



NORC | Data Enclave



NORC Data Enclave

IASSIST 2010 – Session F4: Providing Secure Access to Sensitive Data

Tim Mulcahy – NORC @ University of Chicago

Mulcahy-Tim@norc.org / <http://www.norc.org/DataEnclave>

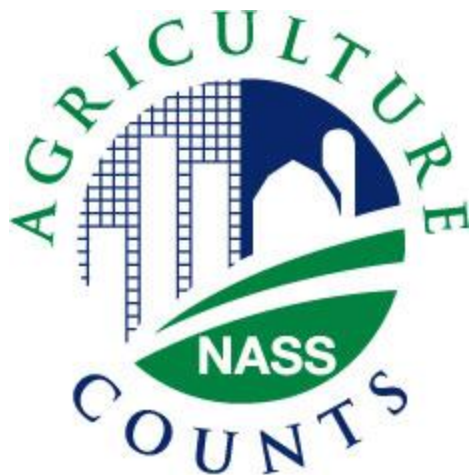
Pascal Heus, Metadata Technology North America

pascal.heus@metadatatechnology.com / <http://www.metadatatechnology.com>



Emergence of the Remote Data Enclave

- Enclave began in July 2006
- Current sponsors
 - US Department of Commerce (NIST-TIP)
 - US Department of Agriculture (ERS/NASS)
 - Ewing Marion Kauffman Foundation
 - National Science Foundation
 - Annie E. Casey Foundation



Portfolio Approach to Secure Data Access

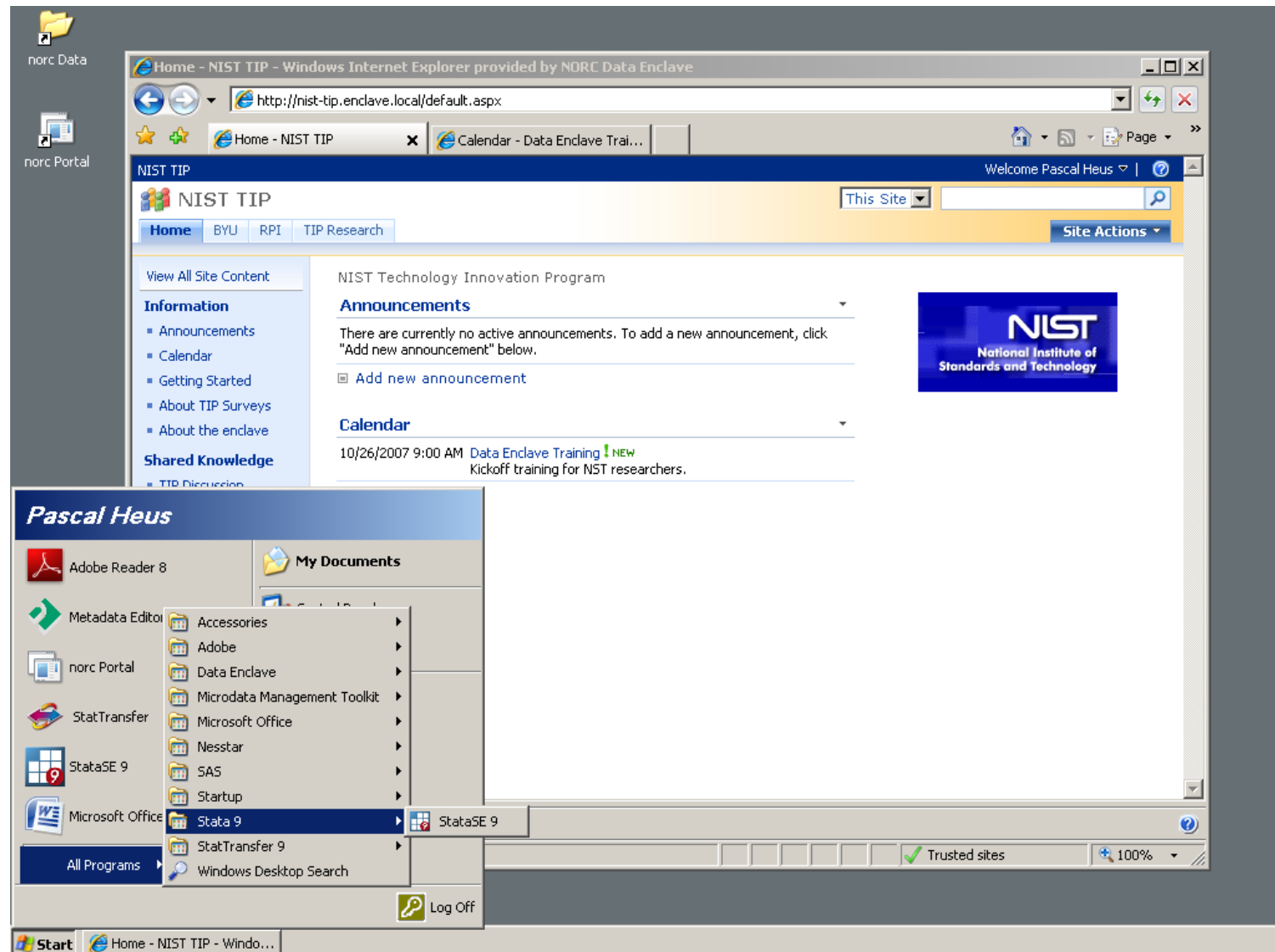
- Set of protection features that can be customized for each data producer
- Legal
 - Data user agreement, NDA, enclave staff certification
- Educational / Training
 - Mandatory (how to use, disclosure, data specific)
 - Delivery: virtual, on site, NORC, workshops
- Statistical
 - Data available in DE is cleared for research (producer specific)
 - Extensive data disclosure review process on output
- Data Protection / Operational / Technological
 - Approved industry standard technology (Citrix), encryption, client/server security, system monitoring
 - Compliance with DOC IT SPP, Section 6.5.2, FISMA, FIPS and other NIST

