

Remote Access to Microdata Challenges and Practical Approaches

Julia Lane and Tim Mulcahy,
NORC/Uof Chicago
Pascal Heus, Open Data
Foundation

Overview

- ▶ Challenges
- ▶ Practical Approach
 - Enclave Structure
 - Portfolio Approach
 - Documentation
- ▶ Metadata Access

Challenges

- ▶ Provide access to micro data
- ▶ Protect Confidentiality
 - Technical
 - Legal
 - Organizational
 - Statistical
- ▶ Archive, Index and Curate Micro-data

Challenges to Providing Access

▶ Public Use Files

- Quality
- Timeliness

▶ Licensing

- Security

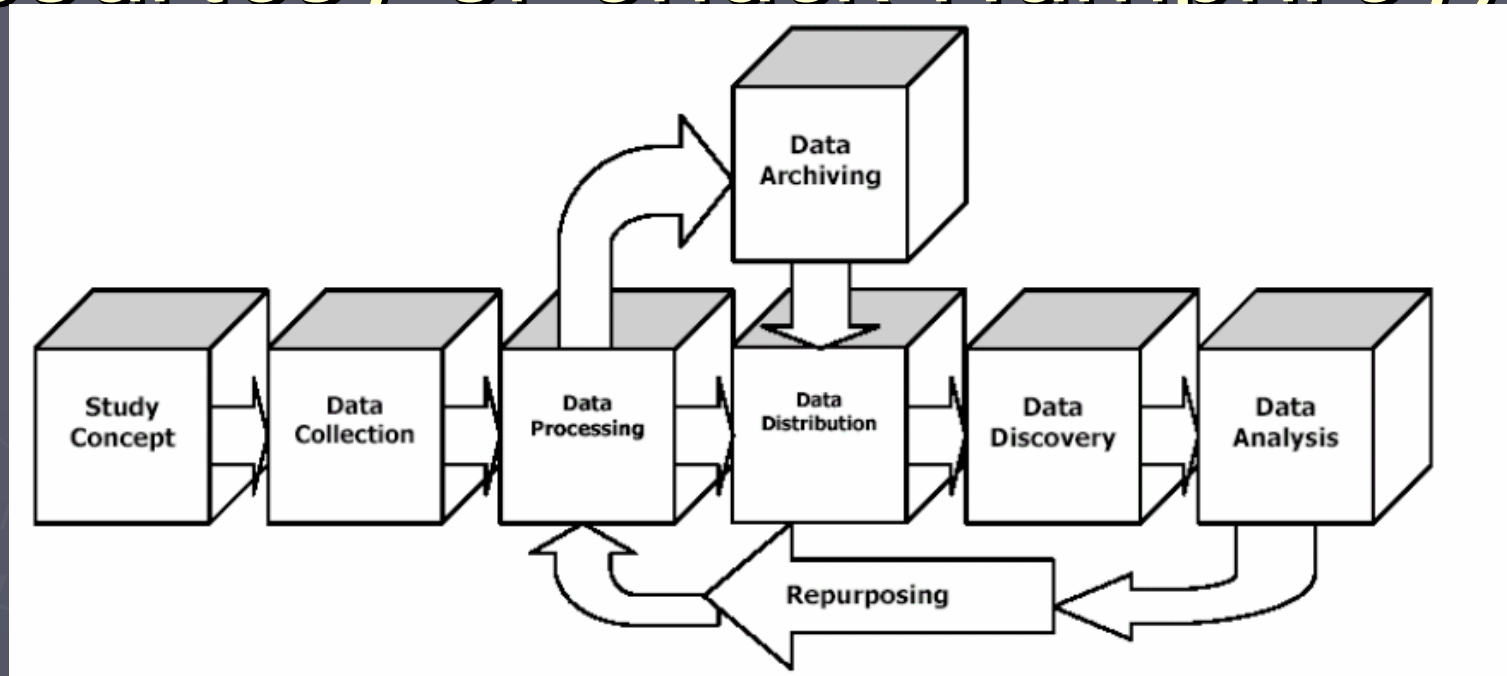
▶ RDC's

- Cost
- Accessibility

Ideal System

- ▶ Secure
- ▶ Flexible
- ▶ Low Cost
- ▶ Meet Replication standard
 - Gary King, Harvard, 1995
 - The only way to understand and evaluate an empirical analysis fully is to know the exact process by which the data were generate
 - Replication dataset include all information necessary to replicate empirical results
 - Metadata crucial to meet the standard
 - ▶ Composed of documentation and structured metadata
 - ▶ Undocumented data are useless
- ▶ Create foundation for metadata documentation and extend data lifecycle

