified by a strong admonition in April 2011 for CIL commercial websites to, among other things, use appropriate market analysis techniques to institute and manage fee-paying provisions, publish more original creations rather than a proliferation of works similar to the ones that are already popular, and guard against piracy.

The proliferation of Chinese academic writings on the topic of net literature reflects the fact that, in a culture that celebrates and fervently debates all aspects of its literary ethos, this online manifestation has been highlighted as a rich and viable subject for teaching and inquiry. In point of fact, “Chinese scholars usually engage in theoretical discussions of the ontology, aesthetics and sociology of web literature.” Sources in the surveyed CIL scholarly literature in PRC do not ever bring up government interference in the online publishing of original literary creations that exceed its tolerance levels in politically and socially sensitive areas.

These censorship concerns, however, have dominated many English journalistic and academic writings on CIL. A 2005 *China Quarterly* piece by Michel Hockx was an early study providing evidence of extensive sudden
disappearances of literature web pages. Hockx in his 2015 book *Internet Literature in China* affirmed that he has actually tried in most of his writings on CIL to resist the trend of talking about only what could not appear on the internet in China, but instead focus on what could. While Hockx admitted to being “intrigued by how difficult it is to study [literary censorship] objectively,” most Western scholarly writings on CIL have been consumed with describing and analyzing the effects of political censorship. Indeed, state control over popular culture has been manifested from the outset through overt pulling-the-plug on websites, as well as the government exerting pressure on writers and cultural brokers to self-censor, with both entities further motivated to do so by commercial incentives.

In his seminal chapter in the 2010 *From Woodblocks to the Internet*, Guobing Yang offers a nuanced reflection, based on his experience in visiting and studying CIL websites, that “overall, political censorship is not as much of a concern on literary sites as in BBS [bulletin board system] forums of more general interest,” and by extension, in personal web pages like blogs. However to illustrate one type of political control effort, he described *Qidian* ["Starting Point"] *Chinese Online* (*Qidian* for short) and other major literary websites as of 2008 displaying the Internet Police logo as well as their own permit numbers. A reader, who already is required to register online in order to post any message on that site, can report to the authorities via this link any information perceived as “unhealthy” on the site, including that in older works in the archive.

Apropos further tightening of state control over online activities, in June 2015 the Singapore *United Daily Website* reported that a Shanghai *New People’s Evening News* reporter learned the following from the Number Three Section of the Public Security Bureau (PSB): in addition to a citizen ID information system already in place, a “PSB Netizen ID Differentiation System” is being rolled out, with digital identity surrogates (“eIDs”) being issued. The “eID” will ostensibly help a netizen avoid disclosure of personal information in online marketing situations, but in reality it would not only add PSB analytical capabilities, but also strengthen the tracking control of relevant departments toward the online activities of that individual. This is evidently part of the sweeping new cybersecurity draft law unveiled July 6, 2015. When implemented, the law will reinforce the loss of anonymity and subsequent vulnerability that literary authors and readers face when forced into real name registration on such websites as *Qidian*.

**Chinese Internet Literature Defined**

What exactly is meant by Chinese Internet Literature? Ouyang You-chuan’s definition of CIL is posited as the best available, given the numerous attempts to encapsulate what it is when scholars and lay people alike discuss the
phenomenon. Distinguished professor-emeritus in the College of Liberal Arts at Central South University in Changsha, Ouyang is considered preeminent in the field of digital culture studies due to his voluminous and influential writings on CIL. He founded and currently heads his university’s Base for Internet Literature Research, a major incubator for many subsequent CIL specialists.

