Gathering and Interpreting Website Changes:
Thailand's Sexual Reassignment Surgery Industry Examined

Jillana Enteen, PhD
Gender & Sexuality Studies
Northwestern University

j-enteen@northwestern.edu

Albert Yan
Research Assistant
Northwestern University

Josh Honn
Digital Scholarship Fellow
Northwestern University

In the past decade, Thailand's medical tourism industry has grown exponentially; Thailand is now the top destination for Sexual Reassignment Surgery/Gender Reassignment Surgery (SRS/GRS). Once an embedded sub-category of cosmetic surgery, SRS is now prominently situated on Thai-generated websites geared to attracting non-Thais for medical tourism. As medical procedures for SRS/GRS have advanced over time, the number of Thai generated websites aimed at Western clientele in Thailand now outnumber the number of cosmetic surgery sites.

Because the growth of modern medical travel has developed subsequent to the shift to internet as a travel information source, the internet serves as the primary repository for information concerning medical tourism, the travel to other countries for elective surgery procedures. Medical tourism/cosmetic surgery travel websites are creating an industry involving thousands of Thais and Western visitors annually, one that has received scant academic scrutiny.

Our research team refined our collection of 150 websites to 70 websites that specifically offer SRS to Westerners; we made this decision in order to track the industry during this unique period of growth. While some websites are offered in languages other than English, our sample is exhaustive of the english-language websites offering SRS in Thailand. The three types of website changes this project is concerned with are: (1) Compare how a single website undergoes changes over time (California Digital Library's Web Archiving Service); (2) Compare within a single timeframe across multiple websites (Devonthink Pro Database Management system); (3) Locate the emergence of new websites (Biweekly Google searches).

Our team uses a combination of services and software to organize the collected data for our sample. California Digital Library's Web Archiving Service allows us to conduct bimonthly crawls of 70 URLs. Web crawling saves websites directories to a CDL server and available anywhere with an internet connection. Crawling a site downloads the websites entire directory or 'website infrastructure' so that it is available locally offline. The Service's 'Change Report' compares two website crawls of the same URL from two different time periods and reports on items newly added, changed, missing, or unchanged. California Digital Library's Web Archiving Service allows us to be precise and exact in monitoring exact changes on websites as they occur.

We employ Devonthink Pro, a database management system that allows us to record, consolidate, organize, and analyze the changes reported by the Web Archiving Service. All 'CDL Change Reports' are logged in Devonthink Pro, along with a collection of website captures in searchable PDF format. Our database contains over 1,200 individual website captures spanning the last ten years. Each PDF file maintains the integrity of the site's formatting, style, and overall presentation, which allows us to qualitatively and quantitatively assess each page. Combining the Web Archiving Service and Devonthink Pro, we are able to conduct vertical analysis of change over time as well as horizontal analysis of any given point in time across the industry. The Web Archiving Service registers all website changes while Devonthink Pro allows us to organize the results. The final product is a database of the online presence of all hospitals, clinics, doctors, and health providers providing Sexual Reassignment Surgery in Thailand from 2009-2012.

Crawling the web gives us a mountain of website directories; these directories vary immensely from one to another because most websites are developed by different web developers. A file called 'robots.txt' presents a challenge as well. 'Robots.txt' tells any 'bot', such as our CDL web crawler, the specific pages that are forbidden from web crawls. As a result, 'Robots.txt' disallows the Web Archiving Service from accessing certain pages of websites. To accommodate this, we have been manually observing the sites on a weekly/biweekly basis to monitor specific pages blacked by the "robots.txt" file. If there is any way to override 'robots.txt', gathering website data would be more streamlined.

For the years 2002-2009, our team used The WayBack Archive to explore past versions of our 70 URLs. Broken links, missing style sheets, and deleted files are scattered throughout our older data; our goal is to retrieve internet website data retroactively for years 2002-2009 to contextualize our three year findings from 2009-2012. Broken links and incomplete website pages hinder our team's ability to analyze website changes over time.