

Technical report with portable systems microscopy format (PSM) specification

Achievement A8.3.1 Systems Microscopy Network
Sampled dataset attached – “Sampled WiSoft DataSet”

The following report describes the Portable Systems Microscopy (PSM) format that is currently the standard format of the WiSoft analysis software by IDEA Bio-Medical.

The data for a cell-based experiment consists of the following elements:

- Raw image data. This data is stored in OME TIFF format, with additional information stored in xml format, for maximum compatibility with other software environments.
- A summary of the experimental run. This data is stored in xml format.
- Dataset. The dataset consists of all elements resulting from an analysis of the raw image data according to a defined protocol. It consists of one or more masks for each field of view, which define regions of each field corresponding to objects of biological significance, an equal number of binary files consisting of the object attributes (measured properties) of each object, an xml file containing some descriptive information including the name of each attribute. It also contains an optional file holding the definition of outliers, if any. There may be any number of datasets, as the same raw data can be analyzed by multiple methods.

The data is stored in a hierarchical format of files and folders as described below in top-down form.

1. Plate

Data for each plate is stored in a single folder. Each PLATE's folder is named for the Plate ID, or using a unique name based on the experiment type plus a date-and-time string. An example Plate folder name is “Protein_Expression_04_11_2012_03_00”. An example of this hierarchy is shown in Appendix I.

1.1. RunSummary.xml

This file contains a description of the acquisition parameters and conditions. The first level of the XML schema is shown below:

2 Attributes:

Name	Value
xmlns:xsi	http://www.w3.org/2001/XMLSchema-instance
xmlns:xsd	http://www.w3.org/2001/XMLSchema

29 Subtags:

Tag name/Text	Text	Unique Subtags
User_Name	New	
Date	2012-11-04T03:00:55.976258+02:00	
Application	Protein Expression	
Plate	Whatman 96-well	
Scan_Region	F2 - G11,	
Well_Fill	100	
Well_Coverage	25	
Fields_per_Well	11	
Time_Lapse		2 unique subtags
Time_Lapse_Summary	1 cycles, every 0 sec.	
Number_Of_Z_Planes	1	
Inter-Plane_Distance	1	
Magnification	10	
Tags_File_Name	Tags_Count_ProtExp_Morph.csv	
Imaging_Info		Biological_Assignment (3 occurrences)
Magnification2	-1	
ObjName1	10_0.4	
ObjName2		
Analysis_Info		AnalysisParams
User		8 unique subtags
Run_Statistics		17 unique subtags
Dataset_Name		
Dataset_Full_Name	DS_Protein_Expression_04_11_2012_03_00	
Dataset_Comments		
NAfPoints	2147483647	
HermesVer	1.0.2.0	
LiveCellControl	false	
Temp	37	
CyclesStartTimes		item

1.2. Images

This folder contains a subfolder for each well for which image data was acquired.

1.2.1. <Well_ID>, (with Well_ID in format "A1")

This folder will contain 2 x W x T x Z files, where W is the number of biological assignments (fluorescence channels), T is the number of timepoints, and Z is the number of stack slices. For each biological assignment, timepoint, and stack slice, there will be one OME TIFF image and one corresponding xml header file.

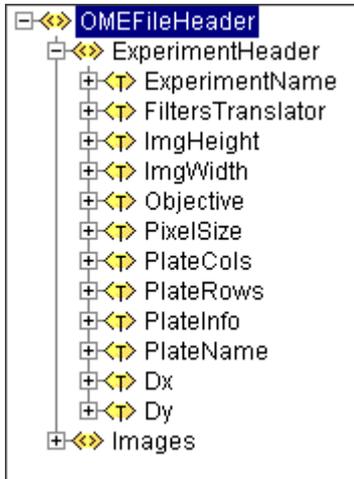
1.2.1.1. <Well_ID>_IX_<Biological_Assignment>_T<timepoint>_Z<stackslice>.ome.tif

This is a multipage TIFF file in OME TIFF format, with one page for each location imaged in the well. <Biological_Assignment> corresponds to one of the string values stored in the "Imaging_Info" sub-table in the RunSummary.xml file, and will usually have a value like "Cell", "Nucleus", "Protein", etc.

<timepoint> is a zero-based integer counter of the number of time points in the experiment, and <stackslice> is a zero-based counter of the number of Z stack slices in the experiment.

