

Risa: Building R objects from local ISA-Tab files

Alejandra Gonzalez-Beltran and Steffen Neumann and Audrey Kauffmann
and Gabriella Rustici and Philippe Rocca-Serra and Eamonn Maguire
and Susanna-Asunta Sansone
isatools@googlegroups.com

May 19, 2021

1 Introduction

The Risa package is part of the ISA infrastructure software suite (<http://isa-tools.org>). It provides functionality to read ISA-Tab datasets, described in the following section. The source code and latest version can be found in the GitHub repository <https://github.com/ISA-tools/Risa>. Please, submit all 'bugs' and feature requests through <https://github.com/ISA-tools/Risa/issues>.

2 ISA-Tab format

The Investigation / Study / Assay (ISA) Tab-delimited (Tab) format is a general purpose framework with which to collect and communicate complex metadata (i.e. sample characteristics, technologies used, type of measurements made) from experiments employing a combination of technologies (<http://isa-tools.org>). In particular, ISA-Tab has been developed for - but not limited to - experiments using genomics, transcriptomics, proteomics or metabol/nomics techniques (the 'omics').

ISA-Tab uses three types of file to capture the experimental metadata:

- *Investigation file*
- *Study file*
- *Assay file* (with associated data files).

The Investigation file contains an overall description of an experiment while all experimental steps are described in the Study and in the Assay file(s). For each Investigation file there may be one or more Study files; for each Study file there may be one or more Assay files.

2.1 Investigation file

In this file, information is reported on a per-column basis and the fields are organized and divided in sections. The Investigation file is intended to meet three needs:

- to define key entities, such as factors, protocols, parameters, which may be referenced in the other files;
- to relate Assay files to Study files; and optionally,
- to relate each Study file to an Investigation (when two or more Study files need to be grouped). The declarative sections cover general information such as contacts, protocols and equipment, and also - where applicable - the description of terminologies (controlled vocabularies or ontologies) and other annotation resources that were used.

2.2 Study file

In this file, information is structured on a per-row basis with the first row being used for column headers. The Study file contains contextualizing information for one or more assays, for example; the subjects studied; their source(s); the sampling methodology; their characteristics; and any treatments or manipulations performed to prepare the specimens.

2.3 Assay file

In this file, as for the Study file, fields are organized on a per-row basis with the first row being used for column headers. The Assay file represents a portion of the experimental graph (i.e., one part of the overall structure of the workflow); each Assay file must contain assays of the same type, defined by the type of measurement (i.e. gene expression) and the technology employed (i.e. DNA microarray). Assay-related information includes protocols, additional information relating to the execution of those protocols and references to data files (whether raw or processed).

For easy transfer, ISA-Tab files and associated data files can be packaged into an ISAarchive, using a standalone Java application named ISAcreator (<http://isatab.sourceforge.net>). In order to facilitate identification of ISA-Tab components in an ISAarchive, specific extensions have been created as follows:

- *i_iname.txt* for identifying the Investigation file
- *s_sname.txt* for identifying Study file (s)
- *a_aname.txt* for identifying Assay file (s)

where 'iname', 'sname', 'aname' are the user-given names for the investigation, study/ies, assay(s), respectively.

3 The Risa package

The Risa package is used to build R objects from an ISA archive or dataset. The output is a list of objects containing, for example, the investigation, studies and assays filenames, the contents of their files, the list of samples, among other things.

These objects can then be used by downstream Bioconductor packages for data analysis and visualization (i.e. xcms). The package currently includes the function `processAssayXcmsSet` that, for a specific mass spectrometry assay, builds an `xcmsSet` object.

3.1 Building an R object from a local ISA dataset

If you have your own ISA archive, you can use the function `readISAtab` to convert it into an R object. The arguments for the function `readISAtab` are:

- `path` the name of the directory containing ISAtab files. The default is the working directory.
- `verbose` a boolean indicating to show messages for the different steps, if TRUE, or not to show them, if FALSE

As an example, we can use the *faahKO* dataset, whose version 1.2.11 contains an ISA dataset describing the experiment. First, it is required to load the *Risa* package, and the *faahKO* package must have been installed.

```
> library(Risa)
> require(faahKO)
```

Then, we read the ISA-Tab data set from the *faahKO* package:

```
> faahkoISA <- readISAtab(find.package("faahKO"))
```

The object `faahkoISA` belongs to the *ISAtab* class, and contains the following elements:

- `path` - the path of the ISA-Tab dataset,
- `investigation.filename` - the name of the Investigation file
- `investigation.file` - a data frame with the contents of the Investigation file
- `study.identifiers` - the list of study identifiers
- `study.fileNames` - the names of the study files
- `study.files` - a list of data frames with the contents of the study files
- `assay.fileNames` - the names of the assay files
- `assay.fileNames.per.study` - the names of the assay files according to the study they belong to
- `assay.files` - a list of data frames with the contents of the assay files
- `assay.files.per.study` - a list of data frames with the contents of the assay files divided per study they belong to
- `assay.technology.types` - a list with the technology types corresponding to each assay
- `assay.measurement.types` - a list with the measurement types corresponding to each assay
- `data.fileNames` - a list with the names of the data files
- `samples` - a list with the names of the samples
- `samples.per.assay.filename` - the samples classified according to the assay filename they belong to
- `assay.fileNames.per.sample` - the names of the assay files classified per sample name
- `sample.to.rawdatafile` - the association between samples and raw data files
- `sample.to.assayname` - the association between samples and assay names
- `rawdatafile.to.sample` - the association between raw data files and samples
- `assayname.to.sample` - the association between assay names and samples

Additionally, the ISA dataset could be compressed in a .zip file. If that is the case, the function `readISAtab` can be used, passing the `zipfile` as parameter. The only condition is that the ISA-Tab files are contained directly into the zip file, i.e. not inside additional folders.

