ROLLING YOUR OWN DIGITAL LIBRARY SYSTEM
Homegrown is not necessarily a bad thing

Many systems begin as “home grown”

Dspace
Fedora
ContentDM
Greenstone
Keystone
Streetprints

I prefer to call it “locally developed”
Reasons to Build Your Own System

You cannot scale you system adequately to meet your various content needs.
- Archival Storage
- Presentation Storage
- Preservation Architecture
- Search Systems

Other solutions do not meet your needs
- Why are your needs special? Are they?
- What would you lose by using an existing system?
- What do you gain by building your own system?
Be Careful of the NIH Syndrome

Not Invented Here syndrome

Be able to justify why you are doing it yourself.

Do you have good institutional support?

Will building it cost as much or more than another solution?

Would it be cheaper to work with an existing product and customize?

Would you actually be able to build the system you want?
UNT Examples

UNT contracted with IndexData in 2003 to customize an infrastructure to support the Portal to Texas History.

We were able to capitalize on this investment by spinning off two other “digital library” projects with the same software.

Code was open source so we could modify stuff.

Code was open source so others gained from our investments.
But we didn’t know what to ask for.

We had no idea what we wanted in a Digital Library system, we thought we did but really... we didn’t.

So we started “fixing” and “making it better”

- Granularity issue with preservation metadata
- Ability to store OCR content at the page level
- Ability to store and present transcriptions at the page level
- Ability to store and present translations at the page level
- Ability to search for content at the page level
- Ability to quickly turn pages of a book
- Ability to jump to a page via a dropdown
... Ability to show textual snippets at the page level
Ability to highlight searched terms on an image of a page
Ability to add a zoom able interface for large images
Ability to add audio to the site
Ability to add video to the site
Ability to quickly view thumbnails for all pages of an object
We changed our In Document Search (IDS) feature two times
We changed our entire indexing system from Zebra to Solr
IOGENE Project

Allowed us to rethink what we wanted in a system.

- Based on a Web framework for all of the “Web stuff”
- We wanted a large community for support
- We wanted to concentrate on the “library stuff” not the other Web stuff
- We wanted to have loosely coupled components
So we are building...

A new system using some of the following components.

Python
Django
Solr
jQuery
Apache
mod_python
ubuntu
MySQL
memcached
Perl
Bal
SVN
Trac
Some things to think about if you decide to build your own.

Standards are your friends
Persistent identifiers from the start
Understand your data models
Understand how you want to search your content
Understand how you want to browse your content
Make sure you work with a toolset that has a good support community
Make sure your administration knows it takes time and money to build your own system
“Re-inventing the wheel is great if your goal is to learn more about wheels.”

James Tauber

Pycon 2007