Three layers constitute the definition: the broad sense is that the web is the disseminating carrier, the literal sense is that the web is the home for original creations, regardless of genre, and the narrow sense is that the web is CIL’s flesh-and-blood. In other words, it is literature propagated over the web, digitally born and first publicly released on the web; moreover, it is capable of passing through hypertext links and merging with multimedia. Lastly it relies on the internet for its very existence. It is a literature that thrives on the interaction of the web and its readers, and they must be able to browse and appreciate it online first and foremost.\(^\text{14}\)

This definition transcends the common trope that CIL is just traditional literature published in a non-traditional format, an assumption that is based on the premise that online works display the same linear mode as print works. As Ouyang states, hypertext links and multimedia writing can occur as part of web literature, and this potentiality is characteristic in defining CIL. To say, as some do, that CIL includes “all web-based writings that are viewed as literature by their authors or readers, regardless of genre,”\(^\text{15}\) brings in a subjective element to which Ouyang’s definition does not have to resort.

**CIL Historical Evolution and Current Status**

The unique historical evolution of CIL based on political and social circumstances is what makes it so different from the original net literature of other countries. Hockx states: “The particular type of [state] censorship practised in China….would support the suggestion that there is such a thing as ‘PRC web culture’ despite the fact that the world-wide web is supposed to work against such nation-based distinctions.”\(^\text{16}\)

The internet spread quickly in China starting in the late 1990s, with steadily increasing literary production and readership on personal sites and bulletin boards by users from the rising middle class, spurred by the mostly-urban affluence from China’s economic boom. Needless to say, these writers and readers found this outlet a handy platform for free expression under pseudonyms, since the state control of the net in the early years was much looser and more irregular.

After commercial literature websites came to be the main CIL purveyors in the mid-2000s, the dynamics completely changed for both “free expression” and “free access”. Writers used to the relaxed openness characterized as
grassroots writing-for-writing’s sake were afforded fewer opportunities for free expression than they had been within the former less-structured web literature platforms. Many readers accustomed to looking online at literature without paying had to switch to web subscriptions and other fee-paying models. Individual literary blogs and the Twitter-like “Weibo” social media accounts continue to this day as some of the non-commercially-driven places where literature can also appear on the internet.

Hockx depicts the current status of CIL as of March 2014 as most likely in decline, “with the number of users of online [literature] sites growing more slowly than the number of Internet users as a whole,...it may well turn out...that the...roughly 2000-2013 [period of years]...has seen the rise, climax and gradual demise of a unique form of Chinese-language cultural creativity.”\textsuperscript{17} He anchors his “CIL time-capsule” characterization on the increasing popularity of mobile platform apps to allow downloading of literature via the Internet. “In future, websites are likely to function more as places where one finds information about writers and their publications and links to where these can be downloaded, rather than as the actual spaces where reading and discussion take place.”\textsuperscript{18} Given the 2015 heightened political measures to control all aspects of the web in China, Hockx’s positing the years 2000-2013 as the apex of Chinese-literature-on-the-web is even more plausible.

**Content Preservation: **\textit{Qidian} as an Example

Among commercial CIL websites that currently exist and still manage to attract and hold readers, \textit{Qidian} can be used as an example of the lack of preservation guarantees by a web platform operation for CIL works that are not transferred to print or another physical medium. \textit{Qidian} pioneered the profit model among commercial sites via its online fee-based reader service. As a prominent member of the \textit{Shengda} [“Grand”] Literary Network Development Company (hereafter referred to as \textit{Shengda}), \textit{Qidian} has led the way in expediting the transfer of many of its popular works into print, e-books, games, films, live theater, and television.

Along the way \textit{Qidian} has come to embody the public perception that such profit-driven enterprises, closely following and acting on the number of online hits for individual works, are redundant and lack quality. The CIL provider takes a work’s popularity into account and churns out repeatedly in cookie-cutter fashion that same product in form and content.