Metadata and the survey life cycle (courtesy of Chuck Humphrey)



- ▶ Data collection is not a static process – it's a lifecycle
- ▶ It dynamically evolved across time and involves many players
- ▶ It extends to aggregate data to reach decision makers
- ▶ Metadata are crucial to capture knowledge

Practical Approaches: NORC Data Enclave

➤ Consortium

- Founding Member NIST (ATP)
- Additional members: USDA (ERS),
Kauffman Foundation

➤ Ongoing discussions with Federal Agencies

NORC Data Enclave

1. Data Protection
 - a) Already collect data for multiple statistical agencies (BLS, Federal Reserve (IRS data), EIA, NSF/SRS etc.) => safeguards in place
 - b) NIST approved IT security plan
2. Data archiving, indexing, and curation
 - a) Microdata Dissemination Toolkit (DDI)
 - b) Curation plan
3. Provision of access – a portfolio approach
 - a) Statistical protection (statistical)
 - b) Researcher training (Educational)
 - c) Dissemination to researcher community (Operational)
 - d) Agency-specific data protection requirements (Legal)
4. Advisory Boards

Provision of Research Access

Menu Options for Agency X (and Study Y)				
Sample Modalities	Legal Options (1,2,3,4)	Statistical (1,2,3,4,5)	Operational (1,2,3,4,5)	Educational (1,2,3,4)
Remote Access	3	1	4	2
	None	2	5	2
Onsite Access	3 with customization	3,5	1	None
Licensing (different levels of anonymization)	2	1	2,3	1,4

Data Protection

The Data Enclave is fully compliant with DOC IT Security Program Policy, Section 6.5.2, the Federal Information Security Management Act, provisions of mandatory Federal Information Processing Standards (FIPS) and all other applicable NIST Data IT system and physical security requirements. This includes

- ▶ Employee Security
- ▶ Rules of Behavior
- ▶ Nondisclosure Agreements
- ▶ IT System Security
- ▶ Applicable Laws and Regulations
- ▶ Network Connectivity
- ▶ Remote Access
- ▶ Physical Access

See

<http://dataenclave.norc.org/Portals/0/NORC%20Data%20Enclave%20Security%20Requirements1.pdf>

IT Security

- ▶ Encrypted connection with the data enclave using virtual private network (VPN) technology. VPN technology enables the data enclave to prevent an outsider from reading the data transmitted between the researcher's computer and NORC's network.
- ▶ Users access the data enclave from specific, pre-defined IP addresses.
- ▶ Citrix's Web-based technology.
 - All applications and data run on the server at the data enclave.
 - Data enclave can prevent the user from transferring any data from data enclave to a local computer.
 - Data files cannot be downloaded from the remote server to the user's local PC.
 - User cannot use the "cut and paste" feature in Windows to move data from the Citrix session.
 - User is prevented from printing the data on a local computer.
- ▶ Audit logs and audit trails

Legal and Statistical Protections

➤ Legal

- Access Agreement signed by institutional and individual researcher
- Approved institutions
- Access limited to data requested and authorized

➤ Statistical

- Remove obvious identifiers and replace with unique identifiers
- Statistical techniques chosen by agency (recognising data quality issues)

Note: Both are at discretion of agency and can go above and beyond the minimum level of protection

Researcher Training

➤ Subjects

- Basic confidentiality (CDAC)
- Agency specific (joint with agency)
- Dataset specific (joint with agency)

➤ Locations

- Onsite
- Webbased
- Researcher locations (AAEA, JSM, AOM, ASA, ASSA, NBER summer institute)

Note: The training is designed to go above and beyond current practice in terms of both frequency and coverage

ARMS Data Enclave Training Agenda

NORC.

