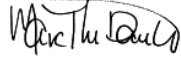


SIP Standard Specification

An Interchange Format Specification for Standard Stochastic Information Packets (SIPs) and Stochastic Library Units with Relationships Preserved (SLURPs)

Version 2.1.1

PROPOSAL APPROVED	
Title	Signature, Date
Chair, Standards Committee	 Marc Thibault, 1 June 2016

CHANGE HISTORY		
Revision	Date	Change Highlights
Version 2.0	2014-10-01	Initial Published Standard
Version 2.0.1	2015-05-05	Minor language, format and title page changes
Version 2.0.2	2015-07-20	Removed excess comma from XML definition. Fix figure C1, add default values. Restrain attribute-name. Add PM_Sips to the Annex B defined names. Add copyright attribute.
Version 2.1.0	2016-02-01	Add Annex E to extend the standard to JSON-formatted SLURPs
Version 2.1.1	2016-06-06	Fix and extend metadata definitions in Annex B

Contents

1.	Background	4
1.1.	Application.....	4
1.2.	Scope	4
1.3.	License	5
2.	Applicable Documents.....	5
2.1.	General	5
2.2.	Order of precedence	5
3.	Definitions	5
3.1.	SIP	6
3.2.	SLURP.....	6
3.3.	Coherence.....	6
4.	General Requirements.....	6
4.1.	SIP Standard Attributes	6
4.2.	Common Optional Attributes.....	7
4.3.	Optional Graph Data.....	7
4.4.	SLURPs	8
4.5.	SLURP Standard Attributes.....	8
4.6.	Data Types	8
4.7.	Versions	8
4.8.	Multi-dimensional SIPs.....	9
4.9.	Domain-specific Attributes.....	9
5.	Abbreviations	9
	Annex A. SIP/XML Exchange Format.....	12
1.	Description	12
2.	SIP Format	12
2.1.	Picture.....	12
2.2.	SIP Schema.....	13
3.	SLURP Format.....	13
3.1.	Picture.....	13

Contents

3.2. SLURP Schema	13
4. Sample SIP File	14
Annex B. Excel SIP Library Workbook	15
1. Description	15
2. Worksheet Layout	16
2.1. Library Attributes.....	16
2.2. SIP Attributes.....	16
2.3. Example of a SIP Library:	17
3. Defined Names.....	18
Annex C. Excel Worksheet SIP/CSV Format	19
1. Description	19
2. Worksheet Layout	19
2.1. Control Key/Value Table.....	20
2.2. SLURP Key/Value Table	20
2.3. SIP Key/Value Table.....	21
2.4. SIP Data Table.....	21
3. Implementation Notes	21
Annex D. Proto SIP/SLURP Format.....	22
1. Description	22
2. Proto-SIP Type 1	22
3. Proto-SIP Type 2	23
Annex E. SIP/JSON Exchange Format	24
1. Description	24
2. SIP Format	24
3. SLURP Format.....	25
4. Sample SIP/JSON File	25
Notes and Resources.....	27

1. BACKGROUND

1.1. Application

Vectors of scenarios or realizations of probability distributions have been used to drive stochastic optimization since at least 1991ⁱ. In 2005, the use of such vectors (dubbed SIPs and SLURPs) was extended to driving interactive simulations for high level decision makers at Royal Dutch Shell by Savage, Scholtes, and Zweidlerⁱⁱ, and the discipline of probability management was formalized. The approach is further described in Savageⁱⁱⁱ and Thibault^{iv}.

There are three primary advantages to representing uncertainties in this manner; communication, calculation, and credibility. First, SIPs provide an unambiguous means of communicating uncertainties across platforms, enterprises, and industries. Second, if statistical relationships are preserved through SLURPs, calculations with uncertainties just involve vector arithmetic, which requires no specialized simulation software. Third, because distributions may be estimated by credible experts, and stored as data with provenance, decision makers are “given permission” to be uncertain within auditable limits.

