I-SPY 1 Data Sharing Dictionary – TCIA Data Collection

MODIFICATION/REVIEW ACTIONS	
Date	By
9/22/2014	Created by: David Newitt
12/12/2014	Reviewed by Nola Hylton
2/8/2015	Modified by David Newitt

Data Dictionary and Object Descriptions

Table of Contents

Scope

This document contains definitions for the DICOM objects and attributes specific to the ACRIN 6657/I-SPY 1 MRI image data collection, and descriptions of associated data provide with this image data set.

Patient and study identification

All patients and studies are identified through standardized, deidentified attributes as shown in Table 1.

- Patient unique identification is provided through a 4 digit identifier set by the I-SPY TRIAL
- Study time-point is identified through the Clinical Trial Visit Codes: T0, T1, T2, and T3
- Imaging center discrimination has been retained (Clinical Trial Site ID) to allow for center-based analysis, but has been anonymized as codes “SITE A”, “SITE B”, etc…
- In general all private DICOM attributes considered HIPPA compliant by TCIA have been retained in the image sharing submission process; but it should be noted that some studies may have been anonymized prior to submission to the analysis and archiving centers, so the presence of scanner manufacturer private attributes that may be required for some analyses can not be guaranteed.

All objects have been deidentified to preserve patient privacy. If any evidence of non-HIPPA compliant patient PHI is found please notify the UCSF Breast Imaging Research Program core lab, c.o. david.newitt@ucsf.edu .

Table 1: Patient and study identification attributes included in all DICOM objects

Variable Name	Variable Description	Format	DICOM Tag
Patient Name	Encoded patient ID (pppp), ACRIN Protocol ID, study name	pppp^6657^ISPY1	(0010,0010)
Patient ID	Study name and patient ID	ISPY1_pppp	(0010,0020)
Clinical Trial Patient ID	Unique identification for patient in the trial	Number: pppp	(0012,0040)
Clinical Trial Site ID	Code of the trial site: "SITE x"	text	(0012,0030)
Clinical Trial Visit Code	ID code for the trial visit T0 baseline T1-n follow-ups	text	(0012,0050)
Clinical Trial Visit Description	Description for the trial visit: baseline early-treatment inter-regimen pre-surgery	text	(0012,0051)

DICOM Objects

Table 2 details the DICOM objects provided by the UCSF core lab in the I-SPY 1 / ACRIN 6698 shared data set on TCIA. Not all objects will be present in all cases due to unanalyzable studies. Original image objects are unprocessed except for necessary conversion to DICOM and deidentification. Derived image objects, including standardized early time-point percent enhancement (PE) and signal enhancement ratio (SER) maps, are provided for all volume-SER analyzable studies. DICOM segmentation objects representing PE analyzable breast tissue and SER volume are also provided. Parameters used in calculating the derived objects are stored in DICOM private group attribute set (0117,10xx) in each object as described below.

Table 2: DICOM Objects in shared data set

Group	Object Group Name	Description	Format	SERIES ID *
Images and Derived Maps				
Original DCE MRI Volumetric Image Sets	MRI pre contrast	pre-contrast image set	DICOM Image Files	S0 (original)
	MRI early post contrast	2'30" (nominally) post-injection image set	DICOM Image Files	S1 (original)
	MRI late post contrast	7'30" (nominally) post-injection image set	DICOM Image Files	S2 (original)
Derived DCE Image Maps	PE early	Percent signal enhancement map at the early (nominally 2'30" post-injection) time-point relative to the pre-injection baseline image: PE=100.0 * MRI(early) / MRI(pre)	DICOM Image Files	(Sref*10000) + 1001
	SER **	Signal enhancement ratio (SER) map between the early (nominally 2'30" post-injection) and late (nominally 7'30" post-injection) time-points: SER=PE(early) / PE(late)	DICOM Image Files	(Sref*10000) + 1000
DICOM Segmentations				Series ID
Fibroglandular tissue	PE_SEG	Segmentation used for early post-contrast PE map	DICOM Segmentation objects	(Sref*10000) + 2001
PE thresholded SER mask	SER_SEG	Segmentation used for SER map	DICOM Image Files	(Sref*10000) + 2000

