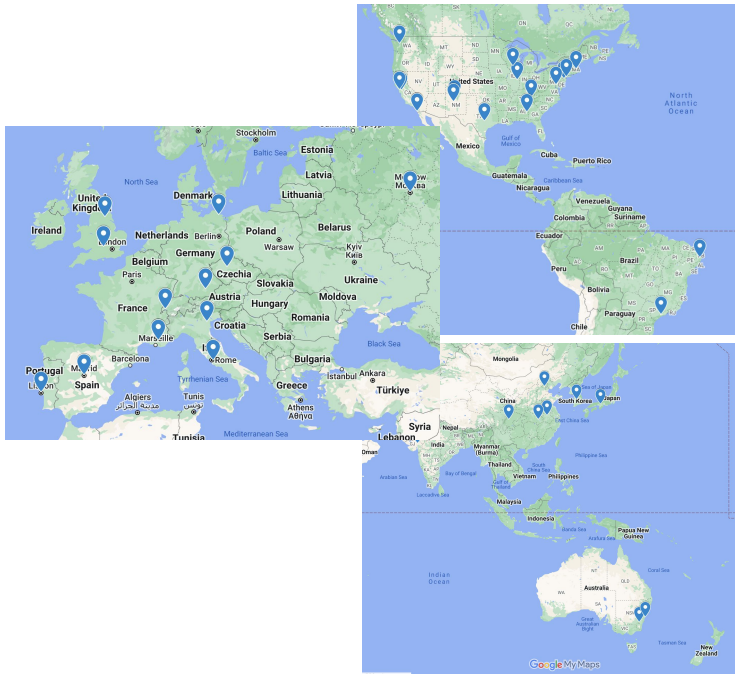
MDSplus version 8.0
The Path to Long Node Names

Mark Winkel, Joshua Stillerman, Stephen Lane-Walsh, Fernando Santoro

14th Technical Meeting on Control Systems, Data Acquisition, Data Management and Remote Participation in Fusion Research

Introduction

- **Introduction to MDSplus**
  - MDSplus is a widely used data management system in magnetic fusion energy research.
  - Provides data storage, management, and remote access through a vector-based interpreter API.
- **Version 8.0 allows long node names**
  - Also supports mixed case
  - Will be released later this year
- **This is a “breaking” change**
  - However, can co-exist with version 7.0



Requirements

Issues with v7.0 (and prior)	Desired	Features for v8.0
12-characters is too short	Long descriptive names (e.g., IMAS)	Expand to 63 usable bytes (maybe more)
Upper case is too restrictive	Mixed case names	Display = mixed case Internal = ignores case
ASCII is too limiting	Allow use of physics symbols. Used world-wide so support other character sets.	Add support for UTF-8 (Unicode)
Large archives in prior format	Must be backwards compatible	Yes, with option to convert
Many sites can't update all computers at once	Mixture of versions for clients and server	Yes, but with some limitations

Two File Formats

- Format v1
  - Data stored in a hierarchy of nodes. Path is a list of nodes delimited by "." and ":" characters.
    - `\CMOD::TOP.MHD.MAGNETICS.MAG_PROCESSED.CURRENT_DATA:IP`
  - Tag has a single name, evaluates to a node in the hierarchy
    - `\IP`
  - Node names limited to 12 uppercase ASCII characters; Tag names limited to 23 ASCII characters
  - Null delimited C string (with one exception)
    - `SHORT`<null> = 6 bytes
    - `LONGEST_NODE` = 12 bytes, note no <null>
  - Used by MDSplus 7.x and prior versions

- Format v2
  - Field size is 64 bytes
  - Always <null> terminated
  - Longest node name is 63 bytes; Tag name is also 63 bytes
    - `ASCII_IS_A_SUBSET_OF_UTF_8_THUS_THE_LONGEST_NAMES_ARE_THIS_LONG`
    - `MDSplus_v8_can_use_UPPERCASE_lowercase_MixedCase`
    - `UTF_8_symbols_βπ→α_and_languages_`山ゟ木شñüçî
  - MDSplus 8.0 can read / write / edit both formats

