



PROPOSED DATA SUBMISSION WORKFLOW FOR NSTX-U

S. M. Kaye

PPPL Data Management Plan approved

- Key component of DMP is making data contained in published figures/tables digitally accessible
 - Include metadata and file description information
- Will use Princeton Univ. DataSpace repository for committed data
- Will use temporary on-site area for review and revision until paper accepted and dataset finalized
- Papers will have to contain ARK (Archival Resource Key) of DataSpace COLLECTION in which the data will reside

DataSpace directory structure

- Communities
 - Collections (can be multiple)
 - Subcollections (can be multiple)
 - Items

- For us:

- PPPL
 - NSTX-U
 - Theory
 - Advanced Projects
 - Stellarator
 - Socio-economics
 - Engineering
 -



Suggested: to be finalized after further discussion within Depts.

NSTX-U COLLECTION will contain “ITEMs”

- Each ITEM consists of data from figures, metadata and file descriptions
 - One item for each publication
- NSTX-U (collection)
 - Item 1
 - Item 2
 - Item 3
- Propose to have only one COLLECTION level, containing all NSTX-U papers (although could break it up into Science and Engineering)
 - We will see why in a few slides

Each ITEM contains

- Metadata
 - Title
 - Authors
 - Keywords
 - Abstract

Preface material to files below

- **Readme.txt**
- **Data (Fig. 1)**
- **Data (Fig. 2)**
- **Data (Table 1)**
- **etc**

Active links to VIEW/DOWNLOAD

Data files can be XLS, CSV, other ASCII, HDF5, JPG, etc.

Can do a DataSpace search for ITEMS (from the COLLECTION level) based on

- Author
- Data (Year)
- Keywords (Macroscopic Stability, Transport and Turbulence, Energetic Particles, etc)
- Title
- etc

- Searches performed at level of NSTX-U “collection”
 - All ITEMS will be stored in this collection (no sublevels necessary)
 - **ARK (or URL) of NSTX-U Collection level will be provided in publication**
 - Allows for paper to be finalized before data is committed
 - *Data cannot be changed once it is committed*

Data will be “reviewed” prior to committing it to DataSpace

- Store in temporary local area to ensure:
 - Enough metadata information
 - Data in files is complete, in correct units, does not look crazy, etc. (no actual review of data and data provenance)
 - README file contains enough description of data.
- For README file, propose figure (table) captions plus whatever additional information author wishes to provide (explaining additional data contained in file that is not in figure, etc)
- Data files can be XLS, CSV, other ASCII, HDF5, JPG, etc.
 - IT will work on some web-based, IDL, MATLAB, PYTHON routines that can be coupled to analysis programs to allow creation of files

Proposed Workflow

