

**MAGETAB Workshop**  
**24-25 January 2008**  
**CBIT (NCICB)**  
**Rockville, USA**  
**Report Author: Helen Parkinson**  
**Version 1.0**  
**5 February, 2008**

Meeting Attendees:

Xiaopeng Bian, caArray  
Catherine A. Ball, Stanford Microarray Database  
Alvis Brazma, EBI  
Ian Fore, NCICB  
B. Gendelman, 5AM Solutions  
Mervi Heiskanen, NCICB (Host)  
Ari Kahn, TCGA, DCC  
Ron Keene, NCI/NIH  
Juli Klemm, NCICB  
Ted Liebfeld, GenePattern  
Junmin Liu, CBIL, University of Pennsylvania  
Subhashree Madhavan, NCICB  
Don Maier, Stanford Microarray Database  
Steve Matyas, 5AM Solutions  
Michael Miller, Rosetta BioSoftware  
Leonie Misquitta, NCI/NIH  
Sandy Orlow, NCI/CIT  
Helen Parkinson, EBI  
Paul Spellman, LBL  
Carl Schaefer, NCICB  
Christian J. Stoeckert Jr., University of Pennsylvania  
Don Swan, caArray  
Eric Tavela, 5AM Solutions  
Joe White, Dana Farber Cancer Center  
Liming Yang, NCICB

Agenda: .....	2
Discussion: .....	2
MAGETAB specification Decisions .....	2
1. Blank Lines .....	2
2. Quotes in Column Headers .....	3
3. Ordering and Cardinalities of SDRF Headings .....	3
4. Namespace format .....	4
5. Hybridizations versus Assays .....	5
6. Protocol REF prefixes .....	5
7. IDF column order and cardinalities.....	5
8. SDRF/IDF Dependence .....	6
9. Restriction of # <comment> line to SDRF .....	6
10. Header Column Order and Cardinalities in the ADF Header, Main Table, and Extended Table .....	7
10.1.1 ADF .....	7
10.1.2. Main and extended tables.....	7
11. Order and Delimiter for ADF Sections .....	7
12. Backward compatibility.....	7
15. Unambiguous conventions – Identifying objects in MAGE-TAB .....	8
16. Connection of Factor Value to Experimental Factors .....	9
18. Use of manufacturer’s name for Array Design Ref .....	9

19. Empty fields.....	9
20. Recommended usage for quantitation types .....	9
21. Term Accession Number" headers in the IDF .....	10
22. Term Accession Number" headers in the ADF .....	10
24. Process for MAGETAB specification e.g. HUPO .....	10
25. Semi colon delimited fields.....	11
27. Documenting technology/type dependencies .....	11
28. ParameterValue[] is uncontrolled.....	11
29. Dealing with Characteristics Columns of the same type .....	12
30. Cross over designs.....	12
32. Three fields to add when specifying an ontology term.....	12
35. Differentially dimensions arrays in same SDRF file .....	13
36. Packaging SDRF/IDF and data files .....	13
37. Annotation Examples in public domain.....	13
38. Improving MAGETAB take up.....	14

## Agenda:

### Discussion:

Don Maier prepared a document detailing multiple issues with the MAGETAB1.0 specification; this was circulated for comment prior to the meeting and was open for community comment. The decisions below refer to issues raised in that document, and these are summarized here for clarity with the problem, solution (both described by Don Maier unless otherwise specified) followed meeting decisions and action items. Each issue was presented at the workshop, a discussion was held and a vote was made on one or more proposals. Some additional issues were raised during the workshop and these are appended at the end of the section below.

## MAGETAB specification Decisions

### 1. Blank Lines

**Problem:** The specification says nothing about the acceptability of blank lines or lines containing only tabs.

#### Proposed Solution:

A completely blank line or row (zero or more space characters) can occur anywhere in any MAGE-TAB file. Within a table that defines a number  $n$  of vertically oriented columns (as in the SDRF or the ADF main or extended table), it is also permissible to have a line containing exactly  $(n - 1)$  tabs. Within the IDF, which may contain any number of data cells on one line, a line containing any number of tabs is permitted. Within the ADF header, which has exactly one "Tag" and one "Value" on each data line, a line containing exactly one tab is permitted. These rules maintain a uniform tab structure within tables, except for the possible presence of entirely blank lines.

