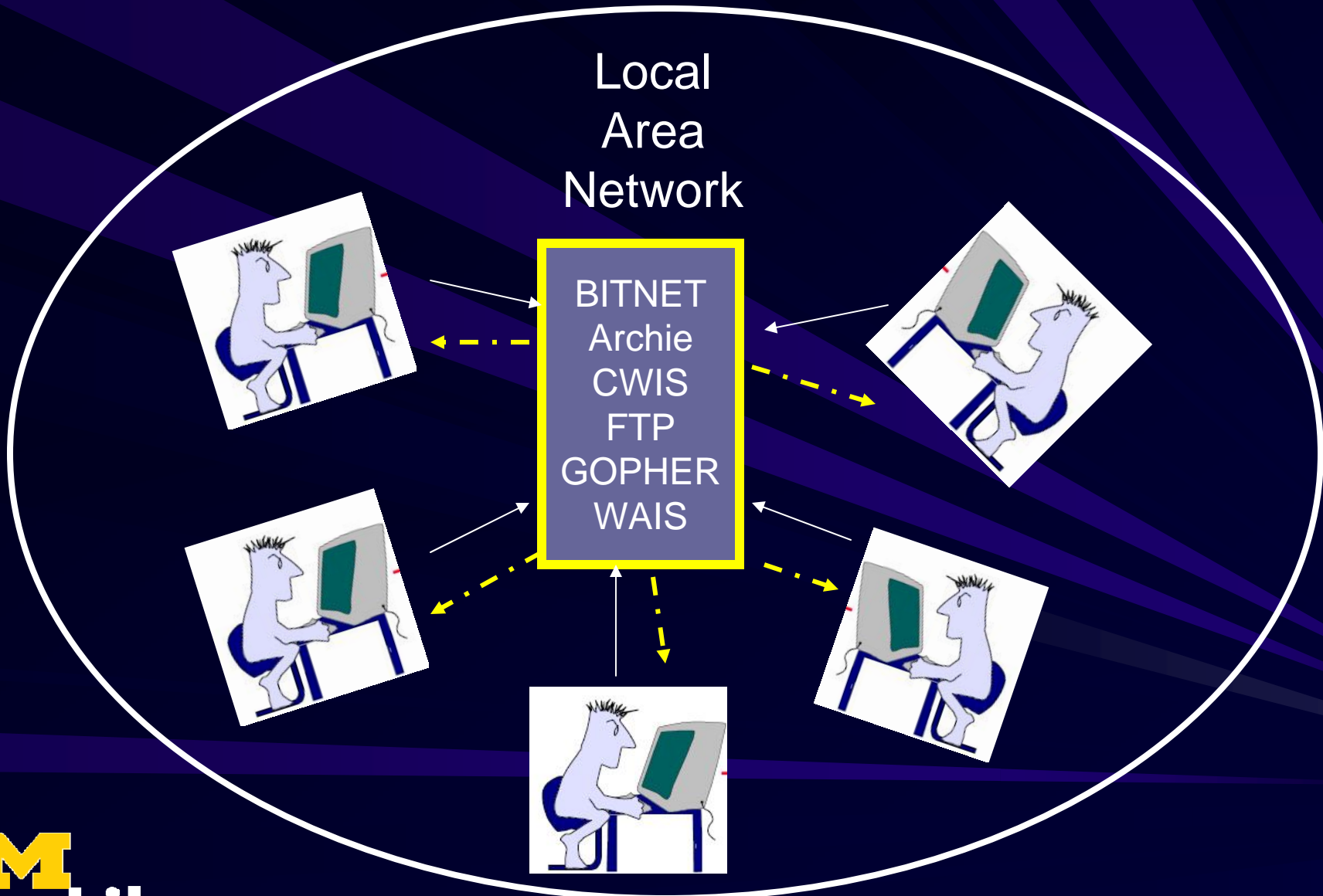


Designing Workflows that Meet Local Needs

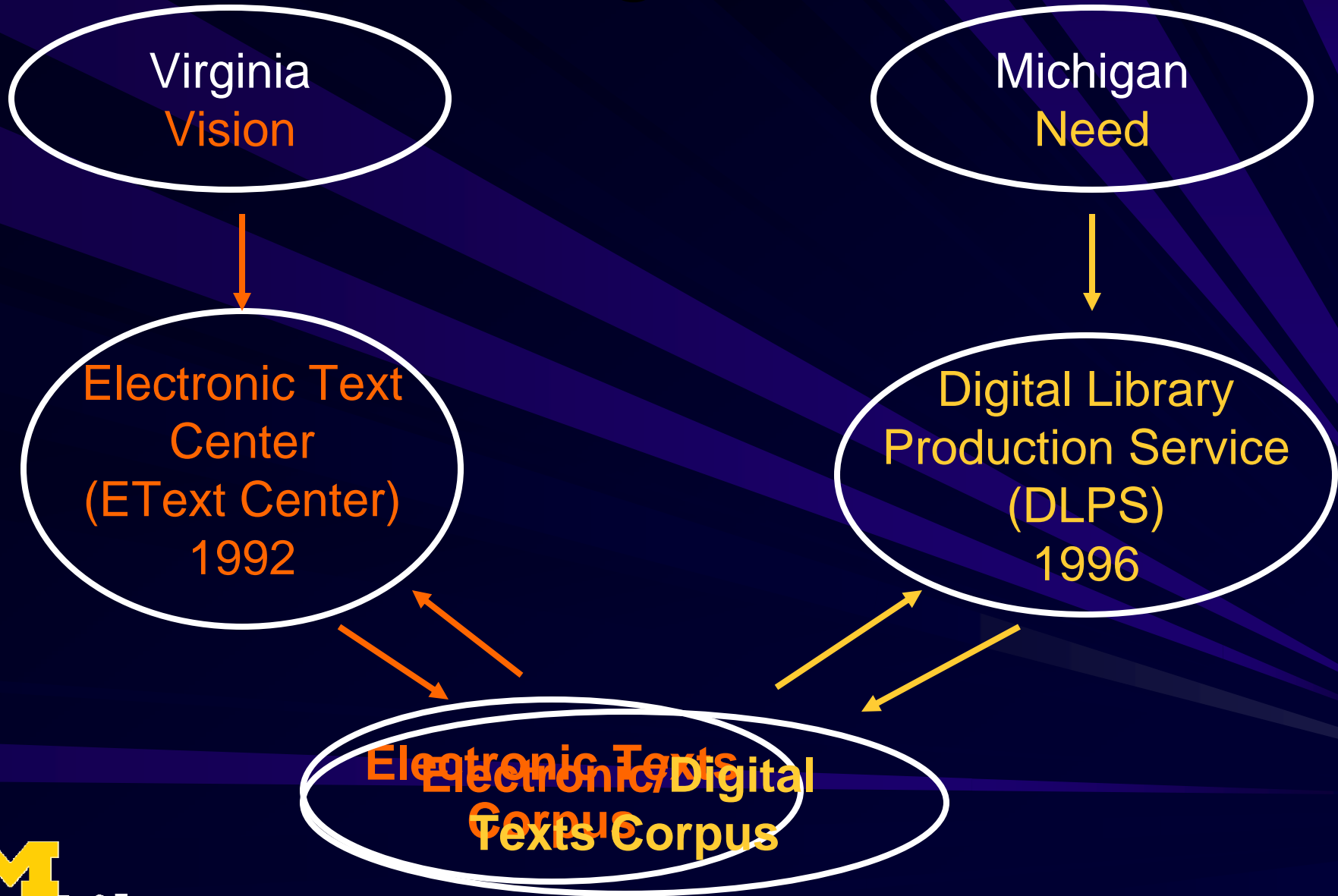
ALA Annual 2007
Washington, D.C.

Jackie Shieh
jshieh@umich.edu
University of Michigan Library

Early 1990's ...



It All Began with ...



Catalysts for Change ...

- Uncharted Territory
- Technology Infrastructure
 - Mosaic Browser (1993)
 - Unix, Perl, CGI
- Collection Building Strategies
- Standards
 - SGML, TEI
 - MARC, AACR
- Data Format
 - SGML tags
<letters and phrases in angled brackets>
 - MARC tags
- New Cataloging Tools
 - OCLC Passport (1991)
 - ILS
- Helper Applications & Tools
 - UKOLN Metadata
 - Perl Toolkits
 - MARCEdit

Cataloging Collection

Virginia

Etext Collection

Encoding texts delivery
supplemented
with images

Michigan

Digital Texts Collection

Production-based
page Image delivery
with OCR searchability

Cataloging
Bibliographic Control
TEI Header
MARC

**Digital
Collection**

Item Request

Exist in
Print?