In this case, the parameters for the function `readISAtab` will be:

- `zipfile` a zip archive containing ISAtab files.
- `path` the name of the directory in which the files from the zip archive will be extracted. The default is the working directory.
- `verbose` a boolean indicating to show messages for the different steps, if `TRUE`, or not to show them, if `FALSE`

Building xcmsSets for mass spectrometry assays

The function `processAssayXcmsSet` allows to build an `xcmsSet` (object defined in the `xcms` package) from the information in an assay file.

The parameters for this function are:

- `isa`: an ISA object, as retrieved by the function `readISAtab`
- `assay.filename` the name of the assay file with information about the relevant assay
- ... extra arguments that can be passed down to the `xcmsSet` function from the `xcms` package

Using the *faahKO* package as an example, we select the name of assay file, and use the `processAssayXcmsSet` to build a object of type *xcmsSet*:

```
> assay.filename <- faahkoISA["assay.names"][1]
> faahkoXset <- processAssayXcmsSet(faahkoISA, assay.filename)
```

Augmenting the ISA-Tab dataset after analysis

The *Risa* package also provides the functionality to augment the original ISA-Tab dataset with more information after analysis.

The function `updateAssayMetadata` allows to modify the metadata in a particular assay file. The arguments are:

- `isa` An isatab object, as retrieved by the `readISAtab` function.
- `assay.filename` the filename of the assay file to be augmented/modified
- `col.name` the name of the column of the assay file to be modified
- `values` the values to be added to the column of the assay file: it could be a single value, and in this case the value is repeated across the column, or it could be a list of values (whose length must match the number of rows of the assay file)

To continue with our example using the *faahKO* data package, we will assume that the results of analysis are stored in the file *faahkoDSDF.txt*. Then, we will update the ISA-Tab dataset adding the result file into the 'Derived Spectral Data File' column of the assay file.

```
> updateAssayMetadata(faahkoISA, assay.filename, "Derived Spectral Data File", "faahkoDSDF.txt" )
```

An object of class "ISATab"

Slot "path":

```
[1] "/Library/Frameworks/R.framework/Versions/4.1/Resources/library/faahKO"
```

Slot "investigation.filename":

```
[1] "i_Investigation.txt"
```

Slot "investigation.file":

	V1
1	ONTOLOGY SOURCE REFERENCE
2	Term Source Name
3	Term Source File
4	Term Source Version
5	Term Source Description
6	INVESTIGATION
7	Investigation Identifier

8	Investigation Title
9	Investigation Description
10	Investigation Submission Date
11	Investigation Public Release Date
12	Comment [Created with configuration]
13	Comment [Last Opened With Configuration]
14	INVESTIGATION PUBLICATIONS
15	Investigation PubMed ID
16	Investigation Publication DOI
17	Investigation Publication Author List
18	Investigation Publication Title
19	Investigation Publication Status
20	Investigation Publication Status Term Accession Number
21	Investigation Publication Status Term Source REF
22	INVESTIGATION CONTACTS
23	Investigation Person Last Name
24	Investigation Person First Name
25	Investigation Person Mid Initials
26	Investigation Person Email
27	Investigation Person Phone
28	Investigation Person Fax
29	Investigation Person Address
30	Investigation Person Affiliation
31	Investigation Person Roles
32	Investigation Person Roles Term Accession Number
33	Investigation Person Roles Term Source REF
34	STUDY
35	Study Identifier
36	Study Title
37	Study Description
38	Study Submission Date
39	Study Public Release Date
40	Study File Name
41	STUDY DESIGN DESCRIPTORS
42	Study Design Type
43	Study Design Type Term Accession Number
44	Study Design Type Term Source REF
45	STUDY PUBLICATIONS
46	Study PubMed ID
47	Study Publication DOI
48	Study Publication Author List
49	Study Publication Title
50	Study Publication Status
51	Study Publication Status Term Accession Number
52	Study Publication Status Term Source REF
53	STUDY FACTORS
54	Study Factor Name
55	Study Factor Type
56	Study Factor Type Term Accession Number
57	Study Factor Type Term Source REF
58	STUDY ASSAYS
59	Study Assay Measurement Type
60	Study Assay Measurement Type Term Source REF
61	Study Assay Measurement Type Term Accession Number

62	Study Assay Technology Type
63	Study Assay Technology Type Term Source REF
64	Study Assay Technology Type Term Accession Number
65	Study Assay Technology Platform
66	Study Assay File Name
67	STUDY PROTOCOLS
68	Study Protocol Name
69	Study Protocol Type
70	Study Protocol Type Term Accession Number
71	Study Protocol Type Term Source REF
72	Study Protocol Description
73	Study Protocol URI
74	Study Protocol Version
75	Study Protocol Parameters Name
76	Study Protocol Parameters Name Term Accession Number
77	Study Protocol Parameters Name Term Source REF
78	Study Protocol Components Name
79	Study Protocol Components Type
80	Study Protocol Components Type Term Accession Number
81	Study Protocol Components Type Term Source REF
82	STUDY CONTACTS
83	Study Person Last Name
84	Study Person First Name
85	Study Person Mid Initials
86	Study Person Email
87	Study Person Phone
88	Study Person Fax
89	Study Person Address
90	Study Person Affiliation
91	Study Person Roles
92	Study Person Roles Term Accession Number
93	Study Person Roles Term Source REF

