





Experiences Integrating HDF5 into DREAM.3D

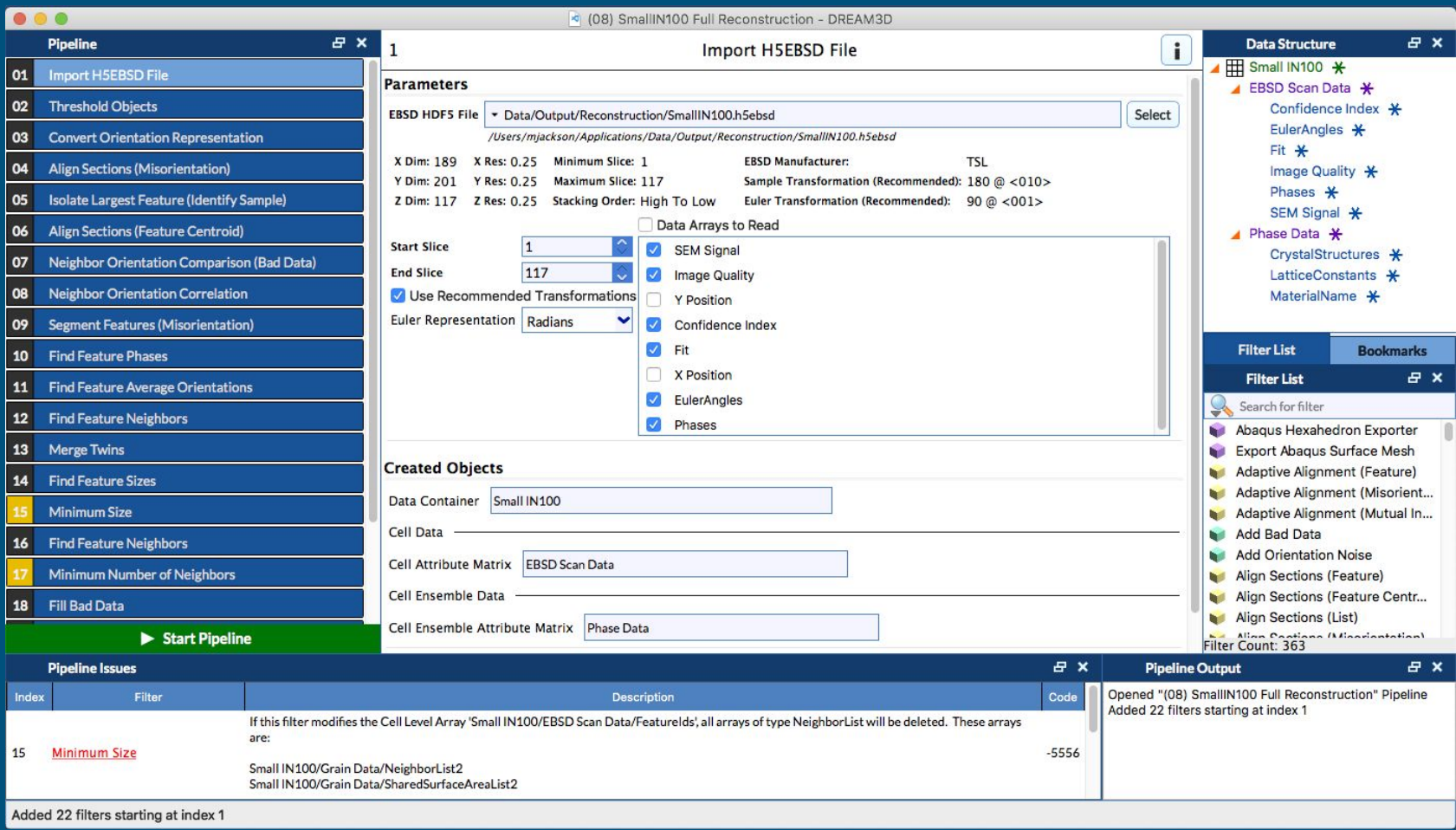


Mr. Michael A. Jackson
Owner/Developer
@ BlueQuartz Software



DREAM.3D

- Air Force Research Laboratory (AFRL) sponsored project
 - Continuously developed since 2009
 - Unified cross platform framework to allow disparate codes to work together
- Materials Science and Engineering (MSE) Data Analysis
 - Qt5 based GUI application
 - Extensible Framework
- Open Source hosted on GitHub
- Foster collaboration between MSE and engineering/design groups
 - Repeatable workflows
- Portable Data
 - Self describing, Fast I/O, available from desktop to HPC
- <http://dream3d.bluequartz.net>
- <http://www.github.com/bluequartzsoftware/dream3d>
- <https://my.cdash.org/index.php?project=DREAM3D>



DREAM.3D User Interface (6.5 version)

Selecting Open Binary File Format

- Spent a fair bit of time evaluating lots of different possibilities
- 2006 Time Frame
 - Images + sidecar file
 - Binary XML
 - Plain Binary Files + sidecar file
 - A few others which I can't remember
- HDF5 checked most of the boxes that we needed
 - Not a lot (at the time) of built up infrastructure for HDF5
 - Fast I/O, Self describing, Flexible
 - All the reasons that you might already be using HDF5
- In the end, HDF5 saves developer frustration, developer time, creates more consistent files that can be exchanged among research groups.

HDF5 & DREAM.3D History

- Started with HDF5 1.6 (2008/2009)