Yes

Electronic
Reproduction

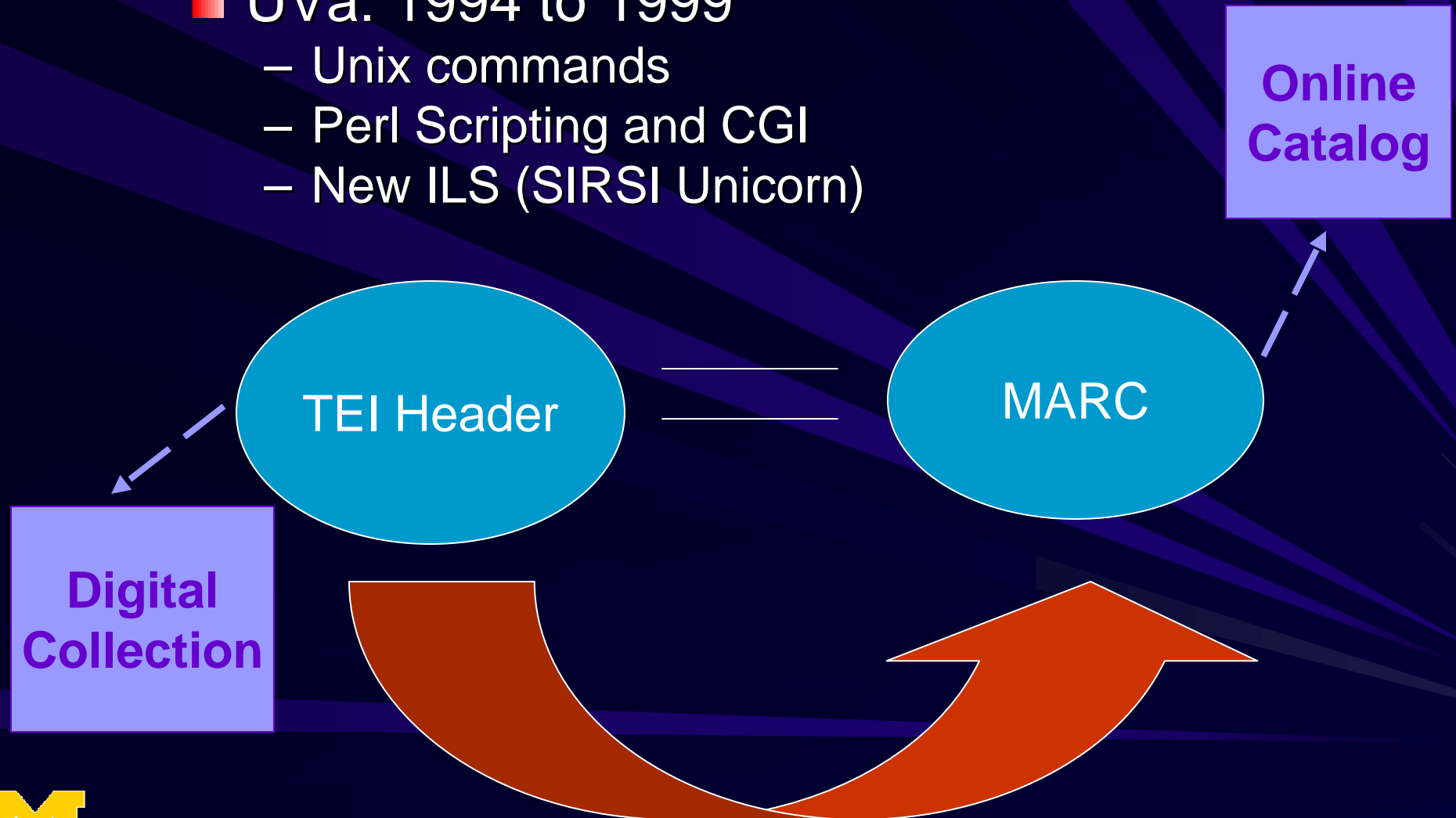
No

Electronic Edition

Yes **BUT**
In Parts

Virginia Workflow

- UVa: 1994 to 1999
 - Unix commands
 - Perl Scripting and CGI
 - New ILS (SIRSI Unicorn)