**Conclusion:** The proposal above was accepted. Blank rows or lines are permitted in MAGETAB IDF, SDRF, and ADF files.

**Action Item:** The 1.1 specification will be amended to state this explicitly.

**Person Responsible:** Don Maier

## 2. Quotes in Column Headers

**Problem:** Version 1.0 of the specification allows that not only data entries, but also column headers may be quoted. However, column headers, unlike data, are keywords that must be lexically recognizable. Something needs to be said in order to ensure this in a straightforward way

**Proposed Solution:** Quoted column headers are identical to non-quoted column headers except for being enclosed in double quotes. This means that no characters other than spaces are permitted between multiple keywords that comprise a header. Also, the opening quotation mark must immediately precede the first character of the first keyword and the closing quotation mark must immediately follow the last character of the last keyword, with no intervening characters.

**Conclusion:** Proposal is accepted as above. There was general agreement that the parser should strip leading and trailing spaces.

**Action Item:** Don Maier will check if there are issues with spaces in Antler that will require a new discussion for future versions

**Person Responsible:** Don Maier

## 3. Ordering and Cardinalities of SDRF Headings

**Problem:** Clarification is needed of what is a required column in a SDRF document. Needed to develop the MAGETAB-OM and also to make the specification clearer

**Proposed Solution:**

Element column headers in the SDRF, except for Protocol REF, must occur in the following order and with the following cardinalities. This is a total ordering. No column (other than Protocol REF) with number  $m$  may occur to the right of a column with number  $n > m$ . An element or attribute that permits a cardinality of 0 may be omitted. An element or attribute that does not permit a cardinality of 0 is required. Currently, all *attributes* may be omitted (are optional).

The attributes of an element or of another attribute must follow the attributed element or attribute without any intervening, non-attributed element or attribute. When an element or attribute has more than one attribute, there is no ordering defined for that set, except for:

- 1) Factor Value must occur after all element nodes and the attributes of those element nodes, as specified in the list of elements immediately above. This rule replaces whatever item 10 in "Notes on Table 7", p. 36 of Version 1.1 might say about the ordering of Factor Values.
- 2) Comment must immediately follow either the element or attribute node for which it is a Comment, or another such Comment. This permits an unambiguous association of a Comment with the element or attribute for on which it comments.
- 3) Term Source REF must immediately follow the ontology term for which it provides the source reference. This permits an unambiguous association of the Term Source REF to the ontology term.

In particular, in a group of attributes for one other element or attribute, there is no requirement to group together all columns of a given type that can occur more than once.

*All attributes are optional \* except Technology Type when used with Assay \* check if this is correct HP*

**Conclusion:** Cardinalities are described in the table below after discussion and were decided on after discussion. These are different from those proposed in Don Maier's discussion document.

Element Nodes and Factor Values	Cardinality	MAGETAB Version
---------------------------------	-------------	-----------------

Source Name	0..1	1.0
Sample Name	0..*	1.0
Extract Name	0..1	1.0
Labeled Extract Name	0..1	1.0
Hybridization Name	0..1	1.0
Assay Name	0..*	1.1 (see Item 5 below)
Scan Name	0..*	1.0
Image Name	0..*	1.0
ArrayData File	0..*	1.0
Derived Array Data File	0..*	1.0
Array Data Matrix File	0..*	1.0
Normalization Name	0..*	1.0
Derived Array Data Matrix File	0..*	1.0
Factor Value	0..*	1.0
Protocol REF	0..*	1.0
<b>Attributes – all are optional</b>	<b>Cardinality</b>	
Characteristics	*	1.0
Provider	*	1.0
Material Type	0..*	1.0
Label	0..*	1.0
Array Design File	0..1	1.0
Array Design REF	0..1	1.0
Technology Type	0..1	1.1 (see item 5 below, required if Assay is used)
Performer	*	1.0
Date	0..1	1.0
Parameter Value	*	1.0
Unit	0..1	1.0
Description	*	1.0
Term Source REF	0..1	1.0
Term Accession Number	0..1	1.0
Comment	*	1.0
		1.0

*\* Tim has pointed out that in Stanford a year ago 2 extracts Chip experiments were discussed and two extracts were proposed to deal with that. In light of that do we want to make extract 1. Two samples could be used instead*

**Action Item:** Add to MAGETAB 1.1 specification

**Person Responsible:** Don Maier

#### 4. Namespace format

**Problem:** Section 3.1.3, "Identifying objects in MAGE-TAB", pp. 25-26 of Version 1.1 suggests this format for namespaces:

<authority>:[<namespace>]:<object>[:<revision>]

Other discussions have indicated that this format is not required and that *any* format should be accepted. However, some restrictions are needed because, for

example, allowing white space of any kind or quotation marks would make it impossible to find an unambiguous grammar.

