Urban Renewal: Some Lessons for HyperCities from the Preserving Virtual Worlds Project*

Matthew Kirschenbaum

This work is produced by The Connexions Project and licensed under the
Creative Commons Attribution License †

External Image

Please see:
http://rup.rice.edu/image/cnx_images/shapes-buybutton.jpg

Seeing the world in a grain of sand is a familiar fantasy for technorati. In 1992, David Gelernter published a book called *Mirror Worlds*, in which he describes, for a lay audience, what it will take to (as his subtitle has it) "put the universe in a shoebox." Gelernter, in addition to being a Yale professor and (only one year after) Unabomber victim, is also an unreconstructed Platonist; he blithely throws around the conceit of a mirror world—"some huge institution's moving, true-to-life mirror image trapped inside a computer"—as if two millennia of philosophical footnoting, culminating in, say, Richard Rorty's *Philosophy and the Mirror of Nature*, hadn't happened. For Gelernter that's okay: the mirror world is simply the natural culmination of a march of technological progress, replacing the metaphor of computers as giant brains with the crystal ball, or palantir. Hands-on, containable (shoeboxes fit easily under the bed), but containing multitudes.

Neal Stephenson was also there, at pretty much the same time. *Snow Crash* (1992) is best remembered for its anticipation of Second Life with the MetaVerse, but in several extended sequences the book's hero-protagonist Hiro Protagonist consults a palm-sized tool called, well, "Earth": "It looks exactly like the earth would look from a point in geosynchronous orbit directly above L.A., complete with weather systems—vast spinning galaxies of clouds, hovering just above the surface of the globe, casting gray shadows on the oceans—and polar ice caps, fading and fragmenting into the sea" (109). The abstraction of information into a visual, spatial, and above all urban representation is, of course, *the* signature cyberpunk trope, back to the granddaddy of them all, William Gibson's lines of light, "like city lights receding." But what Gelernter and Stephenson both have in common is an emphasis on the geographical and the miniature. Both of them, of course, are anticipating the massive contemporary industry of GIS, which has culminated in Google Earth and its rivals, competitor products such as Microsoft's Bing and NASA's World Wind, putting something very much like Stephenson's spectral spinning sphere a mere 10 or 15 MB download away from any desktop.

We can continue to multiply origin stories. There's Buckminster Fuller, for example, and his 1960s Geospace concept, a gigantic globe wired up to receive input from databanks all over the world. But virtual earths and giant electro-mechanical orbs are at best a partial genealogy for Todd Presner and his team's remarkable work on HyperCities, which, as Presner notes, owes as much to the traditions of cultural mapping that emerge from Benjamin's *Arcades* as the panoptical fantasies of Fuller, Stephenson, and Gelernter. In his

*Version 1.2: May 14, 2010 6:06 am -0500
†http://creativecommons.org/licenses/by/3.0/

paper, Presner succinctly catalogs what sets HyperCities apart from more general tools like Google Earth: that it foregrounds *temporal* browsing as a fundamental aspect of the user experience; that the content privileges the interests of humanities scholars, exposing the cultural and historical transformation of space (as opposed to, say, the location of the nearest In-and-Out Burger); and finally, that the entire project is explicitly conceived as a platform for experiments in new forms of scholarly publishing. This last is what I take to be the key feature for purposes of discussion at this meeting.

Indeed, at the heart of the whole enterprise of HyperCities is, it seems to me, the project of curation. The voiceover (one presumes it is Presner) in the video demonstration of the Tehran Election Protests offers the following at 2:20: "This particular project is a massive digital curation project, taking existing resources that are found on the Internet, such as Twitter photos, Flickr photos, YouTube videos, and other documentary reports, and putting it all together, marking it with both a time and a space markup, within a collection that right now is one of the largest existing documents of the election protests in Iran."[1] Curation is also much celebrated in the Digital Humanities 2.0 Manifesto that Presner and colleagues co-authored at UCLA, making the point that digital humanities fundamentally reshapes the relationship between scholarship and curation, with the two activities becoming mutually informing and reinforcing. "Curation," the authors of the Manifesto go on to note, "also has a healthy modesty: it does not insist on an ever more impossible mastery of the all; it embraces the tactility and mutability of local knowledge, and eschews disembodied Theory in favor of the nitty-gritty of imagescapes and objecthood."[2] Nor is the remediation of curation a project limited to HyperCities; one can find other invocations throughout the digital humanities landscape, for example in the ambitious Collex tool being developed by NINES, which allows users "to collect, annotate, and tag online objects and to repurpose them in illustrated, interlinked essays or exhibits."[3]

Much of what one reads in the DH 2.0 Manifesto, and in the general presentation of HyperCities, is in implicit contrast to the technologies and tropes of an earlier ferment of digital humanities, which, like the Web itself at the time, assumed the client-server architecture. Whereas HyperCities seeks to foster "participatory scholarship, open-source models for sharing content and applications, iterative development, and interdisciplinary collaboration" (3), earlier instances of humanities work online were comparatively less open, less participatory, more hierarchical, and more cloistered in their approaches. As Presner notes, HyperCities itself began as a hypermedia "textbook," a rigid silo of links and hardwired geo-codes teetering atop a Flash front-end. Contrast this with the current HyperCities, described by Presner as "a participatory platform that features collections that pull together digital resources via network links from countless distributed databases. . . . the connective tissue for a multiplicity of digital mapping projects and archival resources that users curate, present, and publish" (4). More concretely, and also as Presner relates, HyperCities is a fully realized Web 2.0 environment that relies on feeds, APIs, and markup to present an aggregated and distributed platform for developing collections like the one on display in the Tehran Election video.