*1350 Connecticut Avenue (between N and O), NW, Suite 500,
Washington D.C. 20036*

Monday, September 10

- | | |
|-------------|---|
| 8:30-9:00 | Welcome (NASS/ERS/NORC) |
| 9:00-10:30 | Data enclave navigation (NORC) |
| 10:30-10:45 | <i>Break</i> |
| 10:45-12:15 | Metadata documentation (NORC) |
| 12:15-1:15 | <i>Lunch</i> |
| 1:15-2:45 | Confidentiality and data disclosure (NORC) |
| 2:45-3:00 | <i>Break</i> |
| 3:00-4:00 | ARMS survey overview (ERS) – Jim Johnson, ERS |
| 4:00-4:10 | Confidentiality agreement signing |

Tuesday, September 11

- 8:30-9:00 ARMS data files and documentation (ERS) – Bob Dubman, ERS
- 9:00-10:00 Sampling and weights (NASS and ERS) – Nick Schauer (NASS) and Dave Banker (ERS)
- 10:00-10:15 *Break*
- 10:15-11:15 Item quality control and treatments for non-response (NASS and ERS) - Richard Barton (NASS) and Mitch Morehart (ERS)
- 11:15-12:15 Statistical testing (NASS and ERS) –Bob Dubman (ERS) and Phil Kott (NASS)
- 12:15-12:30 Closing and adjournment

Researcher Responsibilities

- ▶ Serve Agency Mission
- ▶ Metadata documentation
 - Code
 - Information about variables
- ▶ Post research output
- ▶ Cite sources
- ▶ Evaluation and feedback

Accessing the enclave and promoting metadata documentation



Accessin

<https://enclave.norc.org>

*** don't forget it's https://, not http:// ***

Web Interface Log In - Windows Internet Explorer

https://enclave.norc.org/Citrix/AccessPlatform/auth/login.aspx

Web Interface Log In

NORC at the UNIVERSITY OF CHICAGO
Innovative solutions in research and technology

Log in

User name:

Password:

Log In

NORC Data Enclave

Authorized Users Only!

To log in, enter the credentials supplied by NORC, and then click Log In.

The message center will inform you on browser related technical issues. Note that you will first need to install the Citrix Client on you system (a download link will be provided)

Enter your user name and password.
The first time you access, you will need to change your password

User logged in

Citrix Access Platform - Windows Internet Explorer

https://enclave.norc.org/Citrix/AccessPlatform/site/default.aspx

File Edit View Favorites Tools Help

Citrix Access Platform

NORC

at the UNIVERSITY OF CHICAGO

Innovative solutions in research and technology

Data Enclave Desktops

Top Up

DE Desktop - Full Screen

Reconnect Disconnect Log Off

NORC Data Enclave

Data Enclave Desktops

Welcome to the Data Enclave. The Desktops box contains icons to launch the Data Enclave interface. Click an icon to launch an application. Click Refresh to view the latest applications. Click Settings to change your settings. Click a folder icon to display its contents. If you have problems using an application, please contact your NORC representative for more information.

Message Center

Once logged in, the system provides access to the data enclave through a virtual desktop. The full desktop will

Internet 100%

Enclave - Research

The Research Group portal page is displayed when the users logs in

DE Desktop Small - MetaFrame Presentation Server Client [SpeedScreen On]

Home - Brigham Young University - Windows Internet Explorer provided by NORC Data Enclave

http://www.enclave.local/byu/default.aspx

My Documents

byu Data

byu Portal

Data Enclave Portal

View All Site Content

Announcements

Collaborate

Share knowledge

- Shared Documents
- BYU Blog
- BYU Wiki

Get Support

- NORC Enclave
- MIST ATP Documentation
- Technical Support

Recycle Bin

Add new announcement

Team Discussion

Subject

Add new discussion

Shared Documents

Type Name Modified By

NIST National Institute of Standards and Technology

Tasks

Title Assigned To

Add new item

Trusted sites 100%

Start Home - Brigham Youn...