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

37 Enzymes regulate biological processes through the conversion of specific substrates to product

38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75

76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71

72 A 2:1:1 CHCl₃/MeOH/1% NaCl solution (8 mL per brain and 4 mL per spinal cord in 8 mL vials) wa

73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89

90
91
92
93

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72 LC-MS analysis was performed using an Agilent 1100 LC-MSD SL instrument. For the LC analysis,

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

1

2

3

4

5

6

7

8

9

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63

64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

The analysis of the resulting total ion chromatogram was performed manually by generating extr

V6 V7

PATO EFO

v 1.26

Phenotypic qualities (properties) ArrayExpress Experimental Factor Ontology

24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77

labeling
labeling

78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93

V8

1 NA
2 NA
3 NA
4 NA
5 NA
6 NA
7 NA
8 NA
9 NA
10 NA
11 NA
12 NA
13 NA
14 NA
15 NA
16 NA
17 NA
18 NA
19 NA
20 NA
21 NA
22 NA
23 NA
24 NA
25 NA
26 NA
27 NA
28 NA
29 NA
30 NA
31 NA
32 NA
33 NA
34 NA
35 NA
36 NA
37 NA

38 NA
39 NA
40 NA
41 NA
42 NA
43 NA
44 NA
45 NA
46 NA
47 NA
48 NA
49 NA
50 NA
51 NA
52 NA
53 NA
54 NA
55 NA
56 NA
57 NA
58 NA
59 NA
60 NA
61 NA
62 NA
63 NA
64 NA
65 NA
66 NA
67 NA
68 NA
69 NA
70 NA
71 NA
72 NA
73 NA
74 NA
75 NA
76 NA
77 NA
78 NA
79 NA
80 NA
81 NA
82 NA
83 NA
84 NA
85 NA
86 NA
87 NA
88 NA
89 NA
90 NA
91 NA

92 NA
93 NA

Slot "investigation.identifier":
[1] ""

Slot "study.identifiers":
[1] "Global metabolite profiling of faah(-/-) mice"

Slot "study.titles":
[1] "Global metabolite profiling of faah(-/-) mice"

Slot "study.descriptions":
[1] "Enzymes regulate biological processes through the conversion of specific substrates to products"

Slot "study.contacts":
[1] " "

Slot "study.contacts.affiliations":
[1] ""

Slot "study_filenames":
Global metabolite profiling of faah(-/-) mice
"s_Proteomic_profiling_of_yeast_TFs.txt"

Slot "study.files":
\$`Global metabolite profiling of faah(-/-) mice`
Source Name Characteristics[NEWT:Organism LC] Term Source REF
1 Saghantelian_1 Mus musculus (Mouse) NEWT
2 Saghantelian_2 Mus musculus (Mouse) NEWT
3 Saghantelian_3 Mus musculus (Mouse) NEWT
4 Saghantelian_4 Mus musculus (Mouse) NEWT
5 Saghantelian_5 Mus musculus (Mouse) NEWT
6 Saghantelian_6 Mus musculus (Mouse) NEWT
7 Saghantelian_7 Mus musculus (Mouse) NEWT
8 Saghantelian_8 Mus musculus (Mouse) NEWT
9 Saghantelian_9 Mus musculus (Mouse) NEWT
10 Saghantelian_10 Mus musculus (Mouse) NEWT
11 Saghantelian_11 Mus musculus (Mouse) NEWT
12 Saghantelian_12 Mus musculus (Mouse) NEWT

	Term	Accession	Number	Characteristics[tissue]	Term	Source	REF
1			10090	spinal cord		MA	
2			10090	spinal cord		MA	
3			10090	spinal cord		MA	
4			10090	spinal cord		MA	
5			10090	spinal cord		MA	
6			10090	spinal cord		MA	
7			10090	spinal cord		MA	
8			10090	spinal cord		MA	
9			10090	spinal cord		MA	
10			10090	spinal cord		MA	
11			10090	spinal cord		MA	
12			10090	spinal cord		MA	

Term	Accession	Number	Protocol	REF	Sample	Name	Factor	Value[Genotype]
------	-----------	--------	----------	-----	--------	------	--------	-----------------

1	216 sample collection	K01	KO
2	216 sample collection	K02	KO
3	216 sample collection	K03	KO
4	216 sample collection	K04	KO
5	216 sample collection	K05	KO
6	216 sample collection	K06	KO
7	216 sample collection	WT1	WT
8	216 sample collection	WT2	WT
9	216 sample collection	WT3	WT
10	216 sample collection	WT4	WT
11	216 sample collection	WT5	WT
12	216 sample collection	WT6	WT

	Term	Source	REF	Term	Accession	Number
1			NA			NA
2			NA			NA
3			NA			NA
4			NA			NA
5			NA			NA
6			NA			NA
7			NA			NA
8			NA			NA
9			NA			NA
10			NA			NA
11			NA			NA
12			NA			NA

Slot "assay.fileNames":

V2

"a_metabolite.txt"

Slot "assay.fileNames.per.study":

\$`Global metabolite profiling of faah(-/-) mice`

\$`Global metabolite profiling of faah(-/-) mice`[[1]]

[1] "a_metabolite.txt"

Slot "assay.files":

\$a_metabolite.txt

	Sample	Name	Protocol	REF	Extract	Name	Protocol	REF	Labeled	Extract	Name
1		K01	extraction			K01	labeling				NA
2		K02	extraction			K02	labeling				NA
3		K03	extraction			K03	labeling				NA
4		K04	extraction			K04	labeling				NA
5		K05	extraction			K05	labeling				NA
6		K06	extraction			K06	labeling				NA
7		WT1	extraction			WT1	labeling				NA
8		WT2	extraction			WT2	labeling				NA
9		WT3	extraction			WT3	labeling				NA
10		WT4	extraction			WT4	labeling				NA
11		WT5	extraction			WT5	labeling				NA
12		WT6	extraction			WT6	labeling				NA