1.2.1.2. <Well_ID>_IX_<Biological_Assignment>_T<timepoint>_Z<stackslice>.xml

This is an xml file containing descriptive information corresponding to the experiment, and descriptive information for each location imaged in the well. The schema appears as follows:



Typical values of the Experiment Header section look like this:

Tag name/Text	T	Text
ExperimentName		Protein Expression
FiltersTranslator		4:Nucleus [Blue],2:Protein [Green],0:Cell [FarRed]
ImgHeight		1019
ImgWidth		1370
Objective		10
PixelSize		0.583405017852783
PlateCols		12
PlateRows		8
PlateInfo		12,8,c,14,38,11,24,9,9,6,09
PlateName		Whatman 96-well
Dx		799.2649
Dy		594.4897

Typical values of the Images section look like this:

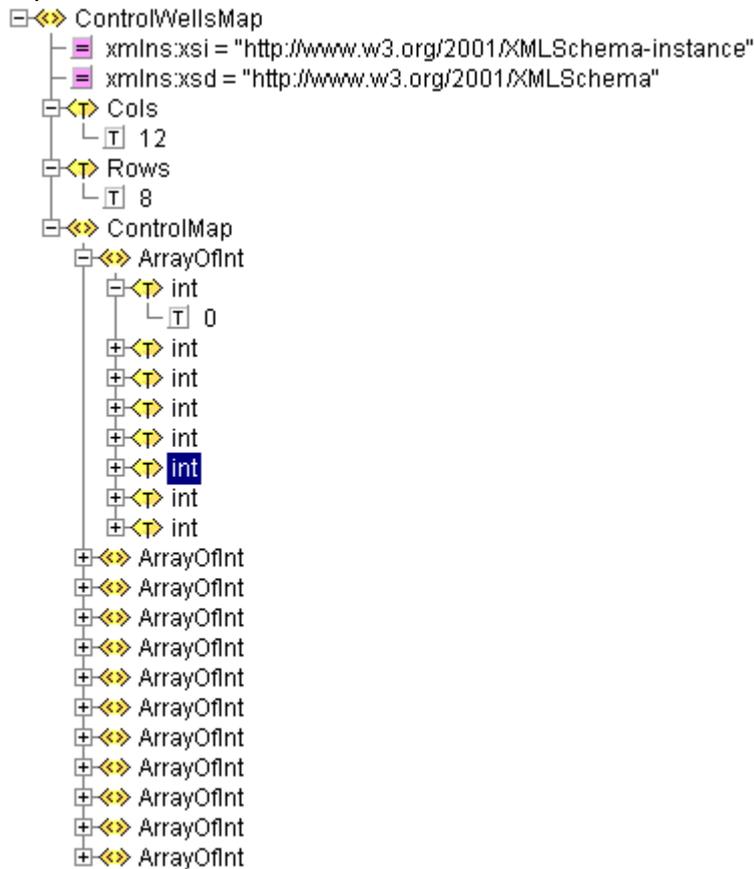
Tag name/Text	Field	Height	Width	TimeIdx	Wavelength	WaveName	WellCol	WellRow	Well	XPos	YPos	ZPlaneIdx	ZPlane	ImageNumInFile	Index
ImageHeader 0	1019	1370	0	2		Protein [Green]	2	6	F2	22580.7342529297	55645.5116271973	0	91	0	22
ImageHeader 1	1019	1370	0	2		Protein [Green]	2	6	F2	22580.7342529297	56239.9978637695	0	90.599609375	1	23
ImageHeader 2	1019	1370	0	2		Protein [Green]	2	6	F2	22580.7342529297	56834.4879150391	0	90.2998046875	2	24
ImageHeader 3	1019	1370	0	2		Protein [Green]	2	6	F2	23380.0010681152	57428.9779663086	0	89.7998046875	3	25
ImageHeader 4	1019	1370	0	2		Protein [Green]	2	6	F2	23380.0010681152	56834.4879150391	0	91	4	26
ImageHeader 5	1019	1370	0	2		Protein [Green]	2	6	F2	23380.0010681152	56239.9978637695	0	90.7001953125	5	27
ImageHeader 6	1019	1370	0	2		Protein [Green]	2	6	F2	23380.0010681152	55645.5116271973	0	90.599609375	6	28
ImageHeader 7	1019	1370	0	2		Protein [Green]	2	6	F2	23380.0010681152	55051.0215759277	0	91.2001953125	7	29
ImageHeader 8	1019	1370	0	2		Protein [Green]	2	6	F2	24179.2640686035	55645.5116271973	0	90.900390625	8	30
ImageHeader 9	1019	1370	0	2		Protein [Green]	2	6	F2	24179.2640686035	56239.9978637695	0	91.900390625	9	31
ImageHeader 10	1019	1370	0	2		Protein [Green]	2	6	F2	24179.2640686035	56834.4879150391	0	91.099609375	10	32