UTF-8 Unicode – Plasma Current

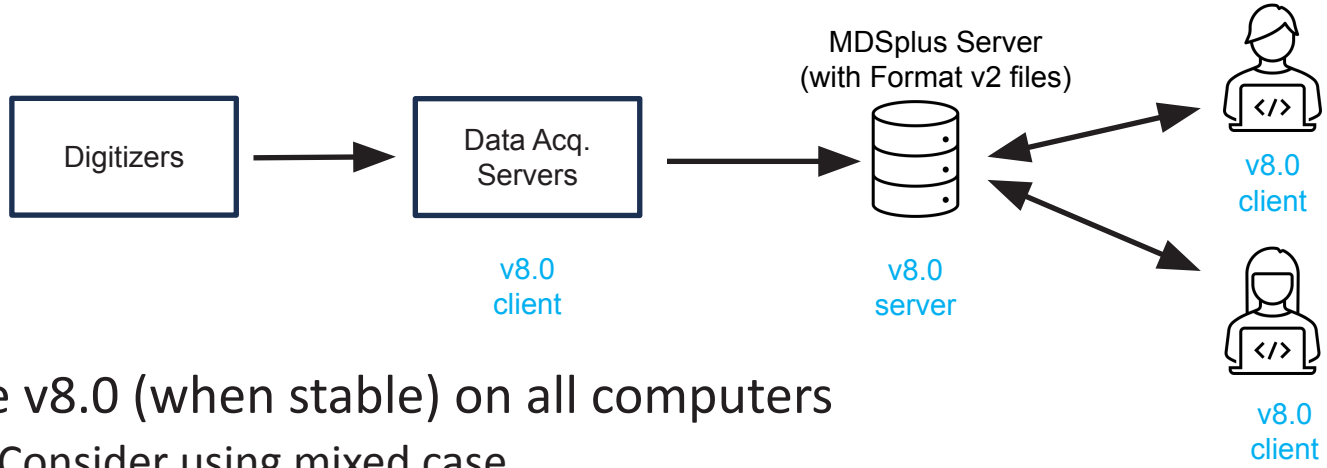
Unicode UTF-8	Language
\corriente_de_plasma	Spanish
\corrente_de_plasma	Portuguese (Brazilian)
\courant_plasma	French
\Plasmastrom	German
\ρεύμα_πλάσματος	Greek
\プラズマ電流	Japanese
\플라즈마 전류	Korean
\等离子电流	Chinese
\اتدفق البلازما	Arabic

- Names must conform to MDSplus syntax
  - no spaces, no dashes, must start with leading letter, etc.
- Collaboration easier if use ASCII for nodes
  - FAIR = Findable, Accessible, Interoperable, Reusable
  - Use English names
  - Perhaps use IMAS
- Recommend UTF-8 only for tags
- Use UTF-8 judiciously
  - Physics symbols
  - Right-to-left languages display poorly

Design Decisions / Issues

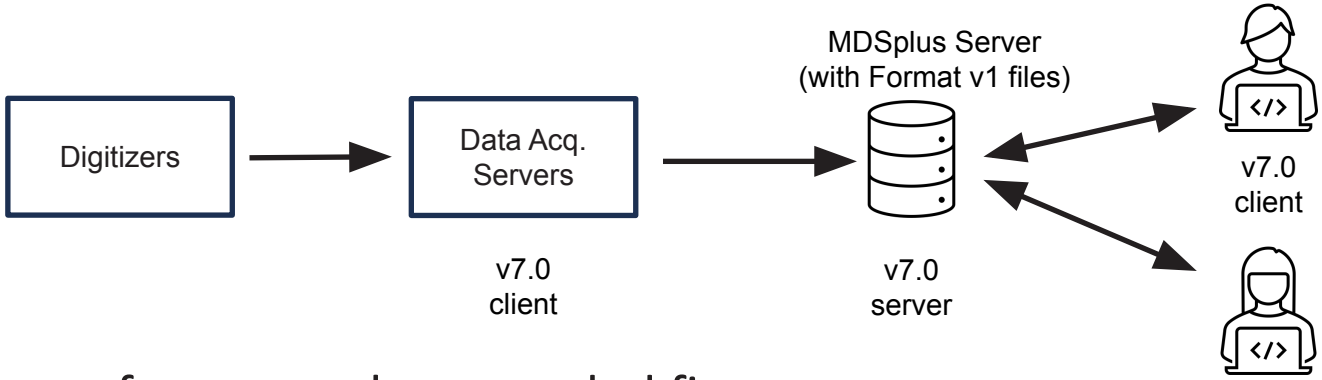
- Concept: v8.0 reads/writes both formats, v7.0 only handles Format v1
- Mixed case is for display only
  - Internally, case is ignored – v8.0 does “case folding” when comparing / evaluating node names
- v7.0 must generate an error message when given a Format v2 file
- v8.0 must read / write / edit both formats
  - Does in-memory expansion when reads a Format v1 file
  - When editing a Format v1 file, must ensure no duplicate node names (must error)
    - THIS_NODE_OK
    - THIS_node_OK_BUT_NOT_REALLY
  - Should also have the ability to convert from Format v1 to Format v2
- UTF-8 multibyte characters pose some challenges
  - Not all characters have an “uppercase” equivalent
  - Affects the parser for the TDI expressions