The current version of the standard addresses the case of equally likely scenarios. Future versions may address weighted scenarios to facilitate the simulation of rare events.

Microsoft Excel™ is prominent in the format annexes of this specification. Although the data architecture and SIP/XML format are platform agnostic, there are millions of Excel users, and many of them will be using Excel to build models with uncertain variables. This makes it effective to make use of Excel as a common language. Having a couple of Excel centric formats improves the odds that Excel implementations will be able to communicate with each other.

1.2. Scope

The purpose of this specification is to define standards for probability distributions as auditable and transportable data. The standards defined herein are the Stochastic Information Packet (SIP) and the Stochastic Library Unit with Relationships Preserved (SLURP), and some interchange formats.

This standard defines a simple, adaptable data architecture that makes it easy to create and use SIP libraries by piggybacking on common data formats including Excel worksheets, XML, JSON and CSV.

This standard defines interchange formats optimized for moving data from one process to another, with the receiving process translating the incoming data stream to whatever internal data structures are appropriate for the application.

1.3. License

This standard is freely available for use without license or fee. It is the copyrighted property of Probability Management, a non-profit corporation. It may be quoted, copied and redistributed, but may not be resold.

The latest version of this specification standard may be downloaded free at ProbabilityManagement.org.

The terms “Stochastic Information Packet,” “SIP,” “Stochastic Library Unit with Relationships Preserved,” “SLURP,” “Proto-SIP,” and “Proto-SLURP” are copyrighted marks and must not be used to describe data elements, except for those data products that comply with this specification.

“SIP Certified” is a status conferred on organizations who have been certified to produce SIPs formatted in accordance with this specification standard. Compliance alone does not constitute certification, which can only be conferred by organizations authorized by ProbabilityManagement.org.

Comments, suggestions, and corrections should be submitted by emailing the address found on the ProbabilityManagement.org website.

2. APPLICABLE DOCUMENTS

2.1. General

The following documents of the exact issue shown form a part of this specification to the extent specified herein.

2.2. Order of precedence

In the event of a conflict between the text of this document and the references cited herein, the text of this document takes precedence. Nothing in this document, however, supersedes applicable laws and regulations unless a specific exemption has been obtained.

3. DEFINITIONS

The use of ‘XML’, ‘JSON’ and ‘CSV’ in this standard does not imply that full compliance with the corresponding standards is a requirement. This document describes the small subsets of those standards actually used.

3.1. SIP

The Stochastic Information Packet (SIP) represents a probability or frequency distribution as a data structure that holds an array of values and some metadata. The values are realizations of possible values of an uncertain variable. The array for a probability distribution is composed so that the probability of each element is $1/N$ where N is the number of elements in the array.

The key benefit of using SIPs is that they are actionable, in that they may be used, as-is, in calculations. If X is a random variable represented by $SIP(X)$, and $F(X)$ is a function of X , then $SIP(F(X))=F(SIP(X))$. That is, the function F is applied sequentially to each element of $SIP(X)$. This means in effect that SIPs and the arithmetic, relational, and logical operators comprise a group.

3.2. SLURP

A coherent collection of SIPs that preserve statistical relationships between uncertainties is known as a Stochastic Library Unit with Relationships Preserved (SLURP).

3.3. Coherence

Two or more SIPs are a coherent set if the values of their corresponding samples are in some way interdependent, and that relationship is preserved in the SIPs' rank orders. For calculations with these SIPs to be valid, the alignment of the samples must be preserved; if one of the SIPs is permuted, the others must be permuted by the same permutation index to preserve coherence.

In this respect, the importance of the SLURP is that any SIP calculated with arithmetic, relational or logical operations on SIPs in a given SLURP will also be coherent and can be collected in that or a separate SLURP.

4. GENERAL REQUIREMENTS

4.1. SIP Standard Attributes

Name	Description
name	Required. A text string identifying the SIP, usually unique in context.
count	Required. The number of samples.

tmax	Optional. The maximum trial index required if the SIP uses a sparse index.
type	Required. The SIP's data encoding format. <i>See Section 4.6.</i>
ver	Required. The version of the SIP's data encoding format.