 - Different build systems on each platform, difficult to get correct on Windows/MSVC
 - <https://github.com/BlueQuartzSoftware/H5Support>
- Moved to HDF5 1.8 (Late 2009)
 - Collaborated with Dr. John Biddiscombe to add CMake support to HDF5 1.8
 - Handed that code off to THG where it has been embraced and maintained since
 - Enables easier integration and use of HDF5 in DREAM3D, and other CMake based projects, from a configuration/compilation point of view
 - Changes to target naming inside of CMake throughout 1.8 series was problematic
- Moved to HDF5 1.10 (Late 2018)
 - Minor code updates in DREAM.3D
 - Generally smooth sailing

H5EBSD: Moving an Industry to HDF5

- EBSD: Electron BackScatter Diffraction
 - Use an electron beam to reveal internal structure of materials
- EBSD equipment vendors all have separate and incompatible file formats
 - Only readable format for external applications is ASCII text
 - Importing ASCII data is slow, prone to failures and has precision (float) issues
 - Advanced users need access to the raw data, not the processed data.
 - Raw data hidden away in proprietary file formats
 - Three main OEMs: EDAX, Oxford Instruments, Bruker
- DREAM.3D already generated HDF5 archives where those ASCII files were converted to a single HDF5 file (.h5ebbsd)
 - Faster loads
 - More descriptive
 - Multiple ASCII files converted to single HDF5 file
- <https://link.springer.com/article/10.1186/2193-9772-3-4>

