Metadata Standards and Technology Development for the NSF Survey of Earned Doctorates

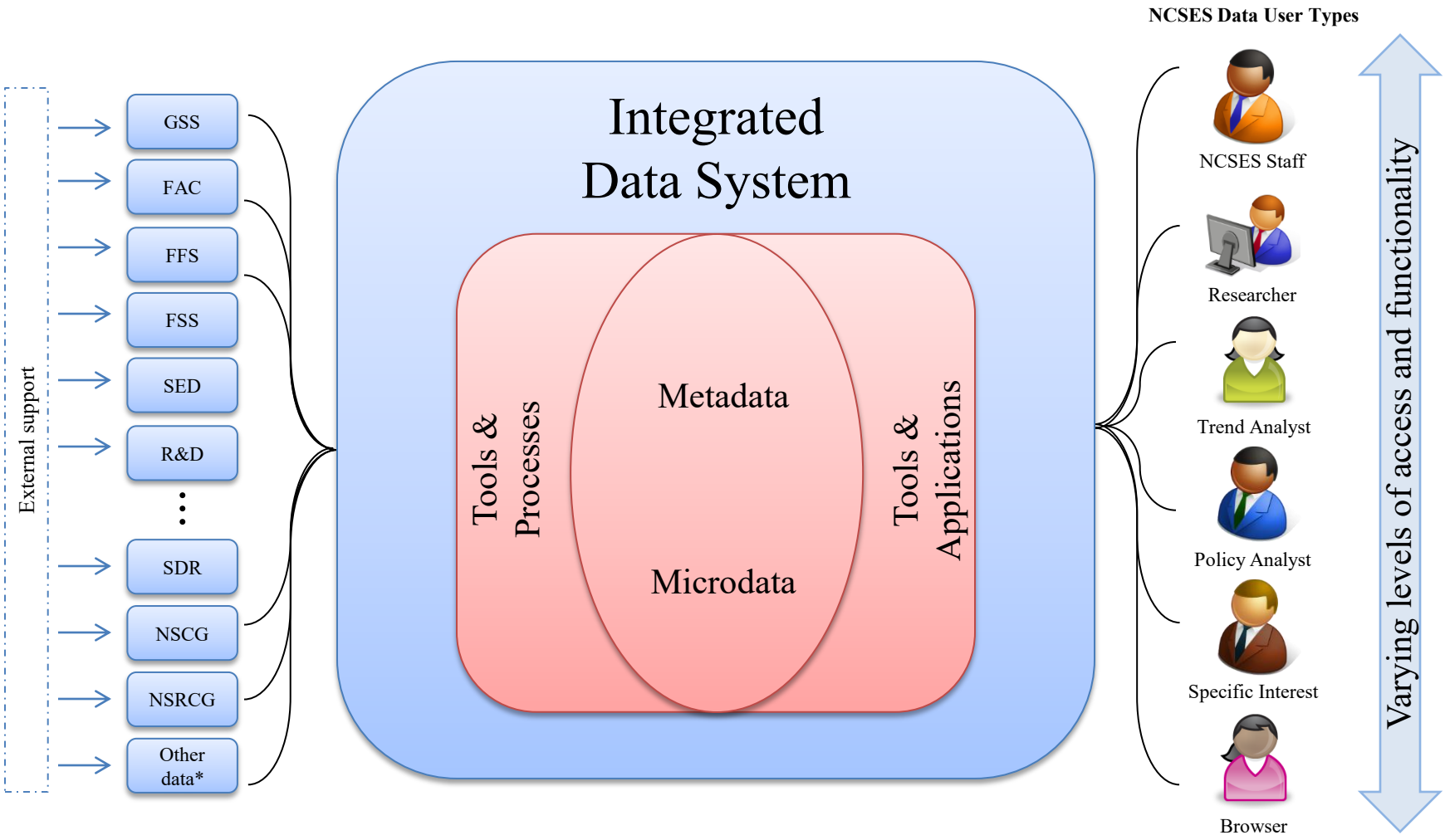
Kimberly Noonan (NSF NCSES)
Pascal Heus (MTNA)
Tim Mulcahy (NORC)

November 5, 2013

National Center for Science and Engineering Statistics (NCSES)

- A federal statistical agency within NSF
- Charged with the mission to provide a central clearinghouse for the collection, interpretation, and analysis of data on scientific and engineering resources
- 12 periodic data collections covering science and engineering
  - Research and Development
  - Education
  - Workforce
- Over 7 contracts for external support
- Building a central data system to store, maintain, and disseminate survey data in a faster, more flexible way

NCSES Data System



* Other data used regularly in NCSES publications

NCSES Data Delivery

- Develop data delivery requirements for all survey microdata and metadata
- Ensure comprehensive documentation
- Standardize delivery formats
- Adopt metadata standards
  - Data Documentation Initiative (DDI)
  - Globally recommended practices
  - Industry standards and technologies
- Automate data processing

NSF Metadata Project: SED a case study

Objectives

- Capture comprehensive survey metadata, in DDI format
- Automate generation of essential documentation, standard reports
- Generate delivery package compatible with the NCSES data systems
- Case study with Survey of Earned Doctorates (SED)
  - Extend to other NCSES surveys

Team

- **NSF NCSES**
  - Building next generation data system and management framework
- **NORC SED team**
  - Survey contractor, years of survey specific knowledge
- **Metadata Technology North America (MTNA)**
  - Domain and technology experts in statistical data management and its challenges