4.2. Common Optional Attributes

Name	Description
about	A description of the SIP or SLURP. Could be a URL.
avg	The average or mean of the SIP sample values before they're encoded into the string.
csvr	The number of digits to the right of the decimal for CSV conversion.
copyright	Any copyright claim
dataver	A number or date indicating the currency of the data in a SIP or SLURP.
dims	The dimensions of a multidimensional SIP. <i>See Section 4.8.</i>
max	The SIP maximum sample value.
min	The SIP minimum sample value.
offset	An offset factor to be applied to a SIP encoded value to get the sample value. The 'b' in $ax+b$. Default is 0.
origin	An arbitrary text string should say something about the institution or project that produced a SIP or SLURP.
provenance	Information about the source and authority of the data. Could be a URL.
scale	A scale factor to be applied to a SIP encoded value to get the sample value. The 'a' in $ax+b$. Default is 1.
units	A text string for the SIP data measurement units e.g. "Dollars".

4.3. Optional Graph Data

hbin	The bin width of a histogram of the SIP
------	-----------------------------------------

hmin	The minimum value in a histogram of the SIP
hnum	The number of bins in a histogram of the SIP
hvalN	The value in the Nth bin in a histogram of the SIP
Ptile	The (P/100) percentile value

4.4. SLURPs

A collection of SIPs can comprise a SLURP if the statistical relationships between SIPs are preserved.

Two attributes are required: name and coherent.

4.5. SLURP Standard Attributes

Name	Description
name	Can be any string, should be a unique identifier in context.
coherent	Must be either “true” or “false”. If false, the coherence of the included SIPs is not assured.

If there’s a “count” attribute, it should refer to the number of SIPs in the SLURP.

4.6. Data Types

The *type* attribute refers to the SIP data encoding format. The attribute *type=“CSV”* says that the data in this SIP is encoded as a basic comma separated values string.

4.7. Versions

Version numbers will follow the generally accepted dotted format *major.minor.patch*. A major version number change will signal a version that doesn’t guarantee backward compatibility; it might break an application. A minor version number change will signal an improvement or upgrade that preserves backward compatibility (e.g. extending the CSV type to handle the European use of dot and comma.) A patch version number change indicates an improvement to the text of the standard that has no functional effect on its implementation.

4.8. Multi-dimensional SIPs

The attribute *dims* holds a comma-delimited list of dimensions for a multi-dimensional SIP. The obvious application is for a time series, where the first dimension is time periods and the last is samples.

The list is in slow-moving-first order, so that the dimensions list matches the indices referring to the last sample. E.g.

```
dims="12,2000"
```

defines a SIP composed with 24,000 samples organized so that the first sample is (1,1), the second sample is (1,2), the 4001st sample is (3,1), and the last sample is (12,2000).

In other words, if the dimensions are (p,q), the index of sample (x,y) is $y+(x-1)*q$.

Also, *dims* and *count* should match so that in this example, `count="24000"`.

The last dimension is always the trials dimension.

Note that a multi-dimensional SIP is one SIP; the SIP metadata applies to the whole SIP, so there's no explicit way, other than order, to distinguish the dimensions.

4.9. Domain-specific Attributes

Communities of interest with different sources of data may require additional attributes. The attributes needed for specific application domains can be standardized to promote open standards and to avoid data fragmentation. Communities of interest or domain specific users should propose the specification items and other relevant resources to Probabilitymanagement.org. Domain-specific extensions to this specification may be included in later updates.

Probabilitymanagement.org has put into place a process for proposing and agreeing to such standards.

5. ABBREVIATIONS

CSV	Comma Separated Value String
JSON	JavaScript Object Notation
SIP	Stochastic Information Packet
SLURP	Stochastic Library Unit with Relationships Preserved

XML	Extensible Markup Language
-----	----------------------------

Annexes

ANNEX A. SIP/XML EXCHANGE FORMAT

1. DESCRIPTION

This format is in active use and has significant open source code to support it.