Label	Term	Source	REF	Term	Accession	Number	Protocol	REF
-------	------	--------	-----	------	-----------	--------	----------	-----

1	NA	NA	NA mass spectrometry
2	NA	NA	NA mass spectrometry
3	NA	NA	NA mass spectrometry
4	NA	NA	NA mass spectrometry
5	NA	NA	NA mass spectrometry
6	NA	NA	NA mass spectrometry
7	NA	NA	NA mass spectrometry
8	NA	NA	NA mass spectrometry
9	NA	NA	NA mass spectrometry
10	NA	NA	NA mass spectrometry
11	NA	NA	NA mass spectrometry
12	NA	NA	NA mass spectrometry

	Parameter	Value[instrument]	Term	Source	REF	Term	Accession Number
1	Agilent	1100 LC-MSD SL			NA		NA
2	Agilent	1100 LC-MSD SL			NA		NA
3	Agilent	1100 LC-MSD SL			NA		NA
4	Agilent	1100 LC-MSD SL			NA		NA
5	Agilent	1100 LC-MSD SL			NA		NA
6	Agilent	1100 LC-MSD SL			NA		NA
7	Agilent	1100 LC-MSD SL			NA		NA
8	Agilent	1100 LC-MSD SL			NA		NA
9	Agilent	1100 LC-MSD SL			NA		NA
10	Agilent	1100 LC-MSD SL			NA		NA
11	Agilent	1100 LC-MSD SL			NA		NA
12	Agilent	1100 LC-MSD SL			NA		NA

	Parameter	Value[ion source]	Term	Source	REF	Term	Accession Number
1	electrospray	ionization			MS		1000073
2	electrospray	ionization			MS		1000073
3	electrospray	ionization			MS		1000073
4	electrospray	ionization			MS		1000073
5	electrospray	ionization			MS		1000073
6	electrospray	ionization			MS		1000073
7	electrospray	ionization			MS		1000073
8	electrospray	ionization			MS		1000073
9	electrospray	ionization			MS		1000073
10	electrospray	ionization			MS		1000073
11	electrospray	ionization			MS		1000073
12	electrospray	ionization			MS		1000073

	Parameter	Value[detector]	Term	Source	REF	Term	Accession Number
1		NA			NA		NA
2		NA			NA		NA
3		NA			NA		NA
4		NA			NA		NA
5		NA			NA		NA
6		NA			NA		NA
7		NA			NA		NA
8		NA			NA		NA
9		NA			NA		NA
10		NA			NA		NA
11		NA			NA		NA
12		NA			NA		NA

	Parameter	Value[ionization mode]	Term	Source	REF	Term	Accession Number
1		positive mode			NA		NA
2		positive mode			NA		NA

3	positive mode	NA	NA
4	positive mode	NA	NA
5	positive mode	NA	NA
6	positive mode	NA	NA
7	positive mode	NA	NA
8	positive mode	NA	NA
9	positive mode	NA	NA
10	positive mode	NA	NA
11	positive mode	NA	NA
12	positive mode	NA	NA

	MS Assay Name	Raw Spectral Data File	Protocol	REF	Normalization Name
1	lc-ms-1	./cdf/KO/ko15.CDF		NA	NA
2	lc-ms-2	./cdf/KO/ko16.CDF		NA	NA
3	lc-ms-3	./cdf/KO/ko18.CDF		NA	NA
4	lc-ms-4	./cdf/KO/ko19.CDF		NA	NA
5	lc-ms-5	./cdf/KO/ko21.CDF		NA	NA
6	lc-ms-6	./cdf/KO/ko22.CDF		NA	NA
7	lc-ms-7	./cdf/WT/wt15.CDF		NA	NA
8	lc-ms-8	./cdf/WT/wt16.CDF		NA	NA
9	lc-ms-9	./cdf/WT/wt18.CDF		NA	NA
10	lc-ms-10	./cdf/WT/wt19.CDF		NA	NA
11	lc-ms-11	./cdf/WT/wt21.CDF		NA	NA
12	lc-ms-12	./cdf/WT/wt22.CDF		NA	NA

	Data Transformation Name	Derived Spectral Data File	Factor	Value[Genotype]
1	NA	faahkoDSDF.txt		KO
2	NA	faahkoDSDF.txt		KO
3	NA	faahkoDSDF.txt		KO
4	NA	faahkoDSDF.txt		KO
5	NA	faahkoDSDF.txt		KO
6	NA	faahkoDSDF.txt		KO
7	NA	faahkoDSDF.txt		WT
8	NA	faahkoDSDF.txt		WT
9	NA	faahkoDSDF.txt		WT
10	NA	faahkoDSDF.txt		WT
11	NA	faahkoDSDF.txt		WT
12	NA	faahkoDSDF.txt		WT

	Term Source	REF	Term	Accession Number
1		NA		NA
2		NA		NA
3		NA		NA
4		NA		NA
5		NA		NA
6		NA		NA
7		NA		NA
8		NA		NA
9		NA		NA
10		NA		NA
11		NA		NA
12		NA		NA

```
Slot "assay.files.per.study":
$`Global metabolite profiling of faah(-/-) mice`
$`Global metabolite profiling of faah(-/-) mice`[[1]]
```