Available Datasets

- Department of Commerce, National Institute of Standards and Technology, Technology Innovation Program
 - » ATP Survey of Joint Ventures
 - » ATP Survey of Applicants
 - » Business Reporting Survey Series
- Department of Agriculture, Economic Research Service, National Agriculture Statistical Service
 - » Agricultural Resource Management Survey (ARMS)
- Kauffman Foundation
 - » Kauffman Firm Survey
- Annie E. Casey Foundation
 - » Making Connections Survey
- National Science Foundation
 - » Survey of Earned Doctorates
 - » Survey of Doctoral Recipients
- Center for MediCare and Medicaid
 - » Claims data



DE Desktop



Tools Available in the Enclave

- Stata/SE 10.0
- StatTransfer 9 (selected users)
- SAS v9.2
- “R” Project for Statistical Computing
- MATLAB
- LimDep / NLogit
- Microsoft office 2007
- LISREL
- Adobe PDF Reader
- IHSN Microdata Management Toolkit /
Nesstar Publisher
(upon request, selected users only)



DE Innovative Research

General approach

- We currently typically use DDI2 based solutions
- Moving towards next generation DDI3 based
 - Enhanced functionalities, reusability, meets broader needs
- We are looking into
 - New DDI3 based products
 - Tools that work with both DDI 2 and DDI 3
 - Conversion tools
- and building tools that
 - Go beyond the traditional use of DDI (archive/documentation) and fully leverage on specification and technology
 - Meet Data Enclave and researchers needs
 - Facilitate reuse and common metadata management