The SIP/XML (SIP over XML) format uses minimal subsets of the XML and CSV standards to hold SIPs and SLURPs. It is intended to be platform-agnostic and easily implemented on commonly available systems and languages.

The SIP/XML format encapsulates an array of sample values and related metadata as strings in a text file or data structure. It has been implemented and tested in MatLab and Excel, and Excel workbooks with code and tests are available.

The XML tag is <SIP>. The value element is the SIP value array formatted as a comma-separated values (CSV) string. The *type* attribute is “CSV”.

Each has required and optional standard attributes in the start tag, and arbitrary attributes can be added to meet specific requirements. As is the norm with XML, any attributes that aren't recognized by a particular application should be silently ignored by that application. *To be fully XML compliant, the first character of the attribute name should not be a digit, “-”(dash) or “.”(period).*

In object-oriented terms, a particular SIP is an instance of the Sample Distribution class, and the XML string is a serialization of the instance state.

A collection of SIPs is encapsulated in a SLURP. Its tag is <SLURP>. The attributes are collection attributes. Text prior to the line starting with the <SLURP tag is not defined by this standard. Each enclosed SIP element must begin on a new line.

The SIP and SLURP schema are presented using Compact Relax NG notation (<https://www.oasis-open.org/committees/relax-ng/compact-20020607.html>).

2. SIP FORMAT

2.1. Picture

```
<SIP name="$ $" count="##" type="CSV" ver="1.0.0" csvr="##" ... >
  CSV Encoded SIP Value array
</SIP>
```

2.2. SIP Schema

```
SIP = element SIP {
  ( attribute name { string } &
    attribute count { integer } &
    attribute type { "CSV" } &
    attribute ver { "1.0.0" } ),
  attribute * { * } * ,
  { string }+ }
```

3. SLURP FORMAT

3.1. Picture

```
<SLURP name="$$" count="###" coherent="true" about="$$" >
  <SIP name= ...
  <SIP name= ...
  ...
</SLURP>
```

3.2. SLURP Schema

```
SLURP = element SLURP {
  ( attribute name { string } &
    attribute coherent { boolean } ) ,
  attribute * { * } * ,
  SIP+ }
```

4. SAMPLE SIP FILE

```
<SLURP name="exampleSLURP" count="2" coherent="true"
provenance="example SLURP provenance">
  <SIP name="Domestic" count="10" type="CSV" csvr="1"
ver="1.0.0" provenance="Data from XYZ Co." average="4.2"
median="4.5">
    3.5,7.4,4.4,4.6,0.7,4.3,4.8,4.7,4.7,2.9
  </SIP>
  <SIP name="Foreign" count="10" type="CSV" csvr="1"
ver="1.0.0" provenance="Data from XYZ Co." average="5.0"
median="4.9">
    6.2,1.1,4.8,5.0,6.0,7.8,7.0,4.5,4.6,3.0
  </SIP>
</SLURP>
```

ANNEX B. EXCEL SIP LIBRARY WORKBOOK

1. DESCRIPTION

This format is in active use and has significant open source code to support it.

The Excel SIP Library is an all-Excel approach to the standard that uses Excel-specific features. A model in one Excel workbook will refer to SIP data in one or more library workbooks accessible as a common resource.

A SIP Library is an Excel workbook containing the following:

- a) Required elements
 - A set of one or more SIPs, each including a name and possibly a provenance string.
 - A *count* of the number of trials in each SIP, stored in a cell named PM_Trials. This is the “count” attribute of the SIP. It applies to all the SIPs in the library.
- b) Optional elements
 - A *coherent* flag (True/False) stored in a cell named PM_Coherent, indicating whether the SIPs in the library are guaranteed to be coherent. Default is True.
 - A *library provenance* string for the Library as a whole, stored in a cell named PM_Lib_provenance, containing information about the source and authority of the data.
 - A table of metadata names and indices for the SIP Library. The metadata names are in a range named PM_Meta. The metadata indices are in a range named PM_Meta_Index. This is also where the provenance of the individual SIPs is contained, as described further in section 2.
 - A *type* cell named PM_Type, containing the value "Excel_range". This applies to all SIPs in the library. If PM_Type is not present, the default is presumed to be "Excel_range".
 - A *version* cell named PM_Ver, containing an identifier for the format version.