A class system is thus set up in which the less popular literary works are ignored to the point that they may ultimately disappear. One could argue that the number of hits a particular work gets is a direct form of peer review by the masses; by this logic, a high-quality online fiction or poetry contribution
would be unlikely to be ignored by so many CIL readers. However, in picturing the mix of literary dregs and the occasional unappreciated gem comprising this category of “remaining-digital due to unpopularity,” an audience of CIL readers and site managers may neglect to evaluate and determine early enough the latter’s literary excellence. For example, pieces that constitute early efforts by authors who later become celebrated and in demand as academic research topics, as well as often-ignored experimental works, are potentially lost forever within this paradigm, simply because they do not get enough hits.\textsuperscript{19}

\textbf{CIL Copyright Protection: Shengda’s Landmark Legal Actions}

Original literary works online, no matter whether they may or may not ever be transferred to print or another physical medium, are vulnerable to copyright infringement involving the outright stealing of all or parts of someone else’s web literature product and passing it off as one’s own. As a result CIL writers and site managers have enlisted digital copyright provisions to protect their intellectual property, guarantee that profits go to the right people, ease transferability to other media, and facilitate foreign sales, all under that same copyright umbrella.

The existence of internet copyright violations is global, and no one has yet found the ideal scheme to counter this online version of an age-old crime. Literary creators and original creation management entities protected by copyright agreements certainly carry the force of law in confronting and prosecuting violators, as exemplified in the 2009 - 2012 successful copyright infringement lawsuit process that \textit{Shengda} carried out against the ubiquitous \textit{Baidu} (literally "100 degrees") \textit{Library} search engine utility and internet portal (known simply as \textit{Baidu}).

In 2010 \textit{Shengda} asked \textit{Baidu} to delete illegal copies; in 2011 a Shanghai court ruled that \textit{Baidu} must stop such violations at once and pay \textit{Shengda} reparation for related economic losses. Even after 2011, \textit{Shengda}’s statistics confirmed a \textit{Baidu} search for popular CIL titles from its legitimate affiliate sites like \textit{Qidian} yielded results rife with links violating copyright. By using a copyright identity confirmation process, \textit{Shengda} followed a trail to more than ten thousand pirate sites and a million pirated links. \textit{Shengda} calculated that under such circumstances there has been each year a four- to six-billion yuan (six- to nine-hundred million dollars) loss to CIL creators.\textsuperscript{20}

In 2012 a Beijing court ordered \textit{Baidu} to compensate wronged writers who had filed a separate law suit, a roster that included the celebrated novelist Jia Pingwa, the outspoken blogger Han Han and the afore-mentioned Murong Xuecun. When Murong originally posted his Chengdu novel in online installments, he had to deal with readers trying to write their own versions of
subsequent chapters, and inserting those after whatever latest chapter he had written and published online. While he was bemused by this early CIL consumer presumptuousness, he later was not so light-hearted about flagrant piracy of this and later online literary creations.

It is instructive to note that Baidu only furnished the platform that others operated within: in and of itself it never directly implemented copyright violations. However, by enabling hordes of netizens to download and share tons of illegal copies for free, it became a “pirate harbor”\textsuperscript{21} whose open-door actions and “ignoring-the-obvious” inactions incurred legal scrutiny and condemnation.

These landmark cases illustrate the seriousness with which the authorities dealt with these violations against individual and management copyright holders. Exasperation increased all around at not being able to control plagiarizing behavior against the born-digital works themselves, beyond the official, yet often easily ignored, digital copyright protection afforded the writers. Safeguarding the original integrity of as many literary works as possible from piracy assumed importance as a corollary to author copyright protection.

**Fighting Web Piracy: The Case of Digital Fingerprinting**

Under this legal and public critical scrutiny, a chastened Baidu therefore felt it had to take action in terms of anti-piracy provisions. One particular anti-theft tool, developed by Baidu programmers\textsuperscript{22} and called “wen xue zhi wen” in Chinese, has been adopted since 2012 by a bevy of commercial CIL and other interested sites to identify illegal copies and plagiarism. Using a DNA identification system, or “digital fingerprinting,” to track specific works, Baidu itself verifies and shuts down the uploading of pieces violating copyright. With the application of this mechanism, not only will illegal works that netizens upload be systematically deleted, but future attempts to upload such works are automatically rejected.\textsuperscript{23}