From Text to HDF5: EBSD Example

```

1 | TEM_PIXperUM      1.000000
2 | # x-star          0.372300
3 | # y-star          0.689300
4 | # z-star          0.978100
5 | # WorkingDistance 5.000000
6 |
7 | # Phase 1
8 | # MaterialName    1.000000
9 | # Formula         3 # x-star 0.372300
10| # Info            4 # y-star 0.689300
11| # Symmetry        5 # z-star 0.978100
12| # LatticeConsta  5.000000
13| # NumberFam      6 # WorkingDistance
14| # hklFamili      7 # Phase 1
15| # hklFamili      8 # MaterialName Nickel
16| # hklFamili      9 # Formula Ni
17| # hklFamili     10 # Info
18| # Symmetry       11 # TEM_PIXperUM 1.000000
19| # LatticeConsta 12 # x-star 0.372300
20| # NumberFam     13 # y-star 0.689300
21| # GRID: SqrGrid 14 # hklFamilies 4 # z-star 0.978100
22| # XSTEP: 0.15  15 # hklFamilies 5 # WorkingDistance 5.000000
23| # NCOLS_ODD   16 # hklFamilies 6 # Phase 1
24| # NCOLS_EVEN 17 # hklFamilies 7 # MaterialName Nickel
25| # NROWS: 201 18 # Categories 0 # Formula Ni
26| # GRID: SqrGrid 19 # Info
27| # OPERATOR:    20 # Symmetry 1 # TEM_PIXperUM 1.000000
28| # XSTEP: 0.2500 21 # XSTEP: 0.2500 11 # LatticeConstan 2 # x-star 0.372300
29| # YSTEP: 0.2598 22 # YSTEP: 0.2598 12 # LatticeConstan 2 # y-star 0.689300
30| # SAMPLEID:    23 # NCOLS_ODD: 18 13 # NumberFamilies 4 # z-star 0.978100
31| # SCANID:      24 # NCOLS_EVEN: 14 14 # hklFamilies 5 # WorkingDistance 5.000000
32| # OPERATOR:    25 # NROWS: 201 15 # hklFamilies 6 # Phase 1
33| # OPERATOR:    26 # hklFamilies 7 # MaterialName Nickel
34| # OPERATOR:    27 # Categories 0 0 # Formula Ni
35| # OPERATOR:    28 # SAMPLEID: 19 # GRID: SqrGrid 10 # Info
36| # OPERATOR:    29 # XSTEP: 0.2500 11 # Symmetry 43
37| # OPERATOR:    30 # YSTEP: 0.2500 12 # LatticeConstants 3.520 3.520 3.520 90.000 90.000 90.000
38| # OPERATOR:    31 # XSTEP: 0.2500 13 # NumberFamilies 4
39| # OPERATOR:    32 # NCOLS_ODD: 189 14 # hklFamilies 1 1 1 0.000000
40| # OPERATOR:    33 # NCOLS_EVEN: 18 15 # hklFamilies 2 0 0 1 0.000000
41| # OPERATOR:    34 # NROWS: 201 16 # hklFamilies 2 2 0 1 0.000000
42| # OPERATOR:    35 # hklFamilies 3 1 1 1 0.000000
43| # OPERATOR:    36 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
44| # OPERATOR:    37 # Categories 0 0 0 0
45| # OPERATOR:    38 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
46| # OPERATOR:    39 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
47| # OPERATOR:    40 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
48| # OPERATOR:    41 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
49| # OPERATOR:    42 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
50| # OPERATOR:    43 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
51| # OPERATOR:    44 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
52| # OPERATOR:    45 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
53| # OPERATOR:    46 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
54| # OPERATOR:    47 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
55| # OPERATOR:    48 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
56| # OPERATOR:    49 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
57| # OPERATOR:    50 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
58| # OPERATOR:    51 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
59| # OPERATOR:    52 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
60| # OPERATOR:    53 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
61| # OPERATOR:    54 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
62| # OPERATOR:    55 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
63| # OPERATOR:    56 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
64| # OPERATOR:    57 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
65| # OPERATOR:    58 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
66| # OPERATOR:    59 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
67| # OPERATOR:    60 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
68| # OPERATOR:    61 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
69| # OPERATOR:    62 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
70| # OPERATOR:    63 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
71| # OPERATOR:    64 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
72| # OPERATOR:    65 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
73| # OPERATOR:    66 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
74| # OPERATOR:    67 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
75| # OPERATOR:    68 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
76| # OPERATOR:    69 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
77| # OPERATOR:    70 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
78| # OPERATOR:    71 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
79| # OPERATOR:    72 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
80| # OPERATOR:    73 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
81| # OPERATOR:    74 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
82| # OPERATOR:    75 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
83| # OPERATOR:    76 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
84| # OPERATOR:    77 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
85| # OPERATOR:    78 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
86| # OPERATOR:    79 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
87| # OPERATOR:    80 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
88| # OPERATOR:    81 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
89| # OPERATOR:    82 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
90| # OPERATOR:    83 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
91| # OPERATOR:    84 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
92| # OPERATOR:    85 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
93| # OPERATOR:    86 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
94| # OPERATOR:    87 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
95| # OPERATOR:    88 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
96| # OPERATOR:    89 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
97| # OPERATOR:    90 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
98| # OPERATOR:    91 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
99| # OPERATOR:    92 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
100| # OPERATOR:   100 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
101| # OPERATOR:   101 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
102| # OPERATOR:   102 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
103| # OPERATOR:   103 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
104| # OPERATOR:   104 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
105| # OPERATOR:   105 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
106| # OPERATOR:   106 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
107| # OPERATOR:   107 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
108| # OPERATOR:   108 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000
109| # OPERATOR:   109 # OPERATOR: Ad 17 # hklFamilies 3 1 1 1 0.000000