Because they are addressed by Excel defined names, these elements need not be located on the same sheet.

2. WORKSHEET LAYOUT

The Library elements are laid out as follows:

2.1. Library Attributes

The *count*, *library provenance* (optional), *coherent* (option), *type* (optional), and *version* (optional) values may be placed in any convenient cells, but usually near the top left corner of a worksheet. The cells holding the values must have the range names PM_Trials, PM_Lib_provenance, PM_Coherent, PM_Type, PM_Version respectively.

2.2. SIP Attributes

The *SIPs* are arranged in a contiguous block of rows or columns. The first element of each row (respectively column) is the SIP name, which should be a valid Excel range name. The 2nd through *count*+1st elements are the values of the SIP. The *count*+2nd element and following may contain SIP metadata, generally statistical data such as averages or percentile values. If there are SIP provenances, they may be put into this metadata section. Each SIP together with its metadata, will be given a separate range name, which should be the same as the SIP name. The top row or leftmost column of the block of sips contains the values 1, 2, 3, ... *count*, 1 being placed above or to the left of the 1st data element of the SIP.

The table of *metadata* indices consists of 2 ranges, the PM_Meta range and the PM_Meta_INDEX range (note that in earlier versions of the standard, these ranges were named PM_IV (for index value) and PM_IV_Index). PM_Meta is a column of cells containing a list of names of metadata elements, e.g. “Average, 10th Percentile, 20th Percentile”, “Provenance” each element in a separate cell. PM_Meta_INDEX is a column of cells which give the location of the respective metadata elements listed in PM_Meta as indices into the block of SIPs, counting the first data element of the SIP as 1. For example, if the SIPs have 10,000 elements each and the average value is put into the 1st cell following the last data element of the SIP, PM_Meta could contain “Average” and the corresponding row of PM_Meta_INDEX would contain 10001. PM_Meta and PM_Meta_INDEX must be adjacent columns, PM_Meta to the left of PM_Meta_INDEX.

2.3. Example of a SIP Library:

	A	B	C	D	E	F
1	PM_Trials	10			PM_Meta	PM_Meta_Index
2	PM_Lib_Provenance	example SLURP provenance			Average	11
3	PM_Coherent	TRUE			Median	12
4	PM_Type	Excel_range			Provenance	13
5	PM_Ver	2.0.0				
	PM_Sips	'B7				
6						
7		Trials	Domestic	Foreign		
8		1	3.5	6.2		
9		2	7.4	1.1		
10		3	4.4	4.8		
11		4	4.6	5		
12		5	0.7	6		
13		6	4.3	7.8		
14		7	4.8	7		
15		8	4.7	4.5		
16		9	4.7	4.6		\
17		10	2.9	3.0		
18		Average	4.2	5.0		
19		Median	4.5	4.9		
20		Provenance	Data from XYZ Co.	Data from XYZ Co.		

Table B.1 Example

In Table B.1, PM_Trials is cell B1. PM_Sips is the top left-hand corner of the SIP table, C7. PM_Lib_provenance is B2. PM_Coherent is B3. PM_Type is B4. PM_Ver is B5. PM_Meta is cells E2:E4. PM_Meta_INDEX is cells F2:F4. C8:C17 is a range named Domestic. D8:D17 is a range named Foreign. C8:C20 is a range named Domestic.MD. D8:D20 is a range named Foreign.MD.

The XML representation of this library is the one shown in Annex A, section 4.