* Sref = reference series number = S0 if S0 < 100 (GE, Siemens), or S0/100 if S0 >= 100 (Phillips)

** See Appendix A for SER derivation

Private attributes

Overview

The following information has been added to some or all of the DICOM objects in the data set:

- Image quality and protocol compliance assessment
- Timing information
- Tumor volume of interest (VOI)
- SER analysis parameters
- Functional tumor volume (FTV) results

All are contained in DICOM group 0117x, labeled with a private creator field:

(0117,0010) UCSF BIRP PRIVATE CREATOR 011710xx

Image quality and protocol compliance assessment

Image quality and protocol compliance were assessed by the UCSF core lab for all submitted image studies. DCE images were assessed for fat suppression, image quality, and artifacts; and then given an overall quality score. In addition, studies were evaluated for protocol violations that would prohibit volume SER analysis. QC ratings are stored in a DICOM sequence, attribute tag (0117,1024), with each separate QA rating contained in an item in this sequence, as described in Table 3. In addition, overall protocol compliance is stored in separate fields as listed in Table 3. Table 4 gives details for the different QA factors.

Table 3: DICOM fields for quality assessment

Name	Description	VR (VM) *	DICOM Tag
QC Sequence	Sequence of items for each QC factor evaluated	SQ	(0117,1024)
> QC Type	Type of quality assessment. Defined terms: GRADE (AA=4.0 to FF=0.0) PF (Pass/Fail or Yes/No) SCORE (integer ratings 1,2,…)	CS	(0117,10C0)
> QC Factor	Quality factor evaluated	LO	(0117,10C1)
> QC value	Numerical quality assessment	DS	(0117,10C2)
> QC meaning	Meaning of quality assessment	CS	(0117,10C3)
> QC comment	Additional quality assessment comments	LT	(0117,10C4)
Protocol compliance	Protocol compliance sufficient for volume SER calculation	CS	(0117,10C5)
Protocol non-compliance reasons	Description of protocol compliance violation(s)	LO (1-n)	(0117,10C6)

* VR = Value representation VM = Value multiplicity
see DICOM Standard Part 3.5 for definitions of value representations:
(http://medical.nema.org/medical/dicom/current/output/chtml/part05/PS3.5.html)

Table 4: Quality assessment factors for the I-SPY 1 / ACRIN 6698 data set

Factor (0117,10C1)	Description	Type (0117,10C0)
Fat sat	Quality of fat suppression. Integer scores: 1 = Poor; 2 = Moderate; 3 = Good	SCORE
Image Quality	Quality of images aside from fat suppression 1 = Poor; 2 = Moderate; 3 = Good	SCORE
Artifact	Presence of imaging artifacts 1 = Present, 0 = Absent	PF
Overall Quality	Overall image quality for volume SER calculation 1 = Low (unusable); 2 = Intermediate; 3 = Good (usable)	SCORE

Timing information

WARNING: Timing information was determined to the best of the core lab's ability based on the meta information in the original images submitted. **Accuracy of the timing information cannot be guaranteed**. In particular, all post-contrast times are based on the assumption that the injection and the start of the 1st post-contrast scan were simultaneous, which could not be confirmed.

Timing information fields are shown in Table 5. Timing information was added to all derived image and segmentation objects.