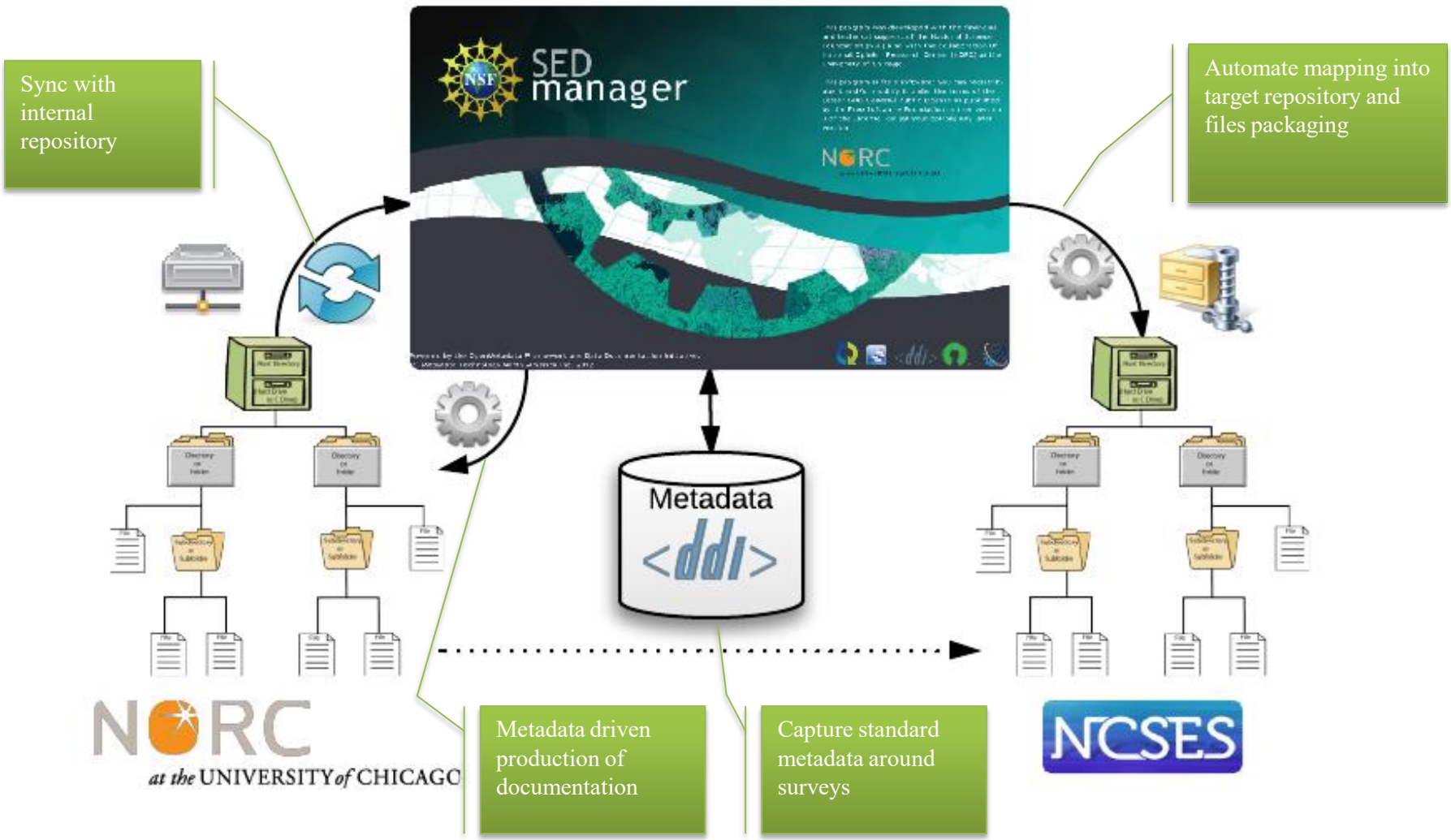
NSF Survey of Earned Doctorates

- **Began in 1958**
- **Annual survey**
- **All individuals receiving research doctoral degrees from accredited U.S. Institutions**
- **Results used to assess characteristics and trends in doctorate education and degrees**
- **Survey is currently conducted by NORC**
- **NCSES disseminates data, reports and documentation**
- **SED sponsors**
  - National Science Foundation
  - National Institutes of Health
  - US Department of Agriculture
  - Department of Education
  - National Endowment for the Humanities
  - National Aeronautics & Space Administration

SED Metadata Project Plan

- **Define SED metadata**
  - Assess current situation, file inventory
  - Capture comprehensive survey metadata
- **Develop SED metadata schema**
  - Develop metadata model
    - Based on metadata standards
    - Aligned with NCSES data system and dissemination needs
- **Prepare metadata for SED 2011**
  - Develop software tool, extend existing MTNA open source application
- **Automate SED metadata preparation for 2008, 2009, 2010**
  - Extend to additional survey cycles
- **Recommend maintenance and future steps**
  - Lessons learned
  - Next steps

SED NSF Manager



NSF SED Manager

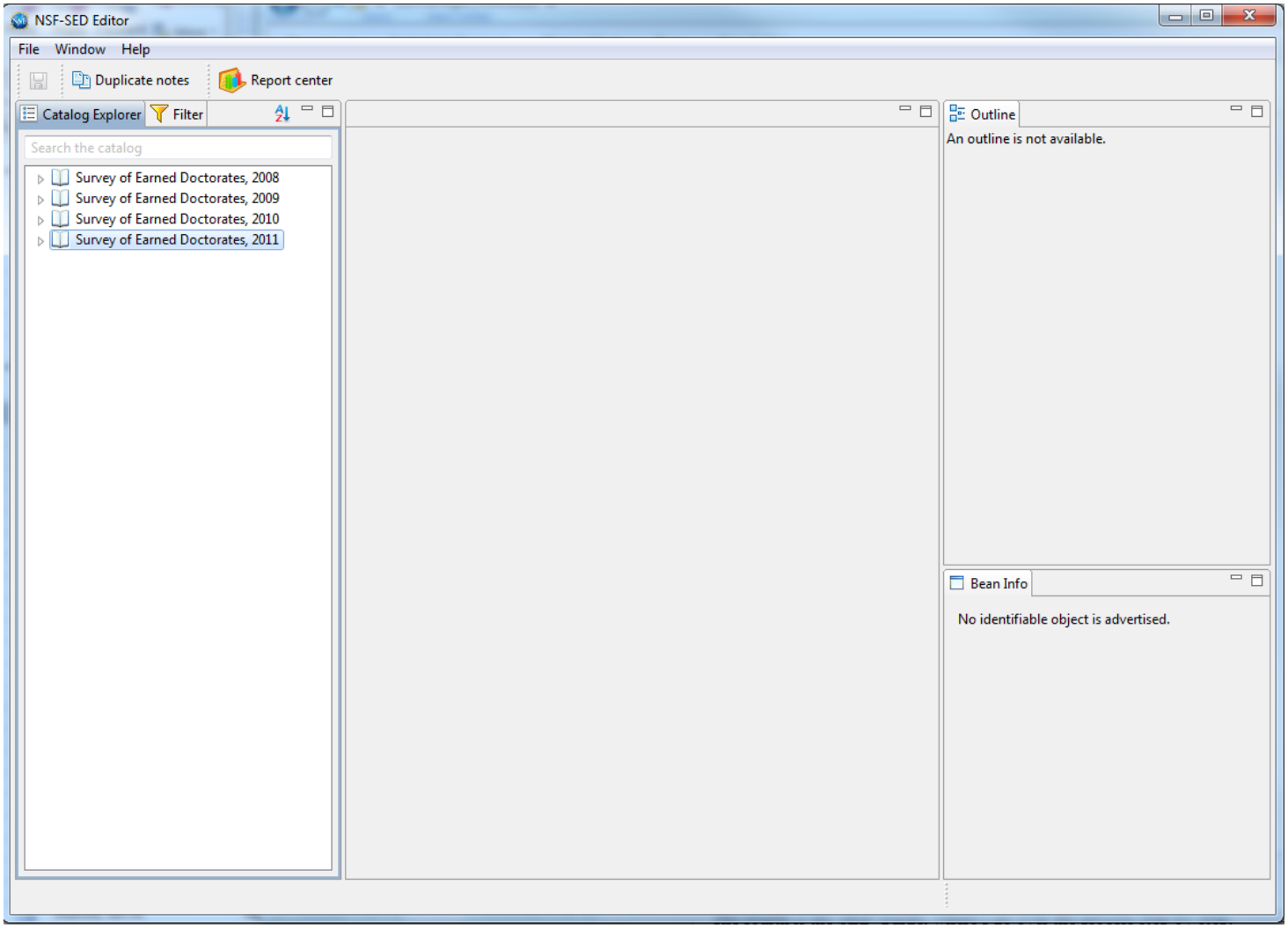
Features:

- **Catalog Explorer**
  - Import/create survey
- **Metadata Editors**
  - Survey, Questionnaire, Classification, Variables, Data, Documentation, Notes,
  - Repository packaging module
- **Report Center**
  - Codebook
  - Comparison reports, e.g. codebook comparison
  - Custom reports

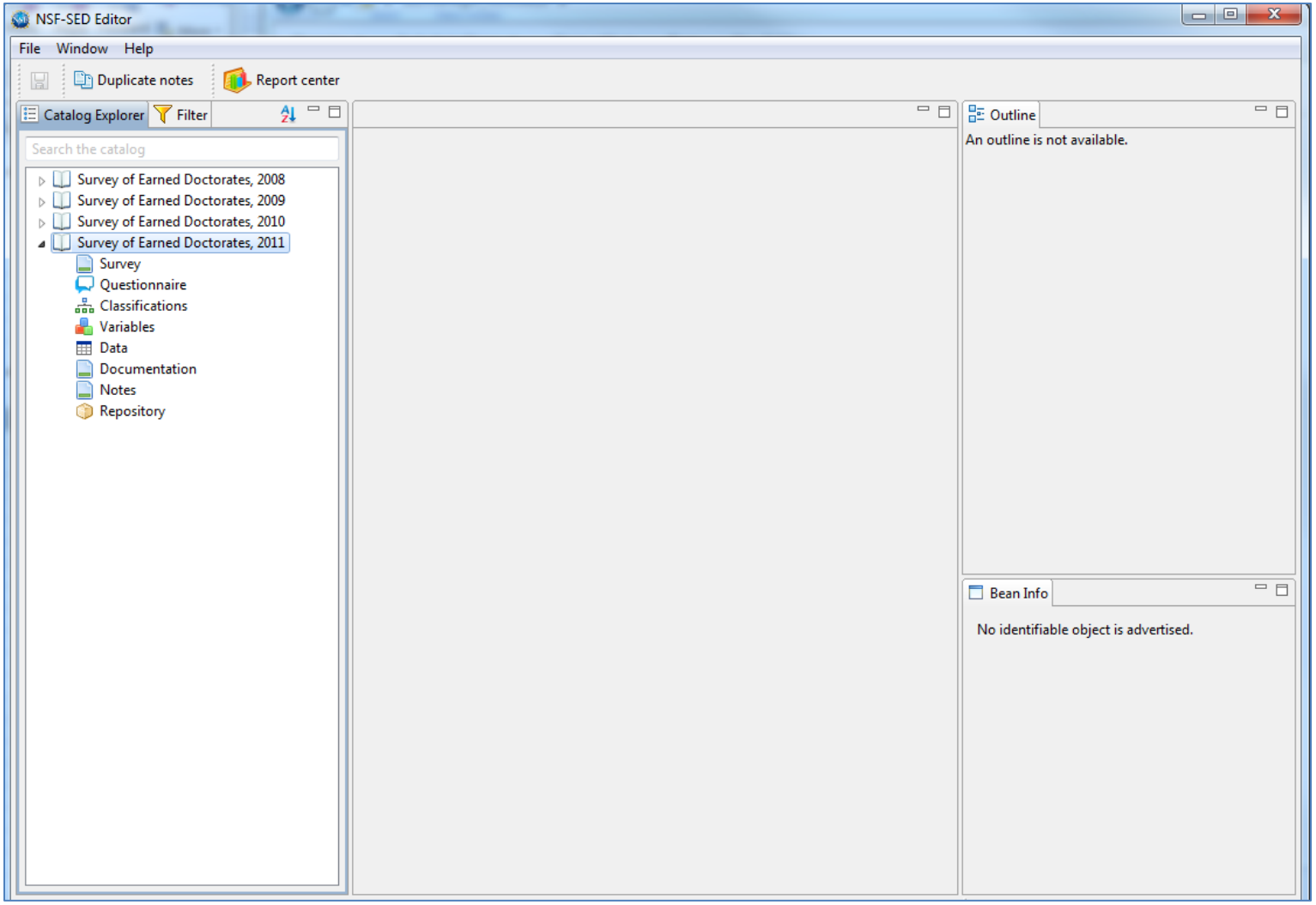
Benefits:

- Metadata are standardized
- Metadata are DDI compliant
- Metadata are automatically captured