3. DEFINED NAMES

NAME	DESCRIPTION
PM_Trials	Single cell containing the count for each SIP in the library. This overrides the SIP count attribute.
PM_Sips	The top left-hand corner of the SIPs table, including the names row.
PM_Lib_provenance	Single cell containing the string describing the SIP library provenance.
PM_Type	Single cell containing type name for the library. By default Excel_range
PM_Ver	Single cell containing version number of the library type.
PM_Coherent	Single cell containing TRUE or FALSE. If absent, the default is True.
PM_Meta	A column of 1 or more cells, each cell containing the name of one type of metadata for the SIPs. This range must have the same number of cells as PM_Meta_INDEX as be placed just to the left of PM_Meta_INDEX..
PM_Meta_INDEX	A column of 1 or more cells, the same size as PM_Meta. Each cell contains the index number of the metadata named in the corresponding cell of PM_Meta. This range must be just to the right of PM_Meta
<i>SIPname</i> .MD	Data and Metadata for a SIP; this will be a row or column of length (height) equal to PM_Trials+(size of PM_Meta)

ANNEX C. EXCEL WORKSHEET SIP/CSV FORMAT

1. DESCRIPTION

SIP/CSV (SIP over CSV) specifies an open standard SIP format compatible with Excel’s worksheet CSV file I/O for exchange. The sheet format includes a layout specification for the SLURP metadata and the SIP data and metadata. The CSV file format is simple and easily generated; any other application could generate the CSV file for consumption by an Excel application, or vice-versa.

2. WORKSHEET LAYOUT

	A	B	C	D	E	F	G	H	I	
1	Sample SIP over CSV file						Always have something in A			
2										
3	Control		SheetName	exampleSLURP			SIPs go on this worksheet			
4			FilePath	SIPCSVsample.csv			Associated file name			
5			SlurpAttrs	C14:C17			Put the SLURP metadata her			
6			SipAttrs	C21:C28			Put the SIP metadata here a			
7			SipTlc	D30			SIPs Array Top Left Corner			
8			ClearFirst	TRUE			Clear the data array before t			
9			numSamples	1000			This many rows of SIP samp			
10										
11										
12										
13										
14	SLURP		name	exampleSLURP						
15			count	2						
16			coherent	TRUE						
17			provenance	exampleSLURPprovenance						
18										
19										
20					1	2				
21	SIP		name	Domestic		Foreign				
22			provenance	Data from XYZ Co.		Data from XYZ Co.				
23			count	1000		1000				
24			type	csv		csv				
25			csvr	1		1				
26			ver	1		1				
27			average	4.2		5				
28			median	4.5		4.9				
29										
30	SIPs			3.5		6.2				
31				7.4		1.1				
32				4.4		4.8				
33				4.6		5				

Figure C.1. Example

This format is in active use for reading SIP libraries into a major Excel/VBA application developed for the Canadian Armed Forces by Lockheed Martin. The library files are built to the SIP/XML standard (Annex A) and read into Excel worksheets laid out using this standard to control and position content.

The worksheet format involves three parts: a control table (C3:D9), a SLURP area and a SIP area. The cell ranges for the last two are defined in the control table.

Figure C.1 shows an example worksheet.

2.1. Control Key/Value Table

The control table is a key/value table with the following items:

SheetName	The name of the sheet that has the SIPs. It's normally the same as the current sheet but it could be a different sheet. This makes it possible to have a worksheet with nothing but SIP elements, keeping the metadata separate.
FilePath	The path name of a file associated with the SLURP. Depending on how the sheet is being used, this could be where the data came from or where it is to be written, or blank.
SlurpAttrs	The cell range defining the SLURP attributes to be included. The range has the attribute names. The attributes are in the column to the right of the attribute names.
SipAttrs	The cell range defining the SIP attributes to be included. The range has the attribute names. The attributes are in the columns to the right of the attribute names, positioned over the corresponding SIPs.
SipTlc	The top left corner cell of the SIP data table. The full extent of the table is determined by the number of SIPs (SLURP.count) and the number of trials (numSamples).
ClearFirst	If this is TRUE, the SLURP and SIP data ranges should be cleared before reloading the data.
numSamples	The number of samples to be taken from each SIP.