	Sample Name	Protocol	REF	Extract Name	Protocol	REF	Labeled Extract Name
1	K01	extraction		K01	labeling		NA
2	K02	extraction		K02	labeling		NA
3	K03	extraction		K03	labeling		NA
4	K04	extraction		K04	labeling		NA
5	K05	extraction		K05	labeling		NA
6	K06	extraction		K06	labeling		NA
7	WT1	extraction		WT1	labeling		NA
8	WT2	extraction		WT2	labeling		NA
9	WT3	extraction		WT3	labeling		NA
10	WT4	extraction		WT4	labeling		NA
11	WT5	extraction		WT5	labeling		NA
12	WT6	extraction		WT6	labeling		NA

	Label	Term	Source	REF	Term	Accession Number	Protocol	REF
1	NA			NA			NA mass spectrometry	
2	NA			NA			NA mass spectrometry	
3	NA			NA			NA mass spectrometry	
4	NA			NA			NA mass spectrometry	
5	NA			NA			NA mass spectrometry	
6	NA			NA			NA mass spectrometry	
7	NA			NA			NA mass spectrometry	
8	NA			NA			NA mass spectrometry	
9	NA			NA			NA mass spectrometry	
10	NA			NA			NA mass spectrometry	
11	NA			NA			NA mass spectrometry	
12	NA			NA			NA mass spectrometry	

	Parameter	Value[instrument]	Term	Source	REF	Term	Accession Number
1	Agilent	1100 LC-MSD SL			NA		NA
2	Agilent	1100 LC-MSD SL			NA		NA
3	Agilent	1100 LC-MSD SL			NA		NA
4	Agilent	1100 LC-MSD SL			NA		NA
5	Agilent	1100 LC-MSD SL			NA		NA
6	Agilent	1100 LC-MSD SL			NA		NA
7	Agilent	1100 LC-MSD SL			NA		NA
8	Agilent	1100 LC-MSD SL			NA		NA
9	Agilent	1100 LC-MSD SL			NA		NA
10	Agilent	1100 LC-MSD SL			NA		NA
11	Agilent	1100 LC-MSD SL			NA		NA
12	Agilent	1100 LC-MSD SL			NA		NA

	Parameter	Value[ion source]	Term	Source	REF	Term	Accession Number
1	electrospray	ionization			MS		1000073
2	electrospray	ionization			MS		1000073
3	electrospray	ionization			MS		1000073
4	electrospray	ionization			MS		1000073
5	electrospray	ionization			MS		1000073
6	electrospray	ionization			MS		1000073
7	electrospray	ionization			MS		1000073
8	electrospray	ionization			MS		1000073
9	electrospray	ionization			MS		1000073
10	electrospray	ionization			MS		1000073
11	electrospray	ionization			MS		1000073
12	electrospray	ionization			MS		1000073

	Parameter	Value[detector]	Term	Source	REF	Term	Accession Number
1			NA		NA		NA

2	NA	NA	NA
3	NA	NA	NA
4	NA	NA	NA
5	NA	NA	NA
6	NA	NA	NA
7	NA	NA	NA
8	NA	NA	NA
9	NA	NA	NA
10	NA	NA	NA
11	NA	NA	NA
12	NA	NA	NA

	Parameter	Value[ionization mode]	Term	Source	REF	Term	Accession	Number
1		positive mode			NA			NA
2		positive mode			NA			NA
3		positive mode			NA			NA
4		positive mode			NA			NA
5		positive mode			NA			NA
6		positive mode			NA			NA
7		positive mode			NA			NA
8		positive mode			NA			NA
9		positive mode			NA			NA
10		positive mode			NA			NA
11		positive mode			NA			NA
12		positive mode			NA			NA

	MS Assay Name	Raw	Spectral Data File	Protocol	REF	Normalization Name
1	lc-ms-1		./cdf/KO/ko15.CDF		NA	NA
2	lc-ms-2		./cdf/KO/ko16.CDF		NA	NA
3	lc-ms-3		./cdf/KO/ko18.CDF		NA	NA
4	lc-ms-4		./cdf/KO/ko19.CDF		NA	NA
5	lc-ms-5		./cdf/KO/ko21.CDF		NA	NA
6	lc-ms-6		./cdf/KO/ko22.CDF		NA	NA
7	lc-ms-7		./cdf/WT/wt15.CDF		NA	NA
8	lc-ms-8		./cdf/WT/wt16.CDF		NA	NA
9	lc-ms-9		./cdf/WT/wt18.CDF		NA	NA
10	lc-ms-10		./cdf/WT/wt19.CDF		NA	NA
11	lc-ms-11		./cdf/WT/wt21.CDF		NA	NA
12	lc-ms-12		./cdf/WT/wt22.CDF		NA	NA

	Data Transformation Name	Derived Spectral Data File	Factor Value[Genotype]
1	NA	NA	KO
2	NA	NA	KO
3	NA	NA	KO
4	NA	NA	KO
5	NA	NA	KO
6	NA	NA	KO
7	NA	NA	WT
8	NA	NA	WT
9	NA	NA	WT
10	NA	NA	WT
11	NA	NA	WT
12	NA	NA	WT

	Term	Source	REF	Term	Accession	Number
1		NA		NA		
2		NA		NA		
3		NA		NA		

4	NA	NA
5	NA	NA
6	NA	NA
7	NA	NA
8	NA	NA
9	NA	NA
10	NA	NA
11	NA	NA
12	NA	NA

Slot "assay.names":

\$a_metabolite.txt

	MS Assay Name
1	lc-ms-1
2	lc-ms-2
3	lc-ms-3
4	lc-ms-4
5	lc-ms-5
6	lc-ms-6
7	lc-ms-7
8	lc-ms-8
9	lc-ms-9
10	lc-ms-10
11	lc-ms-11
12	lc-ms-12