Of course it's particularly apt to be discussing these issues here at Virginia. As a graduate student I would sit in a cubicle in the Institute for Advanced Technology in the Humanities, where as the Blake Archive's project manager I would manually stitch together SGML files arriving via email and FTP from the archive's various editors—distributed across three other institutions—fuse them to image sets that were snail mailed to Virginia from Chapel Hill on CDs, tidy and groom the whole package for the dictates of the Institute's closed source DynaWeb publishing system, and then finally "make book" (as the UNIX command line process was known) to create an electronic edition that would be added to the virtual shelves of the William Blake Archive by way of manual HTML links. Los at his forge this was not, but at various times I still wondered whether anyone else could follow the exact same sequence of hacks, workarounds, kludges, and tweaks I used to get the system to work.

Which brings us to the overarching concern of this gathering: sustainability. Indeed, I would argue that for an enterprise such as HyperCities, sustainability is a moral and ethical imperative as well as a fiduciary and academic one. If we take seriously the claim that the Tehran materials represent "one of the largest

[1] http://www.youtube.com/watch?v=qkEN02dGOlU (<http://www.youtube.com/watch?v=qkEN02dGOlU>).

[2] http://www.digitalhumanities.ucla.edu/images/stories/mellon_seminar_readings/manifesto%2020.pdf (<http://www.digitalhumanities.ucla.edu/images/stories/mellon_seminar_readings/manifesto%2020.pdf>).

[3] http://www.collex.org/?page_id=2 (<http://www.collex.org/?page_id=2>).

existing documents of the election protests in Iran," then we have a responsibility to think about how it will be preserved and accessed as part of the historical record for many years to come.

These issues have been raised before of course, including here at Virginia a decade earlier under the auspices of one of the first data curation studies with a specific focus on humanities content of which I'm aware: the Supporting Digital Scholarship project, also funded by the Mellon Foundation.[4] Among SDS's accomplishments was an exploration of the "significant properties" of several first-generation digital humanities projects developed at IATH (the Rossetti Archive and Salisbury Cathedral) as well as experimental mapping of those projects to METS representations and ingesting these into a Fedora repository. SDS, however, dealt with relatively homogeneous collections of files housed on a common server. In the case of HyperCities, the two most salient points are these: first, that it is a platform, that is a piece of software; and second, that broad swaths of its content are distributed across the Web.

Recently I have been part of a project that has addressed these matters in the neighboring domain of virtual worlds and video games. Virtual worlds, unlike virtual globes, do not traffic in real world geospatial data. Second Life is the archetype, but examples of multi-user persistent virtual spaces go back to the late 1970s, with the multi-user dungeons (MUDs) that followed directly from the archetypal storyworld, Will Crowther's Colossal Cave Adventure. It's also worth noting the increasing prevalence of virtual worlds in serious humanities research: Virginia's own Rome Reborn might well be the most prominent example, but IBM Interactive's Beyond Space and Time, a MMORPG-style recreation of the Forbidden City, is also worth mentioning.[5] For my purposes today I want to focus not on the dungeons and dragons trappings of such things, nor their status as popular culture commodities—Blizzard's World of Warcraft is at $250 million in annual revenue and climbing—but rather ways in which the explorations of the Preserving Virtual Worlds (PVW) project illuminate issues of software preservation as well as the collection and curation of distributed user-generated content.

***

A quick word about the PVW project itself: the Rochester Institute of Technology, Stanford University, the University of Illinois at Urbana-Champaign and the University of Maryland are concluding a two-year exploratory study investigating the preservation of computers games and interactive fiction.[6] Sponsored under the Library of Congress' National Digital Information Infrastructure for Preservation Program (NDIIPP), this project seeks to identify the specific difficulties in the preservation of computer games and interactive fiction that distinguish them from other forms of digital information we wish to preserve, to develop metadata and packaging practices to allow us to manage the long-term preservation of these digital materials in a manner consistent with the Open Archival Information System Reference Model, and to test those practices via ingest of computer games and interactive fiction into a set of functioning digital repositories. Key deliverables include development of metadata schema and wrapper recommendations, the archiving of key representative content and the development of generalizable archiving approaches for preserving this content. Our approach is intended to address both the pressing need to preserve the bits and available representation information of early and significant works now, and the need to begin to address more difficult issues surrounding long-term preservation of more recent multi-player interactive virtual worlds.