**Proposed Solution:** (Proposal 1):

A namespace can be any string that does not contain either white space or double quotation marks. URI that escape white space are permitted.

**Conclusion:** Proposal 1 is accepted. There was a suggestion that an optional source column for such namespaces could be added to a future version to allow global sources such as databases to be specified

**Action Item:** Add to MAGETAB 1.1 specification

**Person Responsible:** Don Maier

## 5. Hybridizations versus Assays

**Problem:** The “Notes on SDRF column headings” in Section 3.3.6, p. 36 of Version 1.1 of the specification states that Assay Name can substitute for Hybridization Name. It doesn’t say whether both Assay Name and Hybridization can appear in the same SDRF. See also the Section on “Ordering and Cardinalities of SDRF Headers”.

**Proposed Solution:** An SDRF may contain Hybridization Name columns or Assay columns, but not both. An SDRF with a Hybridization Name cannot contain a Technology Type, which is an attribute of Assay Name only. Similarly, an SDRF with an Assay Name cannot contain either an Array Design File or an Array Design REF column; these latter are attributes of Hybridization Name only. Assay is allowed in version 1.1. documents to support non array based technologies. Technology Type is required when Assay is used. Hybridization can also have technology type in v1.1. Both Assays and Hybs both can have Array Design files. Technology Type column has an optional Term Source REF.

**Action Item:** Add to MAGETAB 1.1 specification

**Person Responsible:** Don Maier

## 6. Protocol REF prefixes

**Problem:** This notion of a “free-text prefix” is added for Protocol REF headers in Section 3.3.6 under “Notes on SDRF column headings”, third bullet on p. 38 of Version 1.1. We need a definition of “free text” that makes it possible to parse the language with finite lookahead.

**Proposed Solution:** MAGE-TAB already has two ways to do what Protocol REF prefixes do. First, Comment columns can be used to qualify any column, including Protocol REF. Second, if we need to permit qualification of Protocol REF columns themselves, MAGE-TAB already has a canonical way of doing this – namely, the use of square brackets, which pose no language processing issues. Instead of introducing a new and problematic syntax to perform the same or very similar function, we should stick with square bracket syntax:

Protocol REF[<protocol type>]

**Conclusion:** Free text will not be allowed to define protocol types. ‘Protocol REF’ is the only permitted label for the column.

***\*Is this a correct interpretation of discussion***

**Action Item:** Add to MAGETAB 1.1 specification

**Person Responsible:** Don Maier

## 7. IDF column order and cardinalities

**Problem:** The specification is mute on the (vertical) order and allowed number of each type of (horizontally oriented) column in the IDF.

**Proposed Solution:** Every IDF column is optional. Each type of column may occur any number of times – except for the Date of Experiment, Public Release Date, and Experiment Description. (The specification already states that at most one of each of these column types may appear.

Columns may occur in any order. Note that this makes a completely empty file a valid IDF file.

**Conclusion:** Accepted as proposed above

**Action Item:** Add to MAGETAB 1.1 specification

**Person Responsible:** Don Maier

## 8. SDRF/IDF Dependence

**Problem:** The specification is not clear on whether the file references in various MAGE-TAB fields must be valid, or what “valid” means – that is, how they should be resolved. For example, must they be full path names in a local file system? Consider, for example, the Array Data File, Array Data Matrix File, and others in the SDRF; or the IDF’s reference to an SDRF File. In short, we need to specify how file names are to be resolved.

Also, there are a number of references in the SDRF to information defined in the IDF – for example a Protocol REF reference in the SDRF to a Protocol Name in the IDF. It is not clear whether or not it is an error to leave these references unresolved. Of course, checking these references requires that both IDF and SDRF files can be found.

**Proposed Solution:** Each file referenced in the IDF, SDRF and ADF, including the SDRF File in the IDF, must resolve to an existing file.

Considering the coordinating importance of the IDF’s reference to an SDRF, and that these are separate files that can and will be transported around the network, it seems that we must require that any file reference be relative to the file that contains that reference; and further, that all referring files be present in the same directory. This implies that all file references take this form:

file