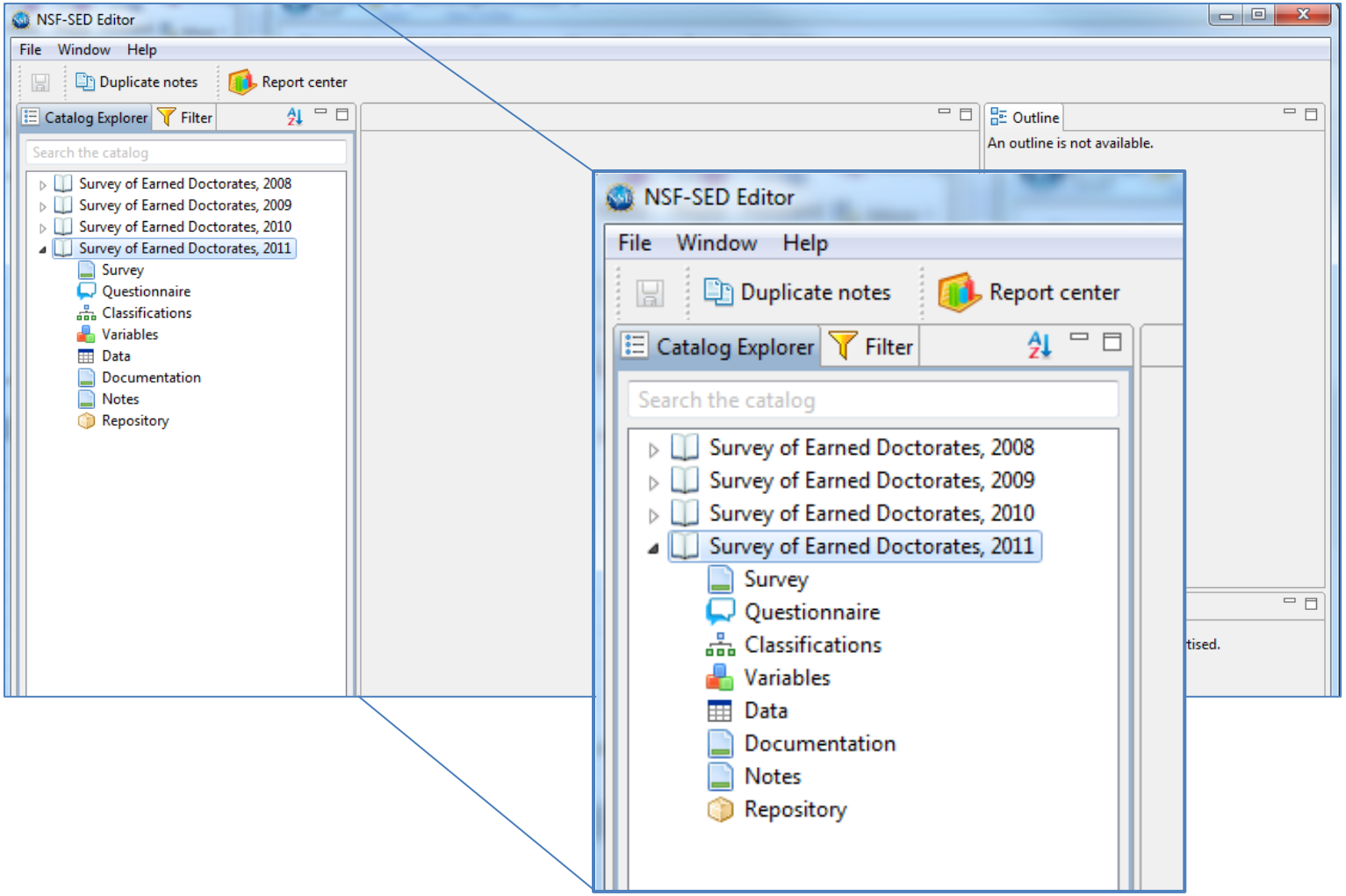
Catalog Explorer



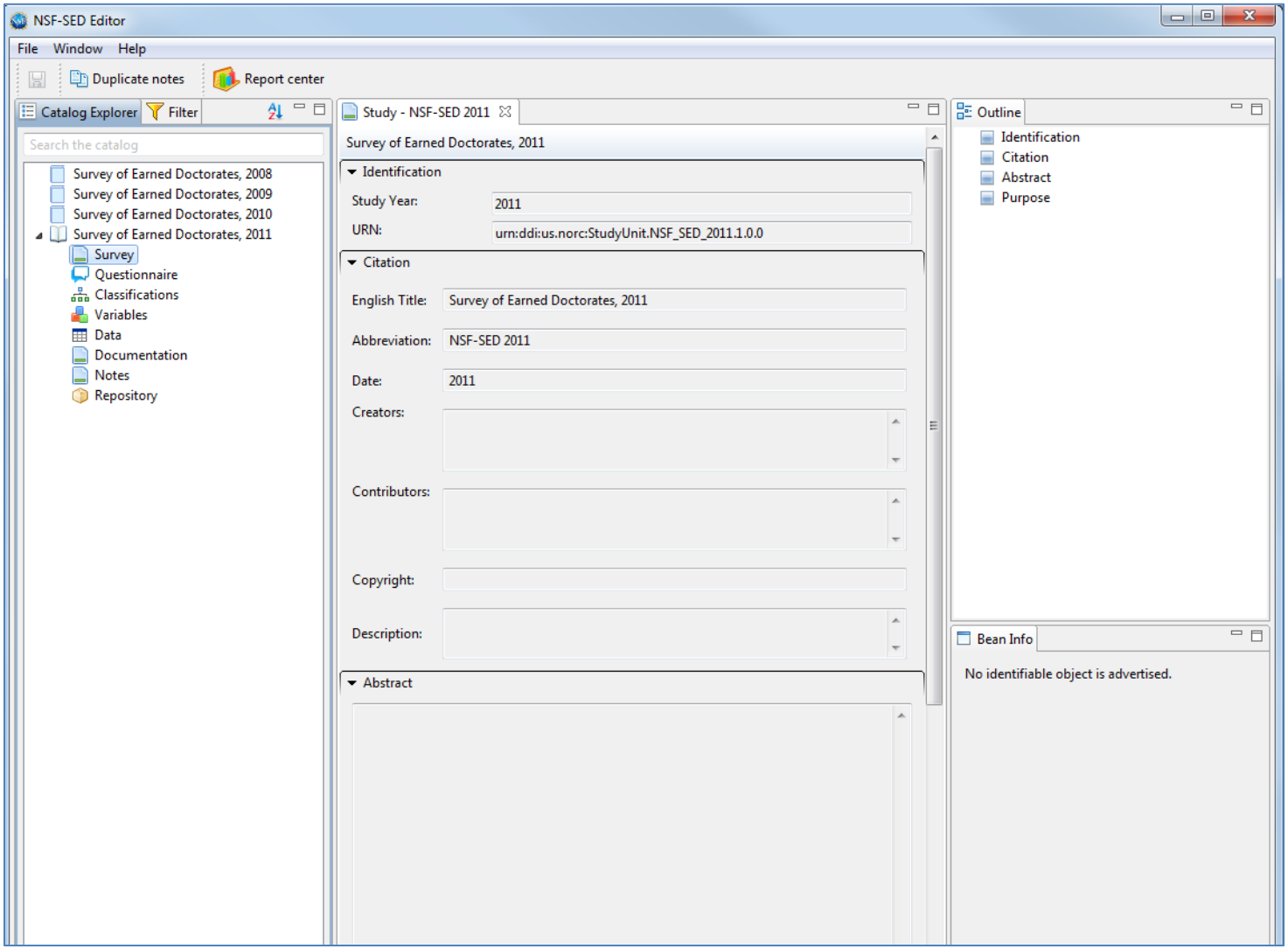
Metadata Editors



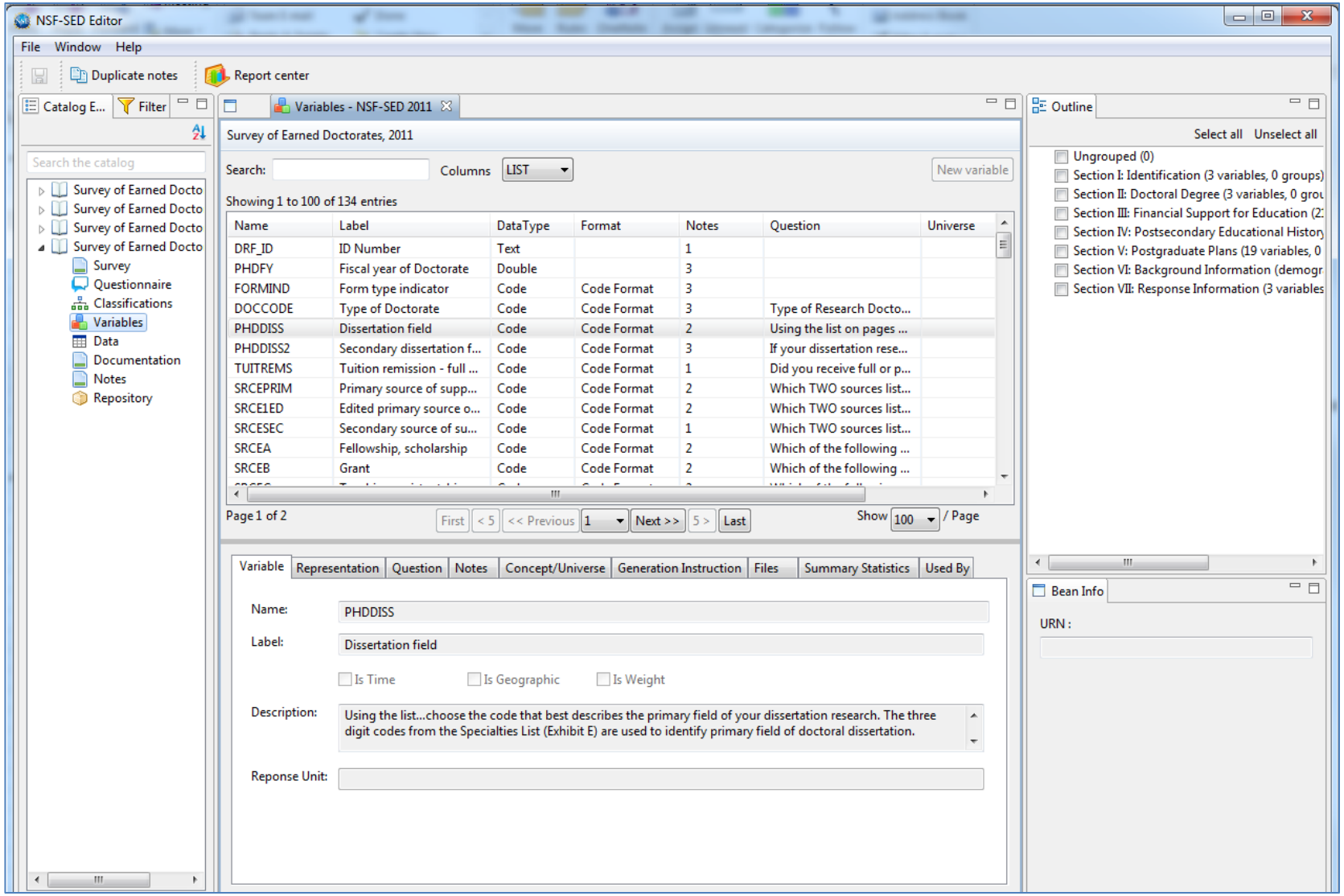
Metadata Editors



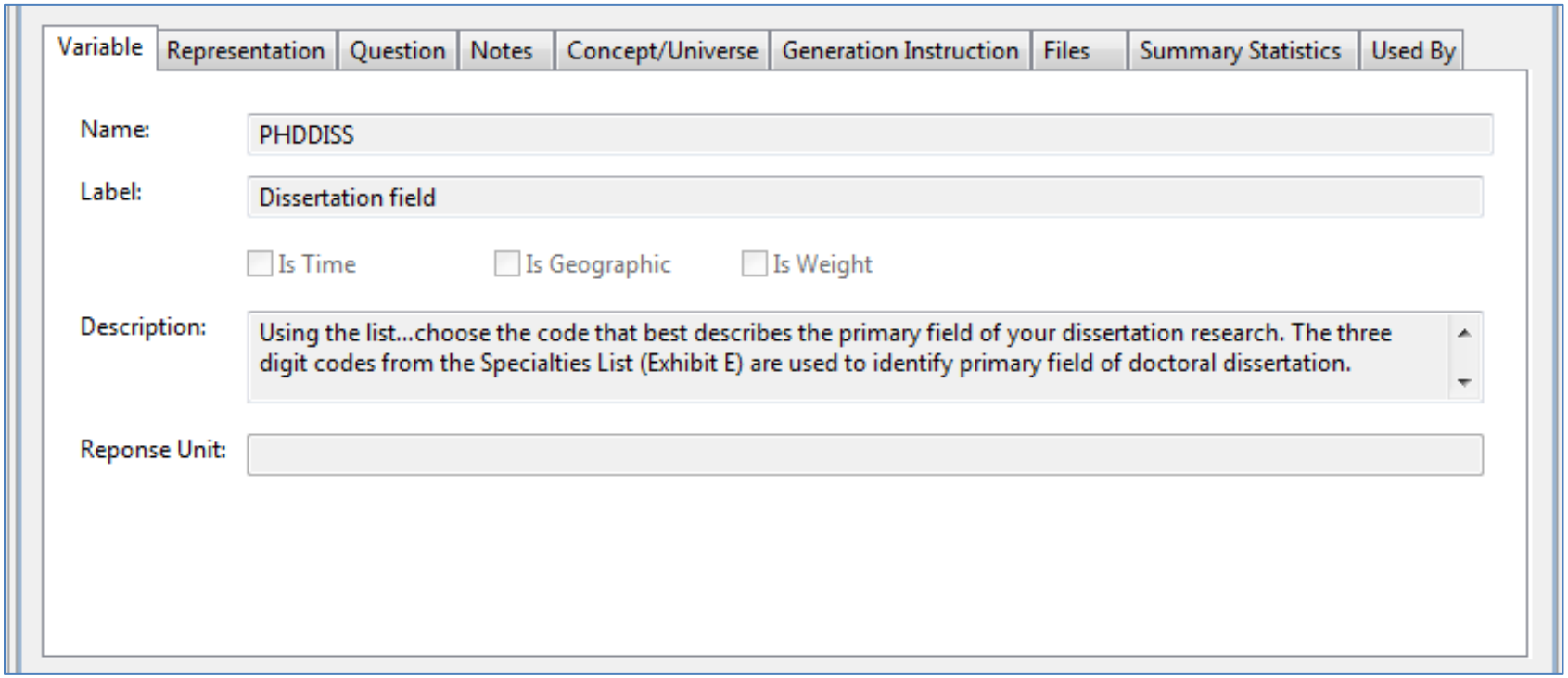
Survey Editor