Literary work fingerprinting establishes the presence of parallel or derivative content in a suspect work; this differs from another anti-piracy tool, “digital watermarking,” which adds artificial information into the original text in order to protect that work, as long as the watermark is not erased or modified. Specifically, digital fingerprinting gathers the patterns of textual content like words, sentences and paragraphs to make a unique values signature characterizing that literary work. This is done through so-called “hashing,” or cryptographic identifier deployment, that converts the data to digital integer values. Later, stored in an operationally central database, the fingerprint can be used to compare whether another literary segment is identical with, or closely mimicking, any segment within a given original source set.\textsuperscript{24}
The fingerprint inquiry can be done any time and involves a conveniently fast extraction. The most telling characteristic is that the original content of the work is not changed whenever a literary segment is manipulated in resizing, compression or other processing operations. This technological tool thus shows great promise for protecting born-digital literary content indefinitely.

In reality only the popular writings on Shengda commercial literary sites evidently undergo fingerprinting. Other online forums that carry original CIL like independent literary sites and blogs most likely miss out entirely due to lack of research-and-development capabilities and funding. However, the digital works anti-piracy movement is continuing to gather momentum, with the application of such anti-plagiarism schemes as literary works fingerprinting increasing as the cost for such additional digital processing lessens.

**Conclusion**

CIL content preservation has been referred to in this paper as a problematic issue rising out of the situation where commercial unpopularity of born-digital remain-digital works leads to neglect and possible disappearance. Whether involving politically correct fiction and poetry, or works flying under the censor radar that approach and even occasionally cross the line into politically, socially or culturally taboo subject areas, the sustainability of online literary publications seems subject to the vicissitudes of crowd appeal, author and website manager perceptiveness and perseverance, and economic considerations.

The original intentions of such CIL authors as Murong Xuecun can be undermined when their texts are corrupted without their permission by censorship or plagiarism, all because of the lack of protection for the content of their born-digital literary works. Interestingly enough, despite Murong’s opposition to copyright abuse as manifested by his participation in the 2012 court action, he insisted during that same period that “a relaxed and free [internet writing] environment is more important than royalties.” He bucked the tide in this pronouncement, since most CIL authors would place theft or distortion of their intellectual property high in their list of professional concerns, since these infringements come down to siphoning off yuan that would otherwise come to the writers and other copyright holders.

Internet literary efforts across the world, not just in China, need to be consistently preserved, just as the born-and-resident-digital intellectual products of such realms as STEM are, whether in commercial website storage space or in digital archives maintained by academic overseers or under other individual or organizational auspices. Consistent archiving of original CIL
documents by the agents of platforms where they appear---for example, by bloggers, social media users or facilitators/managers of commercial literature sites---supplements the efforts of individual readers and the few institutional entities attempting to capture and preserve all these literary works, and is thus highly recommended. However, the so-called “protectors of the creative literary output,” whether the National Library of China, other libraries/archives or independent agencies, certainly need to step up more to describe, facilitate access to and guarantee long-term preservation of original web literary creations.

Digital literary work fingerprinting can preserve born-digital literary content as well as fight web piracy, since it offers a way for a web archive to lock in the original work’s content in a database at the moment it is stored, so that the essential literary elements will not disappear or be altered. This anti-piracy tool thus should be recognized as an effective preservation device and employed as such.

________________________

NOTES


5 May 27, 2014 onsite interview with Changping Hu 胡昌平, professor of digitized information resources management and services at Wuhan University College of Information Management.


15 Yang Guobin, “Chinese Internet Literature and the Changing Field of Print Culture,” 333.


20 Ouyang You-chuan 欧阳友权, “Dang xia wang luo wen xue de shi ge guan jian ci 当下网络文学的十个关键词 [Ten Key Words in Contemporary Network Literature],” *Qiu shi xue kan* 求是学刊 40.3 (May 2013): 128.


22 Confirmed in Wuhan May 27, 2014 by a former employee of Baidu, now a graduate student at Wuhan University College of Information Management.

23 Yang Yuting 杨昱婷, *Wang luo wen xue ji qi ban quan bao hu mo shi yan jiu*, 27.


REFERENCES