This shortcut provides access to the research group data and documentation folders prepared by the enclave along with the group shared directories

These two shortcuts provides access to the research group and NORC portals

Note that the desktop shortcuts are also accessible through the Start menu under the Data Enclave program group

Getting Work Done



Getting Work Done

The screenshot shows a Windows desktop environment. The desktop background is blue with several icons on the left: My Documents, My Home, byu Data, byu Portal, and Data Enclave Portal. The taskbar at the bottom shows the Start button, a taskbar button for 'Home - Brigham Young U...', and a 'Log Off' button. The Start menu is open, displaying a list of programs including Notepad, Metadata Editor, SAS 9.1 (English), StataSE 9, Microsoft Office, Data Enclave Portal, and All Programs. The 'All Programs' folder is expanded, showing sub-folders like Accessories, Administrative Tools, Adobe, Data Enclave, Microdata Management Toolkit, Microsoft Office, Nesstar, SAS, Startup, Stata 9, and StatTransfer 9. The 'Stata 9' folder is expanded, showing 'StataSE 9'. The main window is a web browser displaying the Brigham Young University website. The browser title bar reads 'Home - Brigham Young University - Windows Internet Explorer provided by NORC Data Enclave'. The address bar shows 'http://www.enclave.local/byu/default.aspx'. The website content includes a navigation menu with 'Home', 'BYU Blog', and 'BYU Wiki'. A 'Collaborate' sidebar lists 'Announcements', 'Calendar', 'Team Discussion', and 'Tasks'. The main content area features an 'Announcements' section with a post titled 'Company Size Assymetry vs. Motivation to Form R&D Alliances' by Tim Mulcahy, dated 5/14/2007 1:56 PM. A 'Tasks' section below it states 'There are no items to show in this view of the "Tasks" list. To create a new item, click "Add new item" below.' A 'NIST National Institute of Standards and Technology' logo is also visible on the right side of the page.

You can access your application like you normally do on a regular computer using the Start menu. Note that administrative functionalities have been disabled for security reasons.

Software available in the enclave

- ▶ Stata/SE 9.2
- ▶ StatTransfer 9
- ▶ SAS v9.1
- ▶ Microsoft office 2007
- ▶ Adobe PDF Reader
- ▶ IHSN Microdata Management Toolkit
/ Nesstar Publisher
(upon request, selected users only)



Collaborating: Research Group Portal

Home - Brigham Young University - Windows Internet Explorer provided by NORC Data Enclave

http://www.enclave.local/byu/default.aspx

File Edit View Favorites Tools Help

Home - Brigham Young University

Brigham Young University Welcome Pascal Heus

BYU Brigham Young University This Site

Home BYU Blog BYU Wiki Site Actions

View All Site Content

Collaborate

- Announcements
- Calendar
- Team Discussion
- Tasks

Share knowledge

- Shared Documents
- BYU Blog
- BYU Wiki

Get Support

- NORC Enclave
- NIST ATP Documentation
- Technical Support

Recycle Bin

Announcements

Company Size Assymetry vs. Motivation to Form R&D Alliances 5/14/2007 1:56 PM
by Tim Mulcahy

BYU researchers are investigating whether motivations for forming R&D alliances are different for large vs. small firms and whether the performance outcomes are different. They expect that small firms will be much more likely to commercialize technology...

Add new announcement

Team Discussion

Subject

There are no items to show in this view of the "Team Discussion" discussion board. To create a new item, click "Add new discussion" below.

Add new discussion

Shared Documents

Type	Name	Modified By
Folder	Files for export	Pascal Heus
Folder	Scripts	Pascal Heus
Folder	References	Pascal Heus

Add new document

NIST
National Institute of Standards and Technology

Tasks

Title Assigned To

There are no items to show in this view of the "Tasks" list. To create a new item, click "Add new item" below.