```



SmallIN100.h5ebstd

- Data
 - Confidence Index
 - Fit
 - Image Quality
 - PhaseData
 - Phi
 - Phi1
 - Phi2
 - SEM Signal
 - X Position
 - Y Position
- Header
 - 10
 - Data
 - Header
 - 100
 - 101
 - 102
 - 103
 - 104
 - 105
 - 106
 - 107
 - 108
 - 109

General Attributes

Name	HDF5 Dataset
Path	/1/Header/OriginalHeader
Type	
Object ID	14593
No. of Dimension(s)	1
Dimensions Size(s)	677
Data Type	H5T_STRING

H5EBSD: Moving an Industry to HDF5

- Timeline
 - 2013 First conversations with EDAX, example code sent to EDAX
 - 2014 EDAX Updated release with support for writing HDF5 files
 - 2016 Bruker has tool to convert from .bcf (proprietary format) to HDF5
 - 2019 Oxford Instruments starts to use HDF5 as a supported file format

- Positive feedback from EBSD users regarding the changes
 - Easier to I/O data to/from their custom data analysis programs or control systems

Current Trends in Open Source

- Recent trends with some open source companies is to put the source/binaries behind a “wall”.
 - Paid wall, subscription wall, free account wall
- These walls stop automated scripts in their tracks
 - DREAM3D CI broke, DREAM3D SDK Build Scripts broke.
 - BQ mirrored the HDF5 sources/binaries for DREAM.3D’s use. All that tracking information was lost to The HDF Group
 - DREAM3D seriously considered moving away from HDF5
- We need better ways of financially supporting companies that produce open source software

HDF5: Open Discussions

- **The HDF Group moving to more open development**
 - Enlightening discussion on the HDF Forum.
 - Helped to understand the issues that The HDF Group were/are facing
- **Repositories hosted on GitHub**
 - <https://github.com/HDFGroup/hdf5>
 - Submit bugs/feature requests/code through the “Issues” area
 - Standard PR (Pull Release) workflow
- **Binaries *easily* downloaded from www.hdfgroup.com**
 - <https://www.hdfgroup.org/downloads/hdf5/>
- **Downloads help HDF Group make a case for continued support from funding agencies**

Contributing & Helping HDF5 Ecosystem

- If you are using HDF5 consider the value proposition that HDF5 has brought to your project.
- Consider getting paid support/consulting from The HDF Group
 - However small it might be
- If you have projects/proposals where data storage comes into play, talk to The HDF Group about being a partner/sub on your contract
 - This takes longer term planning and discussions with the groups for whom you work
 - Start those discussions now.
- Most people think that giving back is fixing a bug in code
- Any aspect of development can be a target for your efforts
 - Code, Tests, Documentation, build bots
 - All of these things can help

Takeaways

- HDF5 saves developer time
- Large passionate community of HDF5 users/developers
- Fosters collaboration between research groups
- Be advocates of HDF5 and introduce it into new industries
- HDF5 is not **free** to develop, although it is **free** to use
- You should consider using HDF5 Group on your next project
- The HDF Group are some of the best developers that I have had the pleasure of interacting with.

Setting the context...

- BlueQuartz Software
 - 3 Full Time, 3 part time
 - Primarily funding streams are from DoD sources
- **All** of our software has been open source
- Budgets are tight and funding is even tighter
- Finding funding is sometimes difficult for your own company, but consider the value proposition of HDF5 and your own project(s)
- Strong advocate for HDF5 since 2005'ish