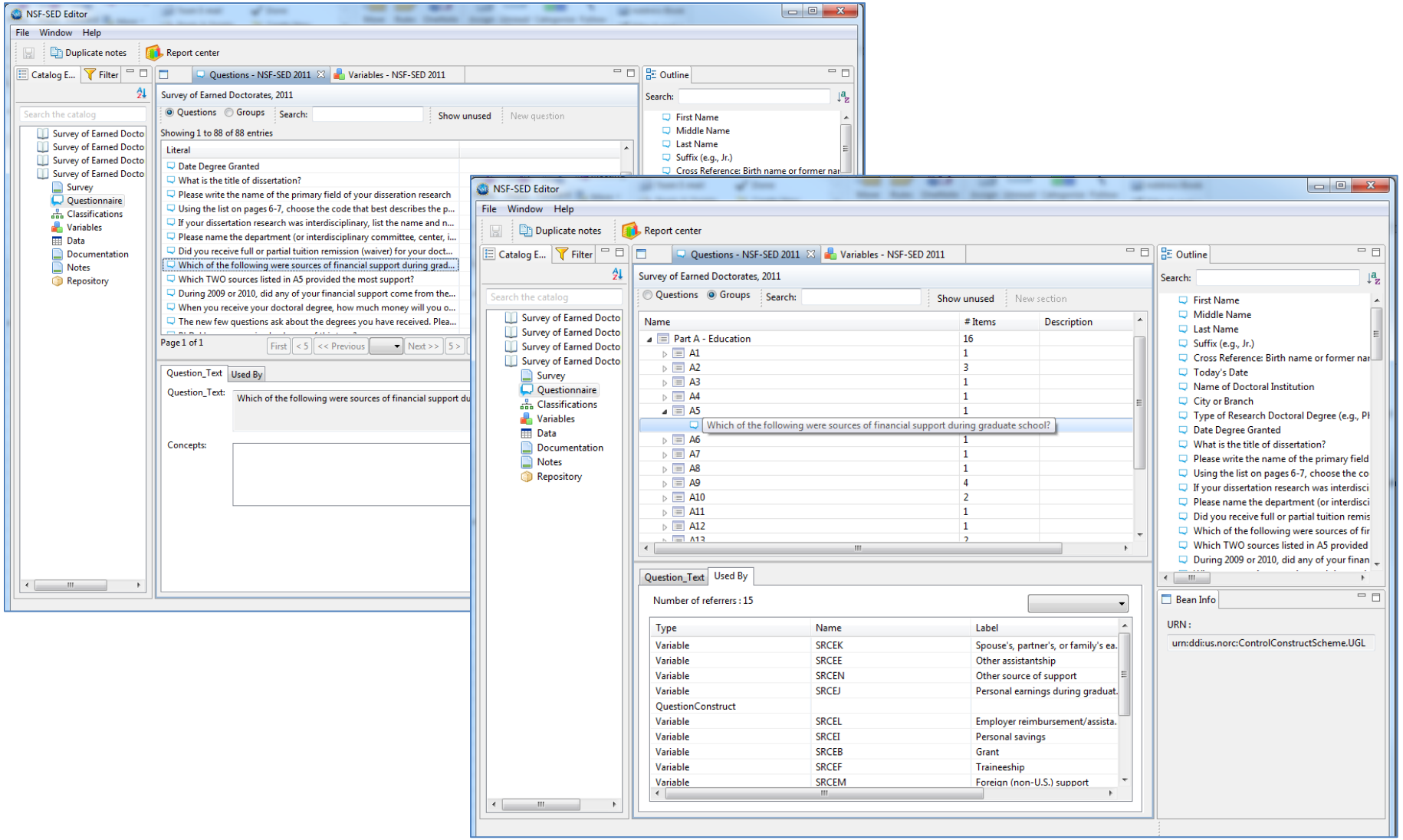
Variable Editor



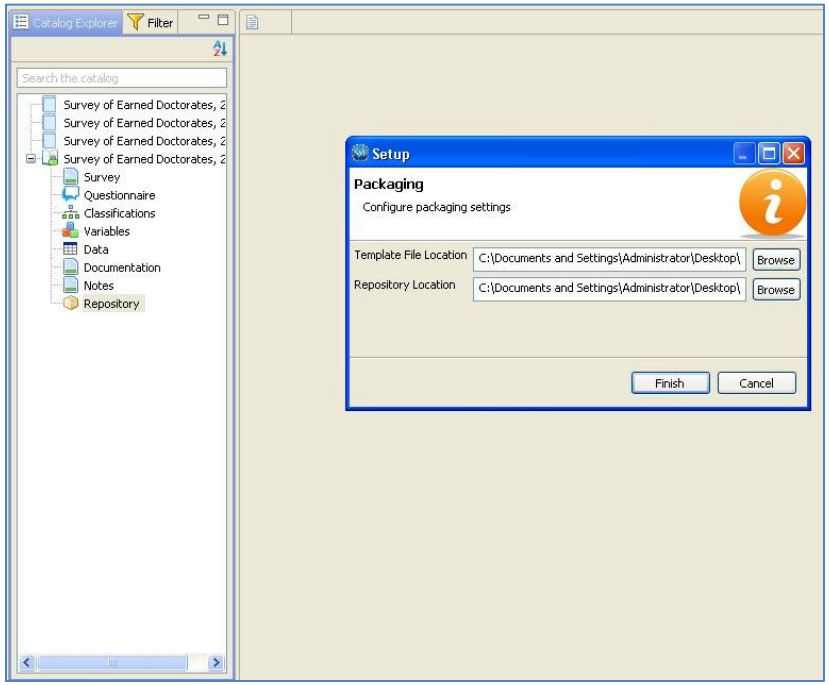
Variable Editor

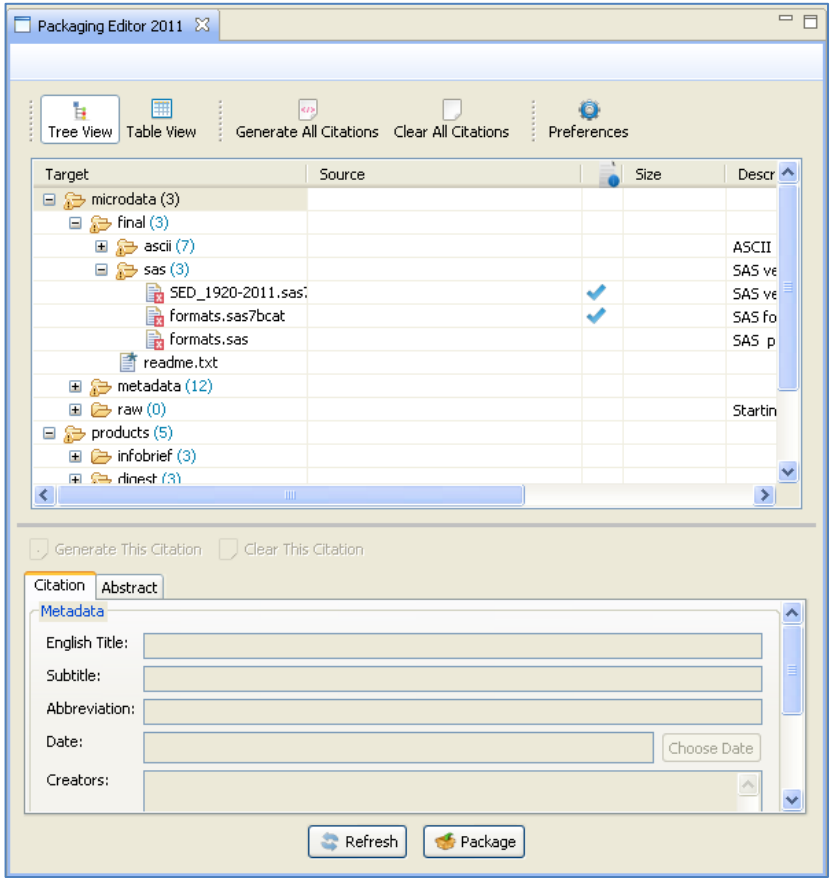


Questionnaire Editor

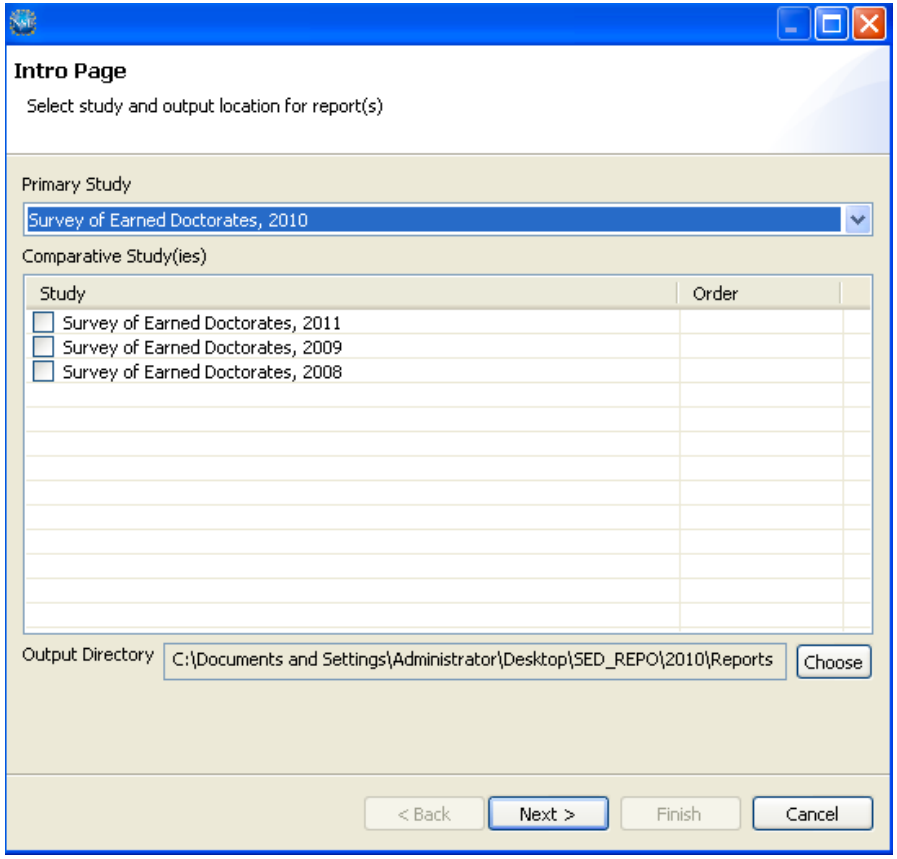


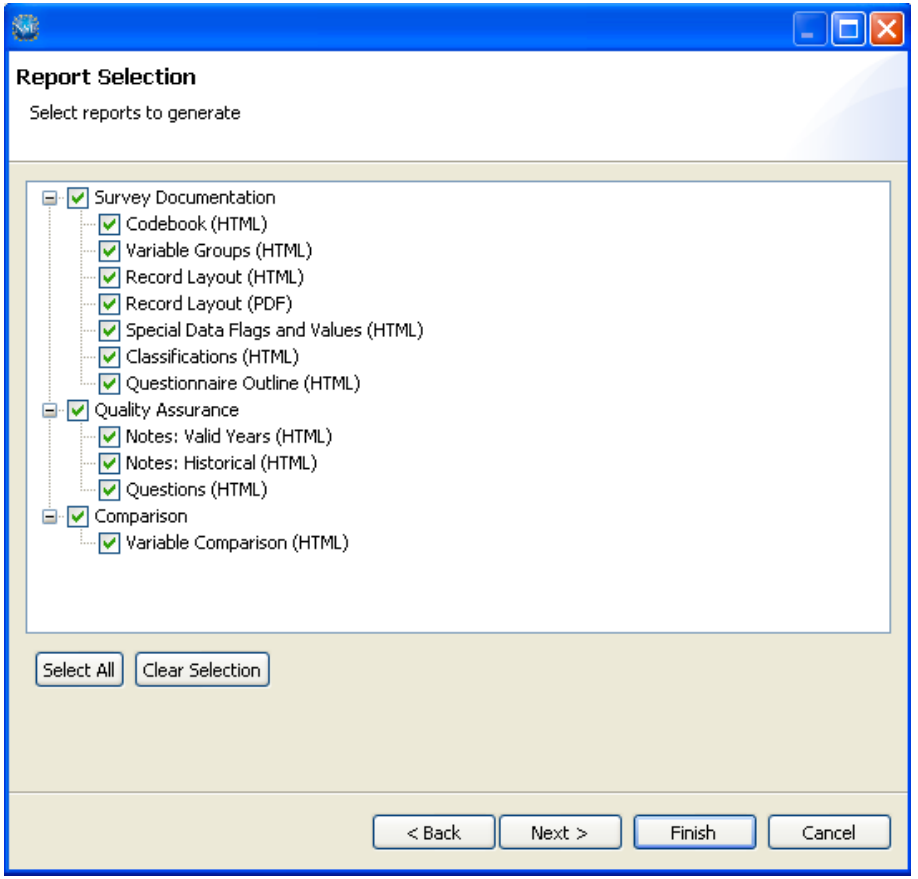
Repository Editor