Add new item

Done Trusted sites 100%

Collaborating: Research Group Portal

[View All Site Content](#)

Collaborate

- Announcements
- Calendar
- Team Discussion
- Tasks

Share knowledge

- Shared Documents
- BYU Blog
- BYU Wiki

Get Support

- NORC Enclave
- NIST ATP
Documentation
- Technical Support

ORGANIZE
&
EXCHANGE

SHARE FILES
CAPTURE EVENTS
& KNOWLEDGE

ACCESS TO GLOBAL
INFORMATION
AND SUPPORT

Overview

- ▶ What is metadata and why is it important
- ▶ What can metadata do
- ▶ Guide to creating metadata
- ▶ Creating metadata in the enclave

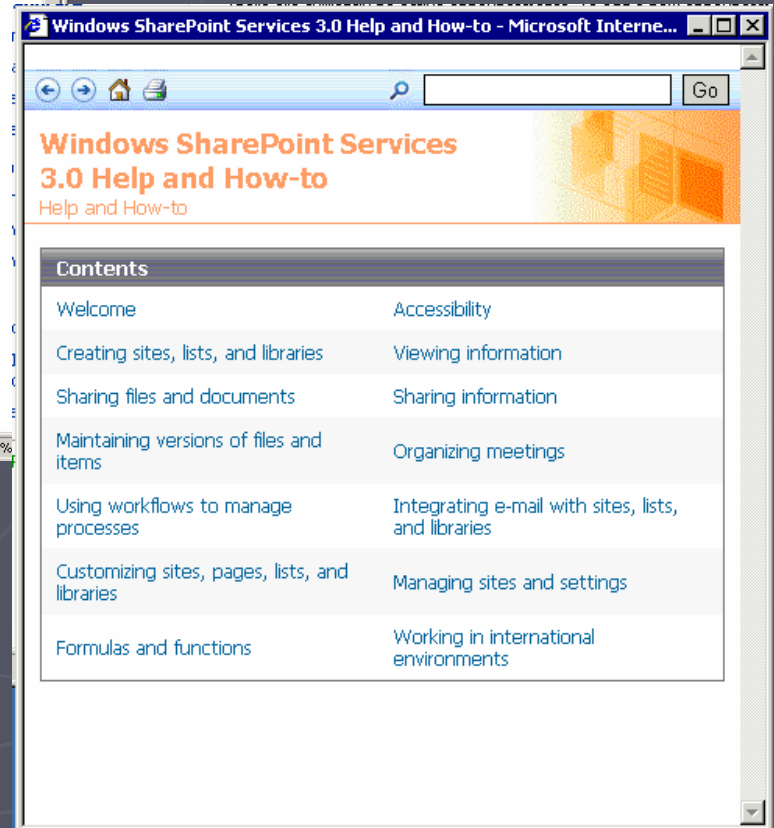
Creating incentives for metadata documentation for researchers

- ▶ Reduce cost of analysis
 - Create collaborative environment, with wiki and FAQ's
 - Capture research process through blogs (replication standard!)
- ▶ Facilitate publication of results and increase visibility of work
 - Facilitate reporting, citations, etc.
 - Facilitate reusability / extend the research
 - Compare results
- ▶ Create value for future research
 - Integrate research results in the survey knowledge