Table 5: Scan timing information fields for dynamic contrast-enhanced (DCE) MRI

Name	Description	VR (VM)	DICOM Tag
Total phases	Number of acquired time points (phases) including a single pre-contrast acquisition	IS	(0117,1030)
Acquisition duration	Single phase acquisition duration	DS	(0117,1031)
Acquisition start times	Starting time delay in seconds for each acquisition relative to the start of the 1st post-contrast acquisition	DS (1-n)	(0117,1032)
Injection time	Assumed injection time per scanner clock	TM	(0117,1033)
Effective acquisition delay	Effective post-injection delay for each acquisition. Non-centric phase encoding is assumed, placing the effective time half way through the acquisition	DS (1-n)	(0117,1034)
SER timing indices	Indices (0-origin) of the 3 acquisitions used in the SER calculation	IS (3)	(0117,1035)
Timing information method	Method used to determine the timing acquisition. Defined terms: AUTO: Automatic based on original image meta data MANUAL: Manually input "best-guess" timing information	LO	(0117,103A)
Timing information comments	Comments on determination of timing information	LT	(0117,103B)

Tumor volume of interest (VOI)

A 3D rectangular VOI enclosing the enhancing tumor region was defined on all cases with acceptable quality and compliance for volume SER analysis. VOI are defined in the DICOM standard patient coordinate system, as defined by the Image Position

Patient (0020,0032) and Image Orientation Patient (0020,0037) fields in the original DICOM image objects. Tumor VOI attributes are described in Table 6, and are included in all derived image and segmentation objects. In cases where significant regions of non-tumor enhancement could not be excluded from the VOI without exclusion of tumor areas , "OMIT" regions of interest (ROI) were defined to mask out these regions. OMIT ROIs were defined either as 3D rectangular VOI analogous to the analysis VOI, or as 2D irregularly shaped ROIs which were projected across the 3D image along one of the 3 orthogonal image axes. OMIT regions are described in private attributes detailed in Table 7.

NOTE: The projected OMIT ROIs were defined on displayed orthogonal maximum intensity projection (MIP) images that had been interpolated to have isotropic voxel dimensions and were transposed where necessary to display in the standard radiologic orientations. Therefore, except for those projected along the z-axis (slice axis, projection axis (0117,1051) = 2) the stored X- and Y- vertices cannot be directly applied to the original images.

Table 6: DICOM Fields for rectangular VOI

Name	Description	VR (VM)	DICOM Tag
VOILPS	Patient coordinate system specified rectangular VOI Sequence	SQ	(0117,1020)
> VOILPS Center	Center of the VOI	DS (3)	(0117,1042)
> VOILPS HalfWidth	1st half dimension vector of the VOI	DS (3)	(0117,1043)
> VOILPS HalfHeight	2nd half dimension vector of the VOI	DS (3)	(0117,1044)
> VOILPS HalfDepth	3rd half dimension vector of the VOI	DS (3)	(0117,1045)
> VOILPS Type	Use for the specified region. Defined terms: VOI Region to be analyzed OMIT Region to be excluded from the analysis	CS	(0117,1046)
VOI_pixel_start *	(x,y,z) coordinates of the first voxel in the VOI	US (3)	(0117, 10A1)
VOI_pixel_end *	(x,y,z) coordinates of the last voxel in the VOI	US (3)	(0117, 10A2)

* VOI_pixel_start and VOI_pixel_end are defined in cases where the Volume SER calculation was done, on the images that were used for the calculation. In out-of-protocol cases where images were acquired in the Axial plane these analyzed images will have been reformatted, cropped and/or resampled to isotropic resolution from the original images.

Table 7: DICOM Fields for description of OMIT regions: rectangular VOI and irregular projected 2D ROIs

Name	Description	VR (VM)	DICOM Tag
OMIT regions	OMIT region sequence. Each item contains either a 3D patient-coordinate system rectangular VOI or a 2D pixel-coordinate projection ROI	SQ	(0117,1022)
> VOILPS ROI flag	Type of VOI: enumerated values: 0 rectangular VOI 1 irregular projected pixel-coordinate ROI	IS	(0117,1041)
> VOILPS item	See Table 5 for attributes for rectangular VOI		
> ProjectedROI npixels	Number of pixels for image used for ROI definition	US	(0117,1050)
> Projection axis	Image pixel axis of projection for the 2D ROI. Enumerated values: 0=x-axis, 1=y-axis, 2=z-axis	IS	(0117,1051)
> ProjectedROI transpose flag	Flag indicating ROI coordinates are defined on a transposed image	IS	(0117,1052)
> ProjectedROI X vertices *	X-axis pixel coordinates defining the irregular ROI	US (3-n)	(0117,1053)
> ProjectedROI Y vertices *	Y-axis pixel coordinates defining the irregular ROI	US (3-n)	(0117,1054)
> ProjectedROI Z range *	Z-axis (plane) range of projection of the ROI. If not present the ROI was projected across all planes in the image.	US (2)	(0117,1055)
> ProjectedROI type	Type (usage) of ROI. Defined terms: OMIT region to be excluded from the analysis	CS	(0117,1056)
> ProjectedROI label	Label for display with the ROI	LO	(0117,1057)