One Stone, Two Birds

teiHeader

fileDesc

titleStmt TITLE STATEMENT

title [main title : subtitle / author statement ; other responsibility statement]

type =

title [electronic resource]

author

editor

respStmt

resp

name

resp

name

resp Conversion to TEI.2-conformant markup:

name

extent

publicationStmt PUBLICATION STATEMENT

publisher University of Virginia Library

pubPlace Charlottesville, Virginia

idno

SUBMIT

```
<teiHeader type="aacr2">
<fileDesc>
<titleStmt>
<title type="245">Rubaiyat of Omar Khayyam of Naishapur / Omar Khayyam.
</title>
<title type="gmd">[electronic resource]
</title>
<author>Omar Khayyam
</author>
<editor>
</editor>
<respStmt>
<resp>
<name>
</name>
</resp>
</resp>
<name>
</name>
</resp>
</resp>
<name>Creation of machine-readable version:
</name>Electronic Text Center
</name>
</resp>
</resp>
<name>Creation of digital images:
</name>Electronic Text Center
</name>
</resp>
</resp>
<name>Conversion to TEI.2-conformant markup:
</name>University of Virginia Library Electronic Text Center
</name>
</resp>
</resp>
</title>
</title>
<extent>ca. 15 kilobytes
</extent>
```

TEI Header

```
UVA MARC RECORD
FORM=MARC
.000. |aamI
.006. |am d
.007. |ac|br|dc|en
.008. |a980831s1997 vaua eng d
.099. |aClick|aon|aURL
.040. |aVA@|cVA@
.049. |aVA@@
.100. 0 |aOmar Khayyam.
.245. 10|aRubaiyat of Omar Khayyam of Naishapur|h[computer file] /
|cOmar Khayyam.
.256. |aComputer text data (1 file : ca. 15 kilobytes).
.260. |aCharlottesville, Va. :|bUniversity of Virginia Library,|c1997.
.516. |aText (SGML) and digital images (GIF and JPEG)
.538. |aMode of access: Internet. Host: etext.lib.virginia.edu
.500. |aTitle from TEI header.
.500. |aPrepared for the University of Virginia Library Electronic Text
Center.
.500. |aCreation of machine-readable version: Electronic Text Center.
.500. |aCreation of digital images: Electronic Text Center.
.500. |aConversion to TEI.2-conformant markup: University of Virginia
Library Electronic Text Center.
.500. |aThis electronic edition consists entirely of page images.
.516. |aTagging checked and parsed against teillite.dtd
.516. |aAll unambiguous end-of-line hyphens have been removed, and the
trailing part of a word has been joined to the preceding line.
.516. |aThe images exist as archived TIFF images, one or more JPEG
versions for general use, and thumbnail GIFs.
```

MARC ASCII

Processing Ascii MARC

MARC ASCII

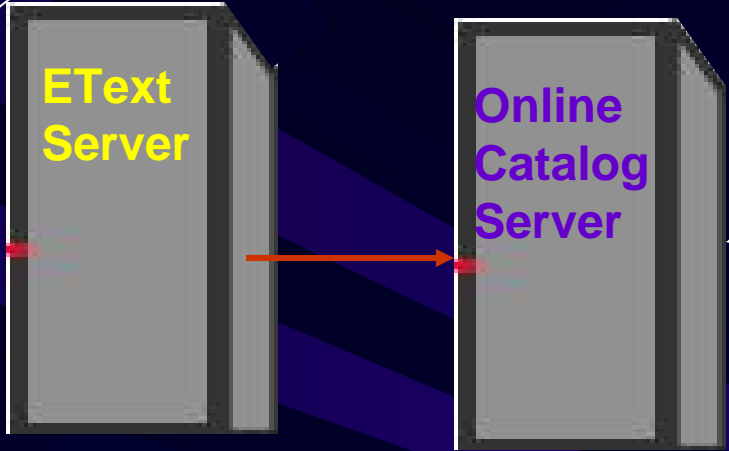
```
001 1001  
005 200101010000.0000  
007 1  
010 1001  
011 1001  
012 1001  
013 1001  
014 1001  
015 1001  
016 1001  
017 1001  
018 1001  
019 1001  
020 1001  
021 1001  
022 1001  
023 1001  
024 1001  
025 1001  
026 1001  
027 1001  
028 1001  
029 1001  
030 1001  
031 1001  
032 1001  
033 1001  
034 1001  
035 1001  
036 1001  
037 1001  
038 1001  
039 1001  
040 1001  
041 1001  
042 1001  
043 1001  
044 1001  
045 1001  
046 1001  
047 1001  
048 1001  
049 1001  
050 1001  
051 1001  
052 1001  
053 1001  
054 1001  
055 1001  
056 1001  
057 1001  
058 1001  
059 1001  
060 1001  
061 1001  
062 1001  
063 1001  
064 1001  
065 1001  
066 1001  
067 1001  
068 1001  
069 1001  
070 1001  
071 1001  
072 1001  
073 1001  
074 1001  
075 1001  
076 1001  
077 1001  
078 1001  
079 1001  
080 1001  
081 1001  
082 1001  
083 1001  
084 1001  
085 1001  
086 1001  
087 1001  
088 1001  
089 1001  
090 1001  
091 1001  
092 1001  
093 1001  
094 1001  
095 1001  
096 1001  
097 1001  
098 1001  
099 1001  
100 1001
```