Using the SharePoint based portal



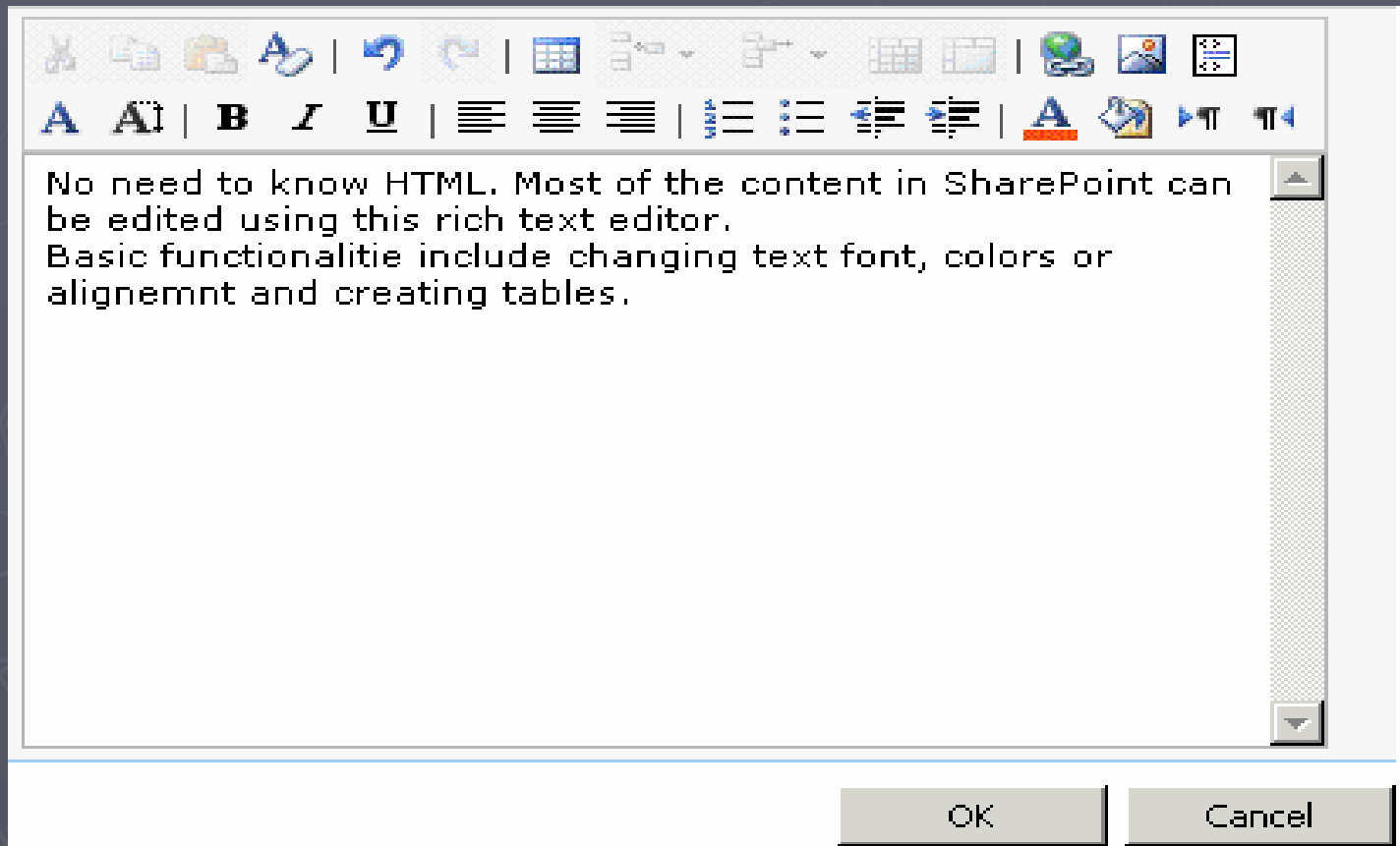
Use the HELP!



But be aware that

- (1) not all documented functionalities are available
- (2) Some functions require administrative access

Editing content



and exchanging ideas

Announcements

New data available from USDA! NEW 6/22/2007 12:02 AM
by Heus-Pascal

The 2004 ARMS survey is now available in the NORC data enclave. Contact your representative for further information.

[Add new announcement](#)

Calendar: New Item

Attach File

enclave training

NORC Washington DC

6/27/2007 9 AM 00

6/27/2007 1 PM 00

enclave training for new enclave users. Introduction to data enclave and metadata.

Make this an all-day activity that doesn't start or end at a specific hour.

Recurrence Make this a repeating event.

Workspace Use a Meeting Workspace to organize attendees, agendas, documents, minutes, and other details for this event.