1.2.1.3. WellMontage.jpg

This is 8 bits per pixel thumbnail (max 2048x2048 pixels) in JPEG format of one channel, showing all of the imaged locations in their positions with respect to each other.

1.2.2. controls.xml

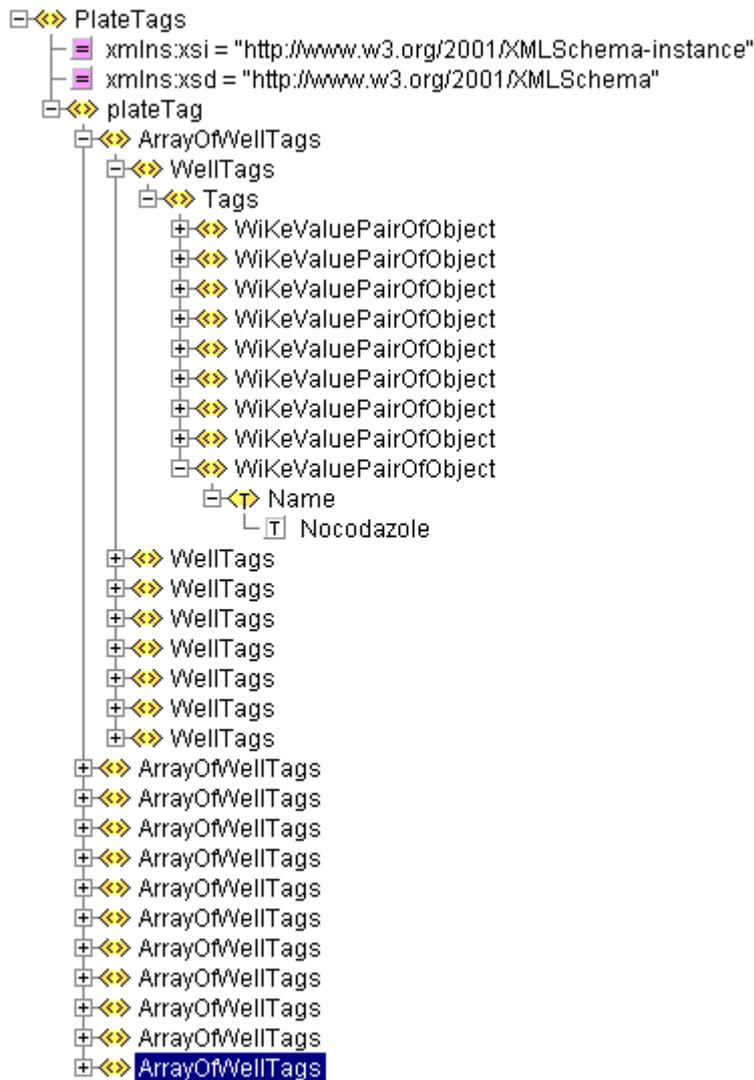
This is an xml file containing the location of positive and negative controls on the plate. The schema is depicted below:



Default values are zero. A value of 1 indicates a positive control. A value of -1 indicates a negative control.

1.2.3. tags.xml

This is an xml file which contains text tags to identify any test conditions or compounds applied to each well. The schema is detailed below:



The number of rows and columns in the schema must correspond to the plate dimensions. All wells which have the same tag will have the result values aggregated statistically for reporting and/or dose/response curve fitting.

1.3. <Dataset>

This is a folder which contains the results of an analysis according to a particular protocol. There are no restrictions on the name of this folder, and there may be any number of such folders, each corresponding to a separate analysis of the raw data in the same parent folder. There are no subfolders. The folder contains 3N+2 files - three files for each well, plus two additional files for the whole experiment. The contents of each file are as follows:

1.3.1. <DataSet>.D_<R>_<C>

This file contains the quantitative information for each object in a well.

<DataSet> is the name of the Dataset folder. <R> and <C> are the row and column indices (one-based) of each well.