The cell ranges in this table should be entered as text with a leading apostrophe (').

2.2. SLURP Key/Value Table

The SLURP table (C14:D16 in Figure C.1) has the desired SLURP attribute keys and values. Note that the count is the number of SIPs.

2.3. SIP Key/Value Table

The SIP table (C20:I21 in Figure C.1) has the desired attribute keys and the values for each SIP.

2.4. SIP Data Table

The SIP data table (D27:I6026 in Figure C.1) has the SIP samples, one SIP per column, one trial per row.

3. IMPLEMENTATION NOTES

This format does not rely on Defined Names or formulas; it can be saved as a CSV file from Excel's SaveAs menu. The resulting file can be opened in Excel and, except for formatting and formulas, it will be restored exactly. Cells with formulas will produce their values (like Paste|Special|Values applied to the whole sheet).

The CSV file, being plain text is easily read or written by code in any programming language.

The key/value tables simplify references to data and the blank rows and columns around them make it easy to identify their extents and to load internal hash tables for efficiency (e.g. VBA's Dictionary object).

In macro-free Excel, LOOKUP() can be used to find values, INDIRECT() to address the cell ranges, and array formulas to process the SIPs.

Always have something in A1, in order to position the start of the CSV encoding.

All the tables should be surrounded by blank rows and columns and the control table should start with the first non-blank cell in column C.

ANNEX D. PROTO SIP/SLURP FORMAT

1. DESCRIPTION

The purpose of this specification is to define standards for the Proto Stochastic Information Packet (Proto-SIP) and the Proto Stochastic Library Unit with Relationships Preserved (Proto-SLURP).

Proto-SIPs and Proto-SLURPs are data structures with data in an array, but which fail to fully comply with SIP and SLURP specification requirements. This appendix defines means by which SIP-like information can be shared, while preserving the potential for the array to be modified and brought into SIP compliance. XML formats will not use <SIP> or <SLURP> tags.

Use of the terms “Proto-SIP” and “Proto-SLURP” should only be used to describe data structures as described in this annex.

Two types of Proto-SIPs are defined in this annex. Proto-SLURPs are not defined. A collection of Proto-SIPs, with relationships preserved, and otherwise conforming to SLURP standards is a Proto-SLURP. Therefore, this annex deals with Proto-SIPs explicitly, but Proto-SLURPs are implicit.

2. PROTO-SIP TYPE 1

The first type of Proto-SIP uses non-conforming delimiters, and applies only to the CSV SIP.

RFC 4180 allows the use of quotes to contain strings which include commas (i.e., commas which are part of the data string, and which are not delimiters). Some database systems cannot generate quotes in this way. In order to deal with issues such as commas in data fields and to provide predictable, statistically improbable delimiters in a Common Format and MIME Type for CSV Files, the following alternative padded delimiters are acceptable alternatives to commas for Proto-SIP CSV usage:

- a. `,|,`
- b. `,*|*,`
- c. `,}^[{,`

Delimiter a. might be generated by instructing the database report generator to output a comma delimited field followed by the character “|” which in turn would be automatically followed by a delimiting comma. The first comma, the character for vertical bar, and the second comma, together make up the three character string which becomes a delimiter. If this three character string could be present in the data, the other delimiters can be used as an alternative.

3. PROTO-SIP TYPE 2

The second type of Proto-SIP fails to provide the complete meta-data descriptions required for SIP compliance. In addition, it may fail to provide compliant CSV delimiters.

The minimum information required to constitute a Proto-SIP is the “name” information, and the data array. For Proto-SIPs, the preferred additional information is the “count” information. For Proto-SLURPS, the coherence designator is preferred additional information.

If a Proto-SIP or Proto-SLURP is generated in Excel or XML, it must comply with formatting and file type standards. In these file formats, missing metadata elements are the distinction between SIPs and Proto-SIPs.