Use the enclave announcement, tasks / todo, and calendar to distribute and organize the research work

Tasks: Run regression of ent

New Item | Edit Item | Delete Item

Title	Run regression of enterprise data
Priority	(2) Normal
Status	In Progress
% Complete	45%
Assigned To	Heus-Pascal
Description	Subset the data to small enterprise and run regression on funding amounts
Start Date	6/22/2007
Due Date	

Created at 6/21/2007 11:59 PM by Heus-Pascal
Last modified at 6/21/2007 11:59 PM by Heus-Pascal

Close

Use the discussion groups to exchange ideas, submit questions, etc

Brigham Young University > Team Discussion

Team Discussion

Use the Team Discussion list to hold newsgroup-style

New Actions Settings

Discussion
Create a new discussion topic.



Team Discussion

Use the Team Discussion list to hold newsgroup-style discussions on topics relevant to your team.

Actions Settings View: Flat

Posted By Post

Started: 6/21/2007 11:48 PM View Properties Reply

Heus-Pascal

How do I use SharePoint?
No need to know HTML. Most of the content in SharePoint can be edited using this rich text editor. Basic functionalities include changing text font, colors or alignment and creating tables.

Posted: 6/21/2007 11:48 PM View Properties Reply

Heus-Pascal

Yes, I love this little editor!
[Show Quoted Messages](#)

Using the blog to capture research events

View All Site Content

Categories

- Category 1
- Category 2
- Category 3

Add new category

Other Blogs

There are no items in this list.

Add new link

Links

- Photos
- Archive
- Archive (Calendar)

Add new link

RSS Feed

Brigham Young University > BYU Blog

5/8/2007

Welcome to your Blog!

To begin using your site, click **Create a Post** under Admin Links to the right.

What is a Blog?

A Blog is a site designed to help you share information. Blogs can be used as news sites, journals, diaries, team sites, and more. It is your place on the World Wide Web.

Blogs are typically displayed in reverse chronological order (newest entries first), and consist of frequent short postings. With this Blog, it is also possible for your site visitors to comment on your postings.

In business, Blogs can be used as a team communication tool. Keep team members in touch by providing a central place for links, relevant news, and even gossip.

Posted at 1:19 PM by Heus-Pascal | [Permalink](#) | [Email this Post](#) | [Comments \(0\)](#)

Admin Links

- Create a post
- Manage posts
- Manage comments
- All content
- Set blog permissions
- Launch blog program to post

- ▶ Research is an iterative, evolving process
- ▶ Capturing ideas and milestone is crucial
- ▶ Personal logs have often been used in the past
- ▶ Blogs is today's version of it

Brigham Young University > BYU Wiki > Wiki Pages > Home

Home

[Edit](#) | [History](#) | [Incoming Links](#)

[View All Site Content](#)

Wiki Pages

- Home
- How To Use This Wiki Site

[Recycle Bin](#)

Recent Changes

- Home
- How To Use This Wiki Site

[View All Pages](#)

Welcome to your wiki site!

You can get started and add content to this page by clicking **Edit** at the top of this page, or you can learn more about wiki sites by clicking **How to use this wiki site** in the Quick Launch.

What is a wiki site?

Wikiwiki means quick in Hawaiian. A wiki site is a Web site in which users can easily edit any page. The site grows organically by linking existing pages together or by creating links to new pages. If a user finds a link to an uncreated page, he or she can follow the link and create the page.

In business environments, a wiki site provides a low-maintenance way to record knowledge. Information that is usually traded in e-mail messages, gleaned from hallway conversations, or written on paper can instead be recorded in a wiki site, in context with similar knowledge.

Other example uses of wiki sites include brainstorming ideas, collaborating on designs, creating an instruction guide, gathering data from the field, tracking call center knowledge, and building an encyclopedia of knowledge.

- ▶ Familiar with Wikipedia?
- ▶ A wiki is a shared web site but does not require programming skills to maintain
- ▶ Multiple authors can add, remove, and edit content (mass authoring).
- ▶ Knowledge grows across time based in community contributions
- ▶ Pages automatically link to each other page on

Summary

- ▶ Goal: To promote access to sensitive micro data while protecting confidentiality
- ▶ Benefits:
 - Secure, low-cost approach to microdata access
 - Archiving, Indexing, and Curation of Micro-data
 - Applicable and Customizable to agency needs and requirements