* ROI vertices are defined on the images that were used for the volume SER calculation. In out-of-protocol cases where images were acquired in the Axial plane these analyzed images will have been reformatted, cropped and/or resampled to isotropic resolution from the original images. Furthermore, for all ROI with projection axis 0 or 1 the transpose flag and npixels values must be used to convert the stored vertices into the original image coordinate system.

SER analysis parameters

Parameters used to specify the Volume SER calculation are stored in a DICOM sequence (0117,1010) described in Table 8. Table 9 lists the parameters used, with each parameter being described in one item in the sequence. See Appendix A for a description of the Volume SER calculation.

Table 8: DICOM sequence for storing analysis parameters

Name	Description	VR (VM)	DICOM Tag
Parameter sequence		SQ (1-n)	(0117,1010)
> Parameter type	Parameter type. Enumerated values: FLOAT, INTEGER, STRING	CS	(0117,1012)
> Parameter name	Identifies parameter	LO	(0117,1014)
> Parameter description	Description of parameter	LT	(0117,1016)
> Floating parameter value	Value of floating point parameter required for type (0117,1012) FLOAT	DS (1-n)	(0117,1018)
> Integer parameter value	Value of integer parameter required for type (0117,1012) INTEGER	IS (1-n)	(0117,1019)
> String parameter value	Value of string parameter required for type (0117,1012) STRING	LO (1-n)	(0117,101A)

Table 9: Parameters for Volumetric Signal Enhancement Ratio (VOLSER) Analysis of Dynamic Contrast-enhanced (DCE) MRI stored in Parameter sequence (0117,1010). Each item in the sequence describes one parameter.

Name (0117,1014)	Description	Type (0117,1012)
tissue_masking_method	Method used for pre-contrast selection of breast fibroglandular tissue regions. Defined terms: NONE No pre-contrast T1 masking employed MANUAL Operator set pre-contrast T1 intensity threshold PERCENT_MAX Pre-contrast T1 intensity threshold set to percentage of 95th percentile intensity in VOI FCM Tissue mask defined by fuzzy C-means analysis	STRING
pre_contrast_threshold	Intensity threshold applied to pre-contrast T1 image to select fibroglandular tissue regions. Required if tissue_masking_method is MANUAL or PERCENT_MAX	INTEGER
PCT_background_threshold	Background masking level percentage Required if tissue_masking_method is PERCENT_MAX	INTEGER
PE_threshold	PE_{thresh}: early percent enhancement threshold	INTEGER
minimum_neighbor_count	Kernel size for a minimum connectivity filter for SER analysis: voxels with fewer than this number of immediate neighbors passing the pre-contrast intensity and PE threshold tests were not included in the SER volume.	INTEGER
ser_time_correct	Flag indicating that SER values were adjusted for scan timing.	INTEGER
target_time_1 target_time_2 time_tolerance ser_correct_amp_1 ser_correct_amp_2 ser_correct_exp_1 ser_correct_exp_2	Parameters used for correction of SER values for acquisitions with significant protocol timing errors. Present if and only if ser_time_correct is present and equal to 1. For a full description see Ka-Loh Li et al, Radiology, 248 (1), July 2008, pages 79-87	FLOAT