Slot "assay.technology.types":

[1] "mass spectrometry"

Slot "assay.measurement.types":

[1] "metabolite profiling"

Slot "data.filenames":

\$a_metabolite.txt

	Raw Spectral Data File	Derived Spectral Data File
1	./cdf/K0/ko15.CDF	faahkoDSDF.txt
2	./cdf/K0/ko16.CDF	faahkoDSDF.txt
3	./cdf/K0/ko18.CDF	faahkoDSDF.txt
4	./cdf/K0/ko19.CDF	faahkoDSDF.txt
5	./cdf/K0/ko21.CDF	faahkoDSDF.txt
6	./cdf/K0/ko22.CDF	faahkoDSDF.txt
7	./cdf/WT/WT15.CDF	faahkoDSDF.txt
8	./cdf/WT/WT16.CDF	faahkoDSDF.txt
9	./cdf/WT/WT18.CDF	faahkoDSDF.txt
10	./cdf/WT/WT19.CDF	faahkoDSDF.txt
11	./cdf/WT/WT21.CDF	faahkoDSDF.txt
12	./cdf/WT/WT22.CDF	faahkoDSDF.txt

Slot "samples":

[1] "K01" "K02" "K03" "K04" "K05" "K06" "WT1" "WT2" "WT3" "WT4" "WT5" "WT6"

```
Slot "samples.per.study":
$`Global metabolite profiling of faah(-/-) mice`
[1] "K01" "K02" "K03" "K04" "K05" "K06" "WT1" "WT2" "WT3" "WT4" "WT5" "WT6"
```

```
Slot "samples.per.assay.filename":
$a_metabolite.txt
[1] "K01" "K02" "K03" "K04" "K05" "K06" "WT1" "WT2" "WT3" "WT4" "WT5" "WT6"
```

```
Slot "assay filenames.per.sample":
[[1]]
[[1]][[1]]
[1] "a_metabolite.txt"
```

```
[[2]]
[[2]][[1]]
[1] "a_metabolite.txt"
```

```
[[3]]
[[3]][[1]]
[1] "a_metabolite.txt"
```

```
[[4]]
[[4]][[1]]
[1] "a_metabolite.txt"
```

```
[[5]]
[[5]][[1]]
[1] "a_metabolite.txt"
```

```
[[6]]
[[6]][[1]]
[1] "a_metabolite.txt"
```

```
[[7]]
[[7]][[1]]
[1] "a_metabolite.txt"
```

```
[[8]]
[[8]][[1]]
[1] "a_metabolite.txt"
```

```
[[9]]
[[9]][[1]]
```



```
[1] "a_metabolite.txt"
```

```
[[10]]  
[[10]][[1]]  
[1] "a_metabolite.txt"
```

```
[[11]]  
[[11]][[1]]  
[1] "a_metabolite.txt"
```

```
[[12]]  
[[12]][[1]]  
[1] "a_metabolite.txt"
```

```
Slot "sample.to.rawdatafile":
```

```
[[1]]  
  Sample Name Raw Spectral Data File  
1      K01     ./cdf/K0/ko15.CDF  
2      K02     ./cdf/K0/ko16.CDF  
3      K03     ./cdf/K0/ko18.CDF  
4      K04     ./cdf/K0/ko19.CDF  
5      K05     ./cdf/K0/ko21.CDF  
6      K06     ./cdf/K0/ko22.CDF  
7      WT1     ./cdf/WT/wt15.CDF  
8      WT2     ./cdf/WT/wt16.CDF  
9      WT3     ./cdf/WT/wt18.CDF  
10     WT4     ./cdf/WT/wt19.CDF  
11     WT5     ./cdf/WT/wt21.CDF  
12     WT6     ./cdf/WT/wt22.CDF
```

```
Slot "sample.to.assayname":
```

```
[[1]]  
  Sample Name MS Assay Name  
1      K01     lc-ms-1  
2      K02     lc-ms-2  
3      K03     lc-ms-3  
4      K04     lc-ms-4  
5      K05     lc-ms-5  
6      K06     lc-ms-6  
7      WT1     lc-ms-7  
8      WT2     lc-ms-8  
9      WT3     lc-ms-9  
10     WT4     lc-ms-10  
11     WT5     lc-ms-11  
12     WT6     lc-ms-12
```

```
Slot "rawdatafile.to.sample":
```

```
[[1]]
```

```
Raw Spectral Data File Sample Name
1      ./cdf/KO/ko15.CDF      K01
2      ./cdf/KO/ko16.CDF      K02
3      ./cdf/KO/ko18.CDF      K03
4      ./cdf/KO/ko19.CDF      K04
5      ./cdf/KO/ko21.CDF      K05
6      ./cdf/KO/ko22.CDF      K06
7      ./cdf/WT/wt15.CDF      WT1
8      ./cdf/WT/wt16.CDF      WT2
9      ./cdf/WT/wt18.CDF      WT3
10     ./cdf/WT/wt19.CDF      WT4
11     ./cdf/WT/wt21.CDF      WT5
12     ./cdf/WT/wt22.CDF      WT6
```

```
Slot "assayname.to.sample":
```

```
[[1]]
```

```
MS Assay Name Sample Name
1      lc-ms-1      K01
2      lc-ms-10     WT4
3      lc-ms-11     WT5
4      lc-ms-12     WT6
5      lc-ms-2      K02
6      lc-ms-3      K03
7      lc-ms-4      K04
8      lc-ms-5      K05
9      lc-ms-6      K06
10     lc-ms-7      WT1
11     lc-ms-8      WT2
12     lc-ms-9      WT3
```

```
Slot "factors":
```

```
[[1]]
```

```
[[1]]$`Factor Value[Genotype]`
```

```
[1] KO KO KO KO KO KO WT WT WT WT WT WT
```

```
Levels: KO WT
```

```
Slot "treatments":
```

```
$`Factor Value[Genotype]`
```

```
[1] KO WT
```

```
Levels: KO WT
```

```
Slot "groups":
```

```
[[1]]
```

```
[[1]][[1]]
```

```
[1] "K01" "K02" "K03" "K04" "K05" "K06"
```

```
[[1]][[2]]
```

```
[1] "WT1" "WT2" "WT3" "WT4" "WT5" "WT6"
```

```

Slot "assay.tabs":
[[1]]
An object of class "MSAssayTab"
Slot "path":
[1] "/Library/Frameworks/R.framework/Versions/4.1/Resources/library/faahKO"

Slot "study.filename":
[1] "s-Proteomic_profiling_of_yeast_TFs.txt"

Slot "study.identifier":
[1] "Global metabolite profiling of faah(-/-) mice"

Slot "assay.filename":
[1] "a_metabolite.txt"

Slot "assay.file":
  Sample Name Protocol REF Extract Name Protocol REF Labeled Extract Name
1      KO1    extraction      KO1    labeling              NA
2      KO2    extraction      KO2    labeling              NA
3      KO3    extraction      KO3    labeling              NA
4      KO4    extraction      KO4    labeling              NA
5      KO5    extraction      KO5    labeling              NA
6      KO6    extraction      KO6    labeling              NA
7      WT1    extraction      WT1    labeling              NA
8      WT2    extraction      WT2    labeling              NA
9      WT3    extraction      WT3    labeling              NA
10     WT4    extraction      WT4    labeling              NA
11     WT5    extraction      WT5    labeling              NA
12     WT6    extraction      WT6    labeling              NA
  Label Term Source REF Term Accession Number      Protocol REF
1     NA              NA              NA mass spectrometry
2     NA              NA              NA mass spectrometry
3     NA              NA              NA mass spectrometry
4     NA              NA              NA mass spectrometry
5     NA              NA              NA mass spectrometry
6     NA              NA              NA mass spectrometry
7     NA              NA              NA mass spectrometry
8     NA              NA              NA mass spectrometry
9     NA              NA              NA mass spectrometry
10    NA              NA              NA mass spectrometry
11    NA              NA              NA mass spectrometry
12    NA              NA              NA mass spectrometry
  Parameter Value[instrument] Term Source REF Term Accession Number
1      Agilent 1100 LC-MSD SL      NA              NA
2      Agilent 1100 LC-MSD SL      NA              NA
3      Agilent 1100 LC-MSD SL      NA              NA
4      Agilent 1100 LC-MSD SL      NA              NA
5      Agilent 1100 LC-MSD SL      NA              NA
6      Agilent 1100 LC-MSD SL      NA              NA
7      Agilent 1100 LC-MSD SL      NA              NA
8      Agilent 1100 LC-MSD SL      NA              NA