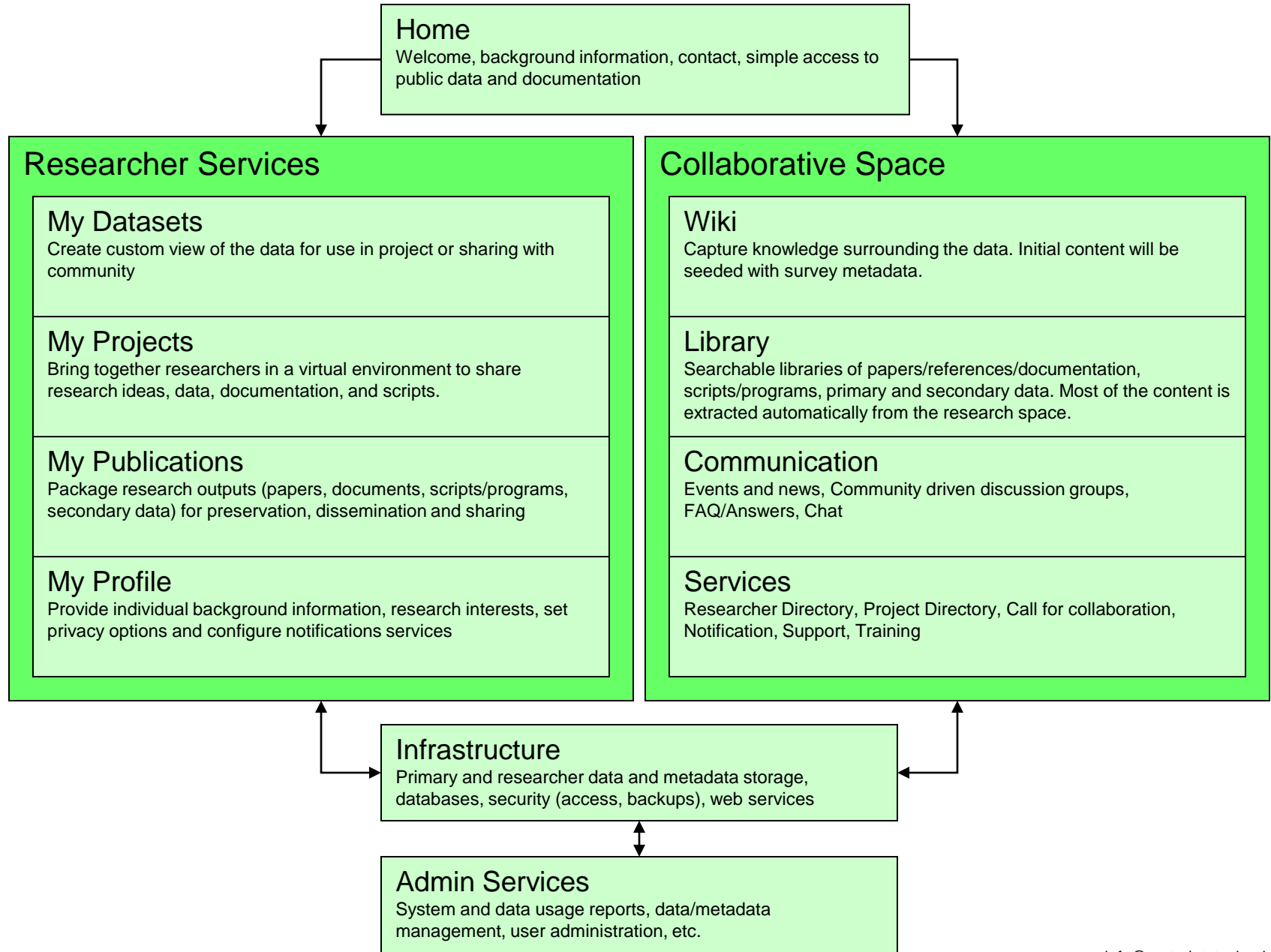
Coodle Project

- Need
 - Capture researcher knowledge and understand data usage
- DocuStat
 - Capture metadata in statistical scripts/program source code
 - Document during research processes
 - Use IT standard practices for source code tagging
 - @author, @version, @reference, @...
 - → Knowledge capture, script indexing, automated reports
- Coodle
 - Build a searchable repository of scripts (for users)
 - Perform DDI metadata driven search for variable occurrence in source code
 - Report on variable / data usage
- Status: prototype ready, scaling up to production

DEER Project

- Need
 - Management of disclosure review processes
 - Facilitate export of files out of environment
- Phase 1
 - Simple request packaging tool for user
 - Request administration tool for enclave manager
 - Support workflow (submit, enclave review, producer, review, delivery)
 - Institutional archiving and audit
- Phase 2
 - Link with DDI metadata to identify data sources and variables
 - Facilitates review processes and understanding of usage
- Status: initiating phase 1

Virtual Centre for Collaborative Research



Thin Client Technology