Functional tumor volume (FTV) results

Functional tumor volume (FTV = FTV(PE_{thresh}, SER_{min}, SER_{max})) is defined as the volume of tissue within the tumor VOI, or otherwise segmented breast tissue region, with a PE greater than or equal to the early PE enhancement threshold (PE_{thresh}) and an SER greater than a specified minimum SER_{min} and less than or equal to a specified maximum SER_{max}. SER_{max} is assumed to be infinite if not specified. Calculated FTV values are stored in the DICOM segmentation objects using the sequence described in Table 10. For the I-SPY 1 / ACRIN 6657 data set two FTV are reported: FTV_{PE} (PE_{thresh}, SER_{min}=0.0, SER_{max}=∞) and FTV_{SER} (PE_{thresh}, SER_{min}=0.9, SER_{max}=∞), where PE_{thresh} was set empirically for each imaging center.

Table 10: DICOM sequence for storing functional tumor volume (FTV) results

Name	Description	VR (VM)	DICOM Tag
FTV Sequence	MRI SER FTV results	SQ (1-n)	(0117,10B0)
> SER Minimum	Minimum value of SER	DS	(0117,10B1)
> SER Maximum	Maximum value of SER: assumed to be infinite if not specified	DS	(0117,10B2)
> Voxel count	FTV number of voxels	IS	(0117,10B3)
> Volume	FTV in cc	DS	(0117,10B4)
> Label	Display label for FTV result	LO	(0117,10B5)

I-SPY Patient Clinical Data

A separate set of files will be available giving a subset of the clinical data collected on the study subjects. Table 11 describes the data provided. More extensive patient data is available through I-SPY [ref].

Table 11: Patient Data Dictionary

Variable Name	**Variable Description**	**Format**
ISPY_ID	I-SPY ID de-identifies a patient's CALGB and ACRIN ID	Integer 1001-1239
DataExtractDt	Date clinical data was downloaded from the CALGB database	Date format mm/dd/yyyy

Patient Demographics		
AgeCat	Patient Age Category 1= 18-30 2= >30-40 3= >40-50 4= >50-60 5= >60-70 6= >70-80 7= >80-<89	Number
Age	Patient Age	Number
Race_id	Patient Race 1=Caucasian 3=African American 4=Asian 5=Native Hawaiian/Pacific Islander 6=American Indian/Alaskan Native 50=Multiple race	Number
Sstat	Survival Status 7=Alive 8= Dead 9=Lost	Number
SurvDtD	Survival date (time from study entry to death or last follow-up; time unit is days	Number

RFS	Recurrence-free survival time – time from neoadjuvant chemotherapy start date until earliest: local or distant progression or death (time unit is days)	Number
RFS_ind	Recurrence-free survival indicator 1=event (local or distant progression or death) 0=censor at last follow-up	Number
ERpos	Estrogen Receptor Status (Allred Score or Community determined), pre-treatment 0=Negative 1=Positive 2=Indeterminate	Number
PgRpos	Progesterone Receptor Status (Allred Score or Community determined), pre-treatment 0=Negative 1=Positive 2=Indeterminate	Number
HR Pos	Hormone Receptor Status, pre-treatment 0=Negative for both ER and PR 1=Positive if either ER or PR was Positive 2=Indeterminate if both ER and PR were Indeterminate	Number
pCR	Pathologic Complete Response, post-neoadjuvant (no residual invasive disease in breast or lymph nodes; presence of only in situ disease are considered disease free): 0= No (did not achieve pCR) 1= Yes	Number

	Blank= no surgery	
RCBClass	Residual Cancer Burden class: 0= 0, RCB index 0 1= I, RCB index less than or equal to 1.36 2= II, RCB index greater than 1.36 or equal to 3.28 3= III, RCB index greater than 3.28 Blank= unavailable or no surgery	Number
RCB Index	Residual Cancer Burden Index: -Numerical value Blank= unavailable or no surgery	Number

Annotated Image Markup (AIM) Files

The FTV results will also be presented in AIM files accompanying the image data sets. [To be available at a future date.]