ANNEX E. SIP/JSON EXCHANGE FORMAT

1. DESCRIPTION

This format is supported by reference code in Excel/VBA and JavaScript.

The SIP/JSON (SIP over JSON) format uses minimal subsets of the ECMA-404 JSON and CSV standards to hold SIPs and SLURPs. It is intended to be platform-agnostic and easily implemented on commonly available systems and languages. See json.org for the JSON standard.

The SIP/JSON format encapsulates a collection of sample values and related metadata as attribute strings in a text file or string data structure.

The SIP and SLURP attribute "*instanceof*" is used to identify the object type as either "SIP" or "SLURP". The SIP "*value*" attribute is the SIP value array formatted as a comma-separated values (CSV) string. The "*type*" attribute is "CSV".

Each has required and optional standard attributes, and arbitrary attributes can be added to meet specific requirements. Any attributes that aren't recognized by a particular application should be silently ignored by that application. *The first character of the attribute name should not be a digit, "-"(dash) or "."(period).*

In object-oriented terms, a particular SIP is an instance of the Sample Distribution class, and the JSON string is a serialization of the instance state.

A collection of SIPs is encapsulated in a SLURP. Its *instanceof* attribute is SLURP. Each enclosed SIP object must begin on a new line.

2. SIP FORMAT

A SIP is encoded as an object with standard attributes. The SIP sample values are encoded as an array.

```
{ "instanceof": "SIP",
  "name": "$$",
  "count": "##",
  "type": "CSV",
  "ver": "1.0.0",
  "csvr": "##",
  Etc.
  "value": [ CSV Encoded SIP value array ]
}
```


Note: “instanceof” must be the first attribute and “value” must be the last.

3. SLURP FORMAT

A SLURP is encoded as an object with standard attributes. Its SIP collection is encoded as an array of SIP objects.

```
{ "instanceof":"SLURP",
  "name":"$$",
  "count":"##",
  Etc.
  "sips":[
    {"instanceof":"SIP", ... },
    {"instanceof":"SIP", ... },
    {"instanceof":"SIP", ... }
  ]
}
```

Note: “instanceof” must be the first attribute and “sips” must be the last.

4. SAMPLE SIP/JSON FILE

```
{ "instanceof":"SLURP",
  "name":"IncomeSources",
  "count":"2",
  "coherent":"true",
  "provenance":"Source Data Provenance",
  "sips":[
    {"instanceof":"SIP",
     "name":"Domestic",
     "count":"10",
     "type":"CSV", "csvr":"1",
     "ver":"1.0.0",
     "provenance":,
     "Data from XYZ Co.",
     "average":"4.2",
```

```
    "median": "4.5",
    "value":
      [3.5, 7.4, 4.4, 4.6, 0.7,
       4.3, 4.8, 4.7, 4.7, 2.9]
  },
  {"instanceof": "SIP",
   "name": "Foreign",
   "count": "10",
   "type": "CSV", "csvr": "1",
   "ver": "1.0.0",
   "provenance": ,
   "Data from XYZ Co.",
   "average": "5.0",
   "median": "4.9",
   "value":
     [6.2, 1.1, 4.8, 5.0, 6.0,
      7.8, 7.0, 4.5, 4.6, 3.0]
  }
]
}
```

NOTES AND RESOURCES

ⁱ Scenario optimization, Ron S. Dembo, Annals of Operations Research, 1991, Volume 30, Issue 1, pp 63-80 - <http://link.springer.com/article/10.1007%2FBF02204809>

ⁱⁱ Probability Management, Sam Savage, Stefan Scholtes and Daniel Zweidler, OR/MS Today, February 2006, Volume 33 Number 1 - <http://www.lionhrtpub.com/orms/orms-2-06/frprobability.html>

ⁱⁱⁱ The Flaw of Averages, Why we Underestimate Risk in the Face of Uncertainty, Sam Savage, John Wiley 2009

^{iv} Calculating Uncertainty: Probability Management with SIP Math, John Marc Thibault, 2013