```

9	Agilent 1100 LC-MSD SL		NA	NA
10	Agilent 1100 LC-MSD SL		NA	NA
11	Agilent 1100 LC-MSD SL		NA	NA
12	Agilent 1100 LC-MSD SL		NA	NA
	Parameter Value	[ion source]	Term Source REF	Term Accession Number
1	electrospray	ionization	MS	1000073
2	electrospray	ionization	MS	1000073
3	electrospray	ionization	MS	1000073
4	electrospray	ionization	MS	1000073
5	electrospray	ionization	MS	1000073
6	electrospray	ionization	MS	1000073
7	electrospray	ionization	MS	1000073
8	electrospray	ionization	MS	1000073
9	electrospray	ionization	MS	1000073
10	electrospray	ionization	MS	1000073
11	electrospray	ionization	MS	1000073
12	electrospray	ionization	MS	1000073
	Parameter Value	[detector]	Term Source REF	Term Accession Number
1		NA	NA	NA
2		NA	NA	NA
3		NA	NA	NA
4		NA	NA	NA
5		NA	NA	NA
6		NA	NA	NA
7		NA	NA	NA
8		NA	NA	NA
9		NA	NA	NA
10		NA	NA	NA
11		NA	NA	NA
12		NA	NA	NA
	Parameter Value	[ionization mode]	Term Source REF	Term Accession Number
1		positive mode	NA	NA
2		positive mode	NA	NA
3		positive mode	NA	NA
4		positive mode	NA	NA
5		positive mode	NA	NA
6		positive mode	NA	NA
7		positive mode	NA	NA
8		positive mode	NA	NA
9		positive mode	NA	NA
10		positive mode	NA	NA
11		positive mode	NA	NA
12		positive mode	NA	NA
	MS Assay Name	Raw Spectral Data File	Protocol REF	Normalization Name
1	lc-ms-1	./cdf/KO/ko15.CDF	NA	NA
2	lc-ms-2	./cdf/KO/ko16.CDF	NA	NA
3	lc-ms-3	./cdf/KO/ko18.CDF	NA	NA
4	lc-ms-4	./cdf/KO/ko19.CDF	NA	NA
5	lc-ms-5	./cdf/KO/ko21.CDF	NA	NA
6	lc-ms-6	./cdf/KO/ko22.CDF	NA	NA
7	lc-ms-7	./cdf/WT/wt15.CDF	NA	NA
8	lc-ms-8	./cdf/WT/wt16.CDF	NA	NA
9	lc-ms-9	./cdf/WT/wt18.CDF	NA	NA
10	lc-ms-10	./cdf/WT/wt19.CDF	NA	NA