Appendix A: *Functional Tumor Volume (FTV)* [1,2,3]

Signal Enhancement Ratio (SER) is a combined enhancement/washout measure derived from dynamic contrast enhanced MRI scans. Three time-points are used: pre-contrast injection, early post-contrast, and late post-contrast. Each acquisition is a high spatial resolution, 3D, T1-weighted scan. Sequential (non-centric) phase encoding is used to ensure that the effective acquisition time for time-points 2 and 3 can be taken as the time from contrast injection to the midpoint of the MRI scan. This time is generally 0.75 to 2.5 minutes after injection for the early time-point, and 7.5 minutes or greater for the late time-point. Initial validation studies and the ACRIN 6657 protocol were done with MRI acquisition duration of 5 minutes, with post-contrast scan timings of 2.5 and 7.5 minutes.

Tumor vascularity can be characterized by the percent enhancement (PE) of a post-contrast time-point S_1, from the pre-contrast time-point S_0, which reflects contrast uptake in the tissue and is given by

$$PE_1 = \frac{S_1 - S_0}{S_0} * 100\%$$.

SER, given by the ratio of the PE at the early post-contrast time to the PE at the late post-contrast time, adds a measure of the washout rate in the tissue. SER is given by:

$$SER = \frac{PE_{early}}{PE_{late}} = \frac{S_{early} - S_0}{S_{late} - S_0}$$.

SER is a three-point approximation of the contrast-enhancement curve that has previously been shown to correlate well with tumor microvessel density and tumor grade, with promising prognostic value for breast cancer. Both PE and SER are calculated on a per-pixel basis.

We calculate functional tumor volume (FTV) using a semi-automated tumor segmentation algorithm based on the PE and SER maps. To avoid including skin and chest wall enhancement and imaging artifacts, analysis is limited to an operator selected rectangular volume of interest (VOI). The VOI is usually drawn on a set of orthogonal maximal intensity projection (MIP) images taken either from the early post-contrast image or from a subtraction image S_1-S_0. For a minority of cases it is also necessary for the operator to draw one or more irregularly shaped exclusion regions to eliminate non-tumor enhancement regions that can not be excluded with the rectangular VOI. All further processing is fully automatic. A map consisting of the SER of each voxel is calculated using 3 levels of filtering: a pre-contrast intensity background mask level set to 60% of the 95th percentile intensity of the VOI is used to reduce spurious noise and to exclude low signal regions such as suppressed adipose tissue and strongly enhancing vessels; a PE threshold, typically 70%, at the early post-contrast time point is applied to segment malignant tissue from normal appearing tissue; a connectivity test is applied to the combined background and PE threshold mask, requiring a minimal number of connected neighboring voxels, to eliminate speckle noise. An SER color map is generated for qualitative assessment, showing areas of strong enhancement and washout (SER>0.9) in a gradation of colors from white to green, while enhancing but non-washing out tissue (SER<0.9) is shown in blue. FTV_{PE} is calculated by summing the volumes of all voxels within the VOI passing all the filtering steps and having a positive SER. Inclusion of the low SER component of the map was found to be beneficial to getting a useable FTV measure in post-chemotherapy pre-surgery examinations where enhancement values are significantly depressed relative to pre-treatment values. FTV_{SER}, measured similarly but with a lower limit of SER > 0.9, giving a volume measure of the washout regions of the lesions, was also investigated.

For further information see:

1. Partridge SC, Gibbs JE, Lu Y, et al: Accuracy of MR imaging for revealing residual breast cancer in patients who have undergone neoadjuvant chemotherapy. AJR Am J Roentgenol 179:1193-9, 2002

2. Hylton NM, Blume JD, Bernreuter WK, et al: Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy--results from ACRIN 6657/I-SPY TRIAL. Radiology 263:663-72, 2012

3. ACRIN PROTOCOL 6657 / CALGB 150007 http://www.acrin.org/6657_protocol.aspx
Contrast-Enhanced Breast MRI for Evaluation of Patients Undergoing Neoadjuvant Treatment for Locally Advanced Breast Cancer