Based on our experiences with this work and lessons learned, I would like to proffer two key potential preservation paths for HyperCities:

- *HyperCities as software*. Presner describes HyperCities as a "platform," and more specifically as "a generalizable, easily scalable data model for linking together and publishing geo-temporal content using a unified front-end delivery system and a distributed back-end architecture" (6). He adds, "The front-end is almost a complete application itself because it contains all the display logic" (7). To

[4] The final report is available here: http://www.iath.virginia.edu/~spw4s/sarah/SDS/SDS_AR_2003.frame.html (<http://www.iath.virginia.edu/~spw4s/sarah/SDS/SDS_AR_2003.frame.html>).

[5] http://www.romereborn.virginia.edu/ and http://www.beyondspaceandtime.org/FCBSTWeb/web/index.html (<http://www.romereborn.virginia.edu/ and http://www.beyondspaceandtime.org/FCBSTWeb/web/index.html>). Earlier forays into scholarly "virtual reality" at IATH include VRML renderings of Rossetti's Cheyne Walk studio and a visualization of Dante's Inferno.

[6] For more on PVW, please refer to the project Web site: http://pvw.illinois.edu/pvw/ (<http://pvw.illinois.edu/pvw/>)

the extent these display logics, presumably calibrated to the needs and imperatives of humanities scholarship, contribute "significant properties" (see Supporting Digital Scholarship) to the end-user experience of HyperCities, what we have here is a problem in software preservation. The challenge is not simply to maintain access to a pile of files—which might be accomplished through periodic migration to more stable formats—but to ensure the possibility of actually running the HyperCities front-end on some future system. Software preservation is at the center of our work in Preserving Virtual Worlds because that is what computer games actually are: software programs. To that end, we have engaged in extensive "packaging" of the original executables, contextualizing them according to the requirements of the Open Archival Information System reference model with so-called "representation information," that is all of the second-order information related to operating system and peripherals that is necessary to recreate the complete environment in which the original program executed. This is not a trivial undertaking: the 1980 interactive fiction game *Mystery House*, a relatively simple text-adventure with some crude vector graphics released by Sierra Online Systems, required almost 400 MB of representation information, none of it rich media files. Nonetheless, such approaches are technically viable, and demonstrate what may lie ahead for HyperCities.

- *HyperCities as distributed or user-generated content.* This is, as I have described it elsewhere, the brick wall at which the whole archival enterprise is currently hurtling at 70 mph. I have suggested and I believe that with hindsight, digital artifacts from the first twenty-five years or so of personal computing will represent an anomalous window as the archivist or digital conservationist will stand some reasonable prospect of actually holding the original storage media in his or her hand.[7] Nowadays, of course, more and more of our online activity takes place in the so-called cloud. HyperCities leverages this to great effect, as Presner describes eloquently in his paper; the contrast to a project such as the Blake Archive as I described it above could not be more striking. Nonetheless, the possibility of gathering into one's arms all of the constituent parts and pieces of one of HyperCities' slices of curative argumentation is virtually impossible (so to speak), not only for technical reasons but also because of the barbed wire thicket of terms of service and end-user license agreements that govern our access to all of the most common "Web 2.0" services. (This matter has gotten some popular media play as next of kin of servicemen and -women killed overseas have had to fight for access to their loved one's email accounts.[8] The best summary of the issues to date is by Simson Garfinkel and David Cox.[9]) We have had some success in Preserving Virtual Worlds using the Internet Archive's crawl services to harvest material from sites devoted to games and gamer culture, as well as spot-setting TwapperKeeper archives and the like. We did not have occasion to look into the current state of archiving KML layers in Google Earth/Maps, but presumably some tools exist. Nonetheless, the primary issues will surely be legal and social, as opposed to technical. This was manifested most dramatically in our attempts to archive content from several Second Life islands: while it proved possible to write scripts to collect rendering information as it was passed from server to client and therefore store aspects of the world's virtual geometry, permission to do so was another matter. Everything in Second Life is user-generated and user-contributed. Therefore, in order to archive the contents of a space we were forced to attempt to reach out to hundreds of individual users with a request for permission; return of response was low. In this respect, Second Life functions as a dramatization of the kind of scenario one might expect from collections of user-contributed data on the 2-D Web.

[7] See Kirschenbaum et al. 2009. Approaches to Managing and Collecting Born-Digital Literary Materials for Scholarly Use. Office of Digital Humanities, National Endowment of the Humanities: http://www.neh.gov/ODH/Default.aspx?tabid=111&id=37 (<http://www.neh.gov/ODH/Default.aspx?tabid=111&id=37>).

[8] http://news.zdnet.co.uk/internet/0,1000000097,39195962,00.htm (<http://news.zdnet.co.uk/internet/0,1000000097,39195962,00.htm>).

[9] Simson Garfinkel and David Cox, "Finding and Archiving the Internet Footprint": http://simson.net/clips/academic/2009.BL.InternetFootprint.pdf (<http://simson.net/clips/academic/2009.BL.InternetFootprint.pdf>).