Digital Resources in Title VI: planning the next 50 years

Doug Cooper  Center for Research in Computational Linguistics
today

one  where do we stand

two where should we be

three open up

four what must be done
Part 1: Where do we stand?
Title VI

**IRS** International Research and Studies  
**TICFIA** Technological Innovation and Cooperation for Foreign Information Access  
**LRC|NRC|...**

funds hundreds of  

- projects  
- dictionaries  
- collections
Increasingly, digital resources are the face of Title VI

digital dictionaries
digital corpora
digital images
digital sound
digital video

but we’re not ready for our close-up
efforts to
“effectively apply developments in technology to
... support of language instruction suffer
from a dispersion of resources”

Title VI programs
“could do more to maximize the
potential created by current technologies”
they were being polite
we’re getting output

but not impact
we’ve lost touch with the zeitgeist

“There's been a culture war between librarians and computer scientists. ... Google won.”

Peter Lyman
Washington Post, 3/3/04
We point to hundreds of books, websites, collections
(remember Yahoo?)

We should point to millions of words, pictures, sounds
(heard of Google?)
we should go from funded by to powered by Title VI
we build beautiful houses

people are more interested in the foundations
we can **maximize** our **potential**

by realizing that **Title VI** builds **platforms**
their platform
POWERED BY
TICFIA/DATA
IMAGES:IRS
TITLE VI

our platform
Part 2: Where should we be?

“On the Internet, nobody knows you’re a dog.”
Title VI **is** language, education, mother pie
it’s also **a brand**

Title VI **competes** against other brands
NSF, NEH, IMSL, NSEP
Google, Wikipedia
Moodle, WebCT

Title VI **needs**
market share
name recognition

but most of all ...
killer apps

we compete for users
students, teachers, surfers

we should compete for developers

because they turn good into great
**moodle** virtual learning environment

60,000 registered sites 2006–2009
60,000 potential Title VI developers

Developers turn **platforms** into **juggernauts**
they turn output into impact
We think the problem is …

“People can’t find our data. We need:

web portals
metadata archives
more publicity
more evaluations

but we shouldn’t have to advertise data
we should be soaking in it
We *think* the problem is …

“Our websites need better Google placement / more bells / whistles / rounded corners / pastels / AJAX / JSON / Java / folksonomies / links / opinion polls / Flash / background music / YouTube / Web 2.0”

between *users*

All these do is keep us in the *middle* and *data*
the problem isn’t
we’re not giving people *enough buttons*

the problem is
people want to make *their own* buttons
the problem isn’t

the **public** can’t find our **data**

the problem is

our **data** can’t find its **public**
Part 3: open up

I live in a house without a door. Each person who visits must bring a door on his back. Install it before sitting down, then take it with him when he leaves. My privacy depends on visits from these people.

Phan Nhien Hao
house without a door
translated by Linh Dinh
we treat digital like paper

go to the **library** / get the **book**

go to the **website** / get the **data**
We understand *paper*

finished products

easily archived

readily reused

It’s our *past*. 
We don’t really get **digital**

**work-in-progress** work
**copyrighted** work
**borrowed** work
**embedded** work
(probably won’t ever)
**get-finished** work

But it’s our **future**.
we need places to store it

we need doors to let queries in
to let data out
storage

open
when possible

embargoed
in

dark archives
when necessary
access

interoperable
from day 1

when projects can’t/won’t release data

it frees the datum but hides the dataset
Google has a webpage
has a database
exposes an API
it’s the doorway for interoperability
Title VI

has a **hundred** dictionaries

why isn’t there an easy
way to look up a
**single word**
in all of them?
interoperunable
Part 4: what must be done

noodle
enable
reward
Noodge

“The primary advantage of noodging is that it does not require an Act of Congress.”

*Bela Abzug*  
(not really)
push projects to address

preservation access interoperability mortality

NSF/NEH/IMLS leads US/ED needs

create a culture of sharing
Enable
Enable

create a path of least resistance

develop best practices for interoperability

do better than Google and the Internet Archive
re-invent ERIC* for a digital age

*Education Resources Information Center
Reward
Reward

re-invent **IRIS***
- **less** reductionist
- **more** responsive

embrace **web**liometrics
- **community**-oriented measures

believe **Title VI** is the **platform**

---

*International Resource Information System*
5. Conclusion
don’t build silos
don’t build silos

apps sell platforms
don’t build silos

apps sell platforms

developers deliver apps
don’t build silos
apps sell platforms
developers deliver apps
smooth the way
the end.
Digital Resources in Title VI:

planning the next 50 years

Doug Cooper   Center for Research in Computational Linguistics
Doug Cooper
Center for Research in Computational Linguistics

**TICFIA** 2005 - 2009
SEAlang Library  Southeast Asian language resources

**IRS** 2006 – 2009
SEAlang Lab  assistive technology for reading, writing, vocabulary in complex-script languages

http://sealang.net


doug.cooper.thailand @ gmail.com