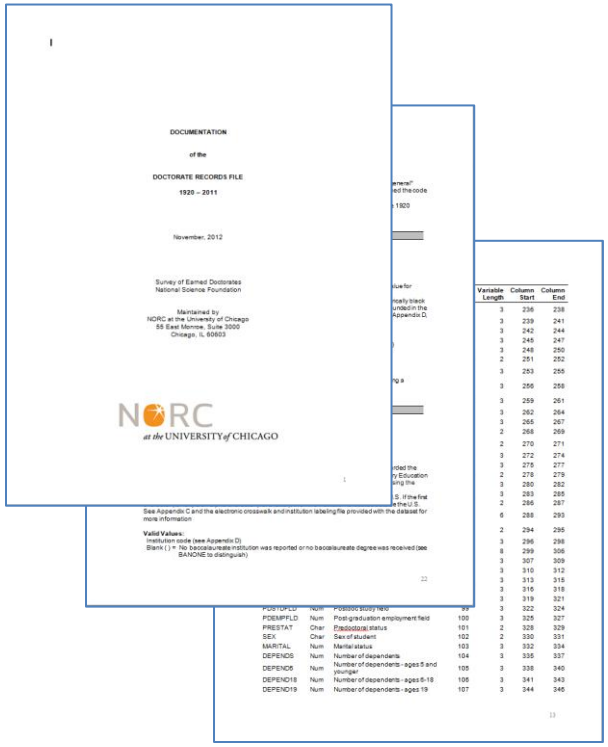
Report Center



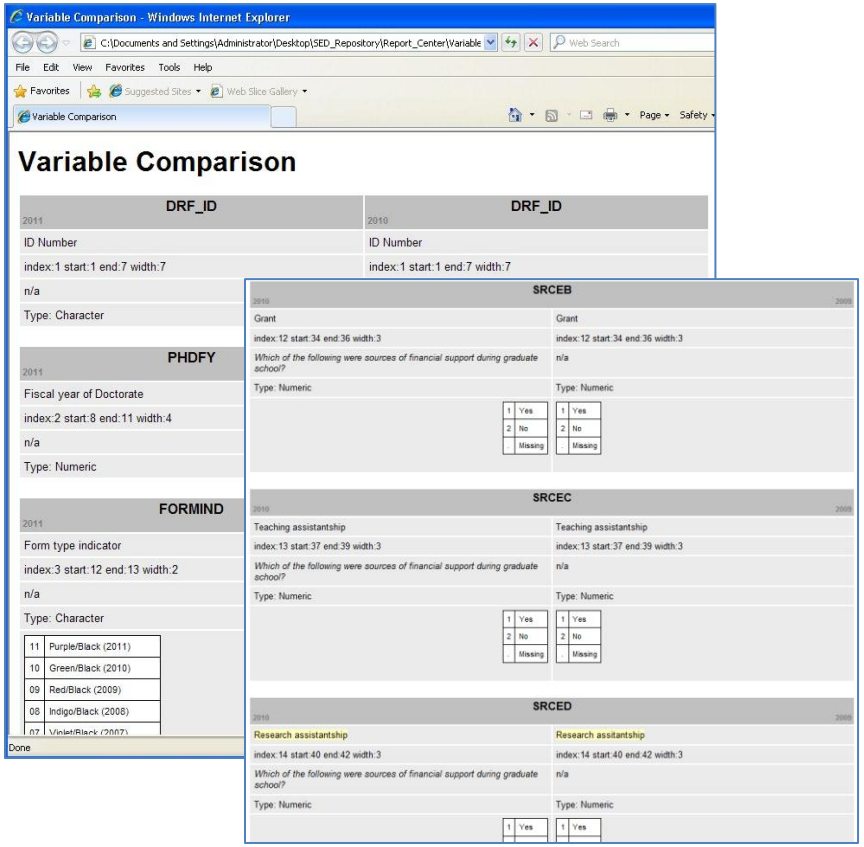


Example Reports

Codebook



Comparison report



Next Steps

- Production and maintenance mode
  - Enhance user manual
  - Apply tool to SED 2012
  - Apply minor fixes and enhancements
- Metadata driven environment
  - Currently data driven
  - Metadata must be considered earlier in process
  - Establish variable/classification/question banks
- Integrate in NCSES Data Repository
- Extend to other NCSES surveys
  - NSF SED Manager based on existing tool
  - Based on common framework, DDI with extensions
  - Open source

Thank you!

Contact information:

Kimberly Noonan
knoonan@nsf.gov

Pascal Heus
pascal.heus@metadatatechnology.com

Tim Mulcahy
Mulcahy-Tim@norc.org