1.3.2. <DataSet>.xml.masks_<R>_<C>

This file contains the mask information for each object in a well.

<DataSet> is the name of the Dataset folder. <R> and <C> are the row and column indices (one-based) of each well.

1.3.3. <DataSet>.xml.masks_<R>_<C>_<Biological_Assignment>

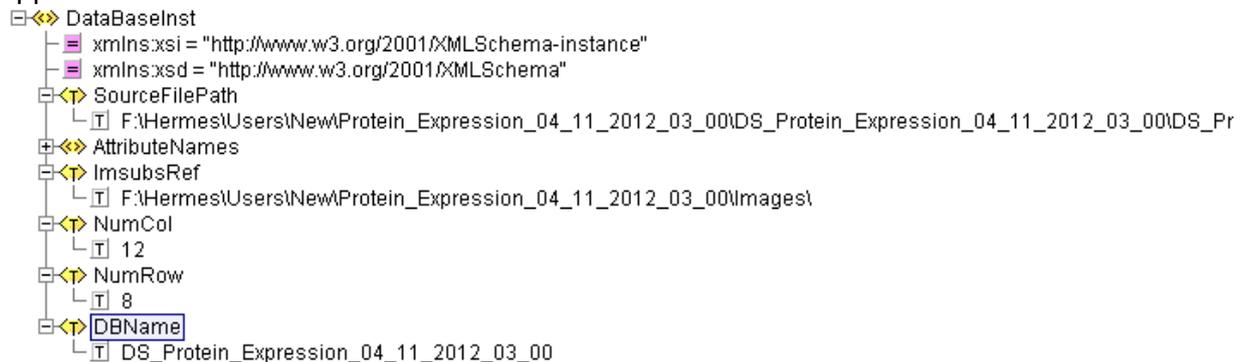
<DataSet> is the name of the Dataset folder. <R> and <C> are the row and column indices (one-based) of each well. <Biological_Assignment> is the string name assigned to the fluorescence channel, as stored in RunSummary.xml.

1.3.4. <DataSet>.outlier

This file contains a binary representation of the status of each object according to a rule-based or selection-based definition of outliers. Outliers are excluded from statistical aggregation into the well-based results.

1.3.5. <DataSet>.xml

An xml file containing the names of all measured attributes of the objects in each image. The schema appears as follows:



The AttributeNames key appears as follows (example of a set of quantify attributes):



Tag name/Text	T	Text
<T> string		ID
<T> string		Flag
<T> string		CmX
<T> string		CmY
<T> string		Area
<T> string		TotInt
<T> string		AvgInt
<T> string		TotInt-Bkgnd
<T> string		AvgBck
<T> string		AvgInt-Bkgnd
<T> string		Perim
<T> string		FramePerim
<T> string		AreaCH
<T> string		PerimCH
<T> string		Variance
<T> string		MinX
<T> string		MaxX
<T> string		MinY
<T> string		MaxY
<T> string		LongAxis
<T> string		ShortAxis
<T> string		AxialRatio
<T> string		LongAxisAng
<T> string		Solidity
<T> string		Cell2HullAreaRatio

Appendix I. Example of plate folder hierarchy

Protein_Expression_04_11_2012_03_00

DS_Protein_Expression_04_11_2012_03_00_script.txt

RunSummary.xml

Images

F2

F2_IX_Cell [FarRed]_T0_Z0.ome.tif

F2_IX_Cell [FarRed]_T0_Z0.xml

F2_IX_Nucleus [Blue]_T0_Z0.ome.tif

F2_IX_Nucleus [Blue]_T0_Z0.xml

F2_IX_Protein [Green]_T0_Z0.ome.tif

F2_IX_Protein [Green]_T0_Z0.xml

wellMontage.jpg [optional]

. [next well]

. [next well]

. [last well]

controls.xml

tags.xml

DS_Protein_Expression_04_11_2012_03_00

DS_Protein_Expression_04_11_2012_03_00.D_6_2

DS_Protein_Expression_04_11_2012_03_00.xml.masks_6_2_Nucleus [Blue]

DS_Protein_Expression_04_11_2012_03_00.xml.masks_6_2

. [same three files for each well, with appropriate _R_C designator]

DS_Protein_Expression_04_11_2012_03_00.xml

DS_Protein_Expression_04_11_2012_03_00.outlier