Deploying – New Installation



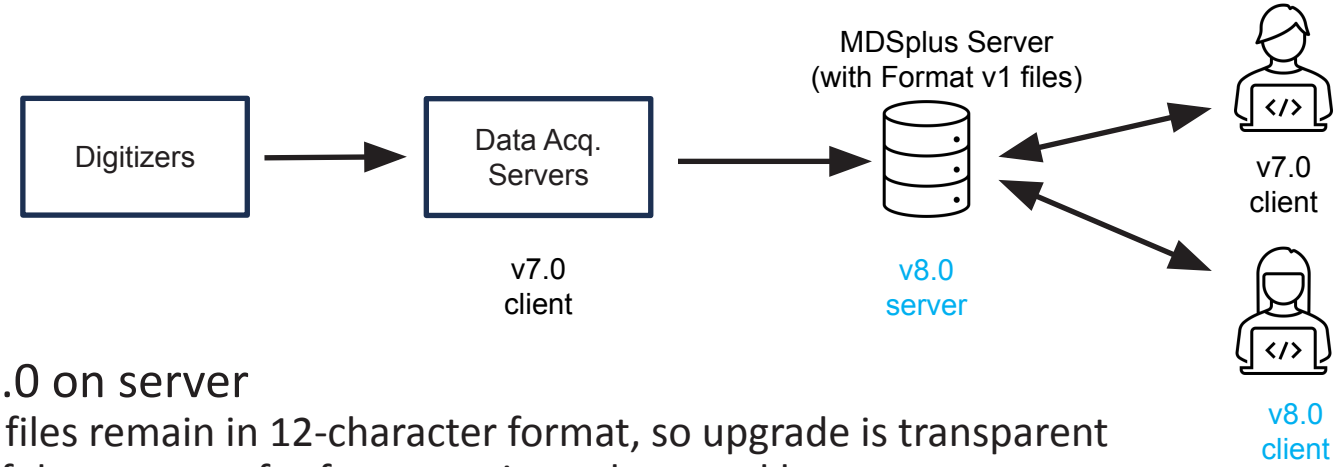
- Use v8.0 (when stable) on all computers
  - Consider using mixed case
  - Good for new projects
- Notes
  - v8.0 is presently in development
  - Try it on a test system before deploying in production

Deploying – Upgrading some Clients



- Client software can be upgraded first
  - Files remain in 12-character format, so upgrade is transparent
  - Some users might download and install v8.0 client
- Notes
  - Establish site policies regarding upgrade of client software

Deploying – Upgrading existing Server



- Use v8.0 on server
  - Old files remain in 12-character format, so upgrade is transparent
  - Useful to prepare for future projects that need long names
- Notes
  - Don’t use both file formats on same project (e.g., old diagnostic)
  - Don’t have Format v1 TDI expressions reference Format v2 nodes
  - If a new diagnostic uses a model in Format v2, then users of that diagnostic must have v8 clients.

Building

- MDSplus.org will eventually switch to the CMake build tools
  - Parallel builds
  - Parallel testing
  - Easier to maintain
- v8.0 thus uses the CMake tools

Current Status

- Working prototype exists
  - Has been demonstrated with some customer projects
  - Passes 90% of the current automated test suite
    - Some tests just need updated “golden” output files
- Still in development
  - Needs more testing to confirm no string overflow
  - Tools (jTraverser, jScope, etc.) need to be updated for long node names
  - Must test with “action servers” and “device drivers”
  - UTF-8 changes still to be done, will require extensive testing

Availability

- Current prototype is available now on GitHub
  - https://github.com/MDSplus/mdsplus
  - Use the “mdsplus8” branch
- We welcome feedback on the prototype
- Plan is for the production version to be released later this year

Questions?

- What are the most valuable features of v8.0 for your site?
- Is UTF-8 important?
- Would it be difficult to deploy v8.0 (when stable) at your site?
- Major change so could also add other features
  - New commands: ls, cd, pwd, mv, et cetera?
  - Command completion?
  - Tab completion of long node names?
- Other?

Coming Soon

- MDSplus for Apple Silicon MacOS

Thank You!

This work was funded under DOE cooperative agreement DE-SC0012470.