11	lc-ms-11	./cdf/WT/wt21.CDF	NA	NA
12	lc-ms-12	./cdf/WT/wt22.CDF	NA	NA

	Data Transformation	Name	Derived Spectral Data File	Factor Value	[Genotype]
1		NA	NA		KO
2		NA	NA		KO
3		NA	NA		KO
4		NA	NA		KO
5		NA	NA		KO
6		NA	NA		KO
7		NA	NA		WT
8		NA	NA		WT
9		NA	NA		WT
10		NA	NA		WT
11		NA	NA		WT
12		NA	NA		WT

	Term	Source REF	Term	Accession Number
1		NA		NA
2		NA		NA
3		NA		NA
4		NA		NA
5		NA		NA
6		NA		NA
7		NA		NA
8		NA		NA
9		NA		NA
10		NA		NA
11		NA		NA
12		NA		NA

Slot "assay.technology.type":
[1] "mass spectrometry"

Slot "assay.measurement.type":
[1] "metabolite profiling"

Slot "assay.names":

	MS Assay Name
1	lc-ms-1
2	lc-ms-2
3	lc-ms-3
4	lc-ms-4
5	lc-ms-5
6	lc-ms-6
7	lc-ms-7
8	lc-ms-8
9	lc-ms-9
10	lc-ms-10
11	lc-ms-11
12	lc-ms-12

Slot "data_filenames":

	Raw Spectral Data File	Derived Spectral Data File
1	./cdf/KO/ko15.CDF	NA
2	./cdf/KO/ko16.CDF	NA

3	./cdf/KO/ko18.CDF	NA
4	./cdf/KO/ko19.CDF	NA
5	./cdf/KO/ko21.CDF	NA
6	./cdf/KO/ko22.CDF	NA
7	./cdf/WT/wt15.CDF	NA
8	./cdf/WT/wt16.CDF	NA
9	./cdf/WT/wt18.CDF	NA
10	./cdf/WT/wt19.CDF	NA
11	./cdf/WT/wt21.CDF	NA
12	./cdf/WT/wt22.CDF	NA

For an example for a real use case, please refer to <https://github.com/sneumann/mtbls2/>.

Writing ISA-Tab datasets

The Risa package offers functions to write the whole ISA-Tab dataset or part of it back to disk. These functions are `write.ISAtab`, `write.investigation.file`, `write.study.file`, `write.assay.file`.

So, after updating the assay file as indicated above, we can save it back to disk, using the following command:

```
> temp = tempdir()
> write.ISAtab(faahkoISA, temp)
> #write.assay.file(faahkoISA, assay.filename, temp)
```

Session Info

```
> toLatex(sessionInfo())
```

- R version 4.1.0 RC (2021-05-10 r80283), x86_64-apple-darwin17.0
- Locale: C/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
- Running under: macOS Mojave 10.14.6
- Matrix products: default
- BLAS:

/Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.dylib
- LAPACK:

/Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
- Base packages: base, datasets, grDevices, graphics, methods, parallel, stats, stats4, utils
- Other packages: Biobase 2.52.0, BiocGenerics 0.38.0, BiocParallel 1.26.0, MSnbase 2.18.0, ProtGenerics 1.24.0, Rcpp 1.0.6, Risa 1.34.0, S4Vectors 0.30.0, affy 1.70.0, biocViews 1.60.0, faahKO 1.31.0, mzR 2.26.0, xcms 3.14.0
- Loaded via a namespace (and not attached): BiocManager 1.30.15, DBI 1.1.1, DEoptimR 1.0-8, DelayedArray 0.18.0, GenomeInfoDb 1.28.0, GenomeInfoDbData 1.2.6, GenomicRanges 1.44.0, IRanges 2.26.0, MALDIquant 1.19.3, MASS 7.3-54, MassSpecWavelet 1.58.0, Matrix 1.3-3, MatrixGenerics 1.4.0, MsCoreUtils 1.4.0, R6 2.5.0, RANN 2.6.1, RBGL 1.68.0, RColorBrewer 1.1-2, RCurl 1.98-1.3, RUnit 0.4.32, SummarizedExperiment 1.22.0, XML 3.99-0.6, XVector 0.32.0, affyio 1.62.0, assertthat 0.2.1, bitops 1.0-7, clue 0.3-59, cluster 2.1.2, codetools 0.2-18, colorspace 2.0-1, compiler 4.1.0, crayon 1.4.1, digest 0.6.27, doParallel 1.0.16, dplyr 1.0.6, ellipsis 0.3.2,

fansi 0.4.2, foreach 1.5.1, generics 0.1.0, ggplot2 3.3.3, glue 1.4.2, graph 1.70.0, grid 4.1.0, gtable 0.3.0, impute 1.66.0, iterators 1.0.13, lattice 0.20-44, lifecycle 1.0.0, limma 3.48.0, magrittr 2.0.1, matrixStats 0.58.0, munsell 0.5.0, mzID 1.30.0, ncd4 1.17, pcaMethods 1.84.0, pillar 1.6.1, pkgconfig 2.0.3, plyr 1.8.6, preprocessCore 1.54.0, purrr 0.3.4, rlang 0.4.11, robustbase 0.93-7, scales 1.1.1, tibble 3.1.2, tidyselect 1.1.1, tools 4.1.0, utf8 1.2.1, vctrs 0.3.8, vsn 3.60.0, zlibbioc 1.38.0

Further information

For further information about the ISA software infrastructure, please visit our website <http://isa-tools.org>.