MARC ASCII

```
001 1001  
005 200101010000.0000  
007 1  
010 1001  
011 1001  
012 1001  
013 1001  
014 1001  
015 1001  
016 1001  
017 1001  
018 1001  
019 1001  
020 1001  
021 1001  
022 1001  
023 1001  
024 1001  
025 1001  
026 1001  
027 1001  
028 1001  
029 1001  
030 1001  
031 1001  
032 1001  
033 1001  
034 1001  
035 1001  
036 1001  
037 1001  
038 1001  
039 1001  
040 1001  
041 1001  
042 1001  
043 1001  
044 1001  
045 1001  
046 1001  
047 1001  
048 1001  
049 1001  
050 1001  
051 1001  
052 1001  
053 1001  
054 1001  
055 1001  
056 1001  
057 1001  
058 1001  
059 1001  
060 1001  
061 1001  
062 1001  
063 1001  
064 1001  
065 1001  
066 1001  
067 1001  
068 1001  
069 1001  
070 1001  
071 1001  
072 1001  
073 1001  
074 1001  
075 1001  
076 1001  
077 1001  
078 1001  
079 1001  
080 1001  
081 1001  
082 1001  
083 1001  
084 1001  
085 1001  
086 1001  
087 1001  
088 1001  
089 1001  
090 1001  
091 1001  
092 1001  
093 1001  
094 1001  
095 1001  
096 1001  
097 1001  
098 1001  
099 1001  
100 1001
```

MARC ASCII

```
001 1001  
005 200101010000.0000  
007 1  
010 1001  
011 1001  
012 1001  
013 1001  
014 1001  
015 1001  
016 1001  
017 1001  
018 1001  
019 1001  
020 1001  
021 1001  
022 1001  
023 1001  
024 1001  
025 1001  
026 1001  
027 1001  
028 1001  
029 1001  
030 1001  
031 1001  
032 1001  
033 1001  
034 1001  
035 1001  
036 1001  
037 1001  
038 1001  
039 1001  
040 1001  
041 1001  
042 1001  
043 1001  
044 1001  
045 1001  
046 1001  
047 1001  
048 1001  
049 1001  
050 1001  
051 1001  
052 1001  
053 1001  
054 1001  
055 1001  
056 1001  
057 1001  
058 1001  
059 1001  
060 1001  
061 1001  
062 1001  
063 1001  
064 1001  
065 1001  
066 1001  
067 1001  
068 1001  
069 1001  
070 1001  
071 1001  
072 1001  
073 1001  
074 1001  
075 1001  
076 1001  
077 1001  
078 1001  
079 1001  
080 1001  
081 1001  
082 1001  
083 1001  
084 1001  
085 1001  
086 1001  
087 1001  
088 1001  
089 1001  
090 1001  
091 1001  
092 1001  
093 1001  
094 1001  
095 1001  
096 1001  
097 1001  
098 1001  
099 1001  
100 1001
```



Automatic Process:

- Unit Records collated into one single file
- Transfer File on a timer set by cron daemon

Staff manually uploads and runs indexing

Other Projects

- ICPSR Codebook (1995)
 - Unix cron daemon, Perl scripts
- Electronic Theses and Dissertation Pilot Project (1997)
 - Perl scripts, CGI

Michigan Workflow



Helper Applications

- Michigan: 1996-Present
 - Perl Scripting
 - marc2tei
 - marc2teiutf (2005)
 - XSLT
 - Perl Modules
 - MARC.pm (1999-2003)
 - MARC::Record (2003-present)
 - Windows-based
 - MARCEdit (2000)

One Swap, Many Flies

ILS
Export MARC

DLPS
1996

SPO
Scholarly
Publishing
Office
2001

ERRWPC
Elec Res
Received
Without
Physical
Container

Processing MARC

Online
Catalog
Server

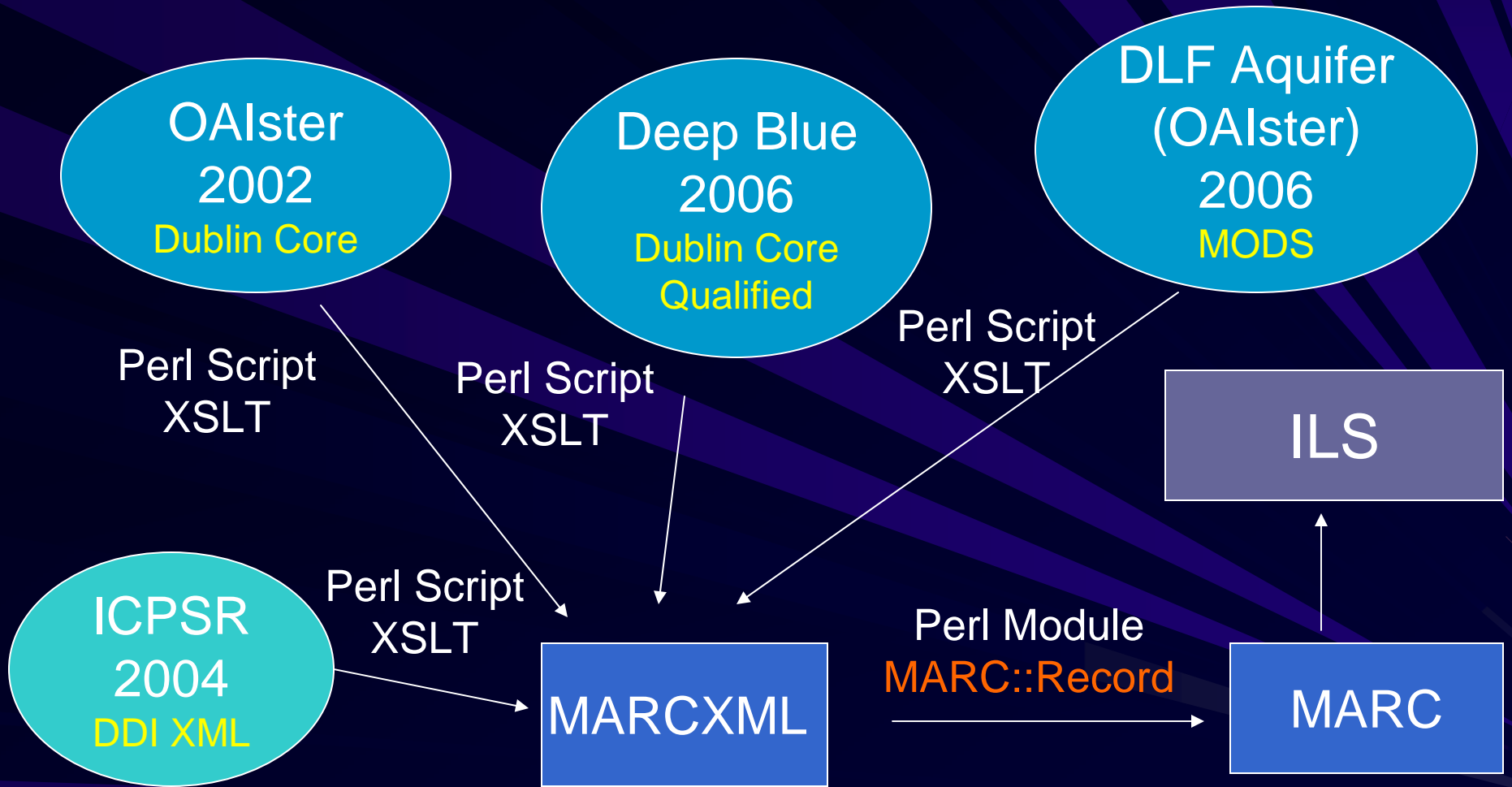
MARC
00750mam 22002051
45000010013000000080041000130100017000540350025000710
35002400096040001300120050001600133100004200149245007
40019126000400026530000430030561000740034865101000042
2998001400522999000900536ULAAS9842 840921s1866
ctuac 00010 eng a 02012793 a(RLIN)
MIUG84-B54284 a(CaOTULAS)160188259 cmIudMIU1
aE499.5.27th10aSheldon, Winthrop Dudley,d1839-
1931.04aThe "Twenty-seventh."ba regimental
history.cBy Winthrop D. Sheldon.0 aNew-Haven,bMorris
& Benham,c1866. a144 p.bfront., illus. (port.)c19
cm.10aUnited States.bArmy.bConnecticut infantry
Regiment 27th, 1862-1863. 0aUnited
StatesxHistoryyCivil war, 1861-1865xRegimental
historiesxConnecticut infantryx27th. cJF s9124
abuhr

marc2tei.pl
marc2teiutf.pl

TEI Header

```
<TITLE TYPE="245" 12="4">The "Twenty-seventh." A regimental history. By Winthrop
D. Sheldon.</TITLE>
<AUTHOR> Sheldon, Winthrop Dudley, 1839-1931.</AUTHOR>
</TITLESTMT>
<EXTENT> 800dpi TIFF 04 page images</EXTENT>
<PUBLICATIONSTMT><PUBLISHER>University of Michigan Library</PUBLISHER>
<PUBPLACE>Ann Arbor, Michigan</PUBPLACE>
<DATE>2005</DATE>
<IDNO TYPE="dips">AA39442 </IDNO>
<IDNO TYPE="loclesso">E499.5.27th</IDNO>
<AVAILABILITY><P>These pages may be freely searched and displayed. Permission must
be received for subsequent distribution in print or electronically. Please go to
http://www.uofl.umich.edu/ for more information.</P></AVAILABILITY>
</PUBLICATIONSTMT>
<SOURCEDESC>
<BIBLFULL><TITLESTMT><TITLE TYPE="245" 12="4">The "Twenty-seventh." A regimental
history. By Winthrop D. Sheldon.</TITLE>
<AUTHOR> Sheldon, Winthrop Dudley, 1839-1931.</AUTHOR>
</TITLESTMT>
<EXTENT>144 p. front., illus. (port.) 19 cm. </EXTENT>
<PUBLICATIONSTMT>
<PUBPLACE>New Haven.</PUBPLACE>
<PUBLISHER>Morris & Benham.</PUBLISHER>
<DATE>1866.</DATE>
</PUBLICATIONSTMT>
</BIBLFULL>
</SOURCEDESC>
</FILEDESC>
<ENCODINGDESC><PROJECTDESC><P>Header created with script marc2teiutf.xml on
2005-05-27.</P>
</PROJECTDESC>
<EDITORIALDECL N="1"><P>This electronic text file was created by Optical Character
Recognition (OCR). No corrections have been made to the OCR-ed text and no editing
has been done to the content of the original document. Encoding has been done usi
ng the recommendations for Level 1 of the TEI in Libraries Guidelines. Digital pa
ge images are linked to the text file.</P>
</EDITORIALDECL></ENCODINGDESC>
</PROFILEDESC>
<LANGUAGE ID="eng"><LANGUAGE>eng</LANGUAGE></LANGUAGE>
<TEXTCLASS><KEYWORDS>
<TERM>United States -- History</TERM>
<TERM>United States. -- Army.</TERM>
</KEYWORDS></TEXTCLASS>
```

Outgrowths



Where Do We Go from Here...

- Extend Derivative Processes
- Expand Cataloging Boundaries
- Engage and Aid in Tool Creations in preparation for RDA, DCMI, FRBR, Semantic Web

Thank you!

Designing Workflows that Meet Local Needs

ALA Annual 2007
Washington, D.C.

Jackie Shieh
jshieh@umich.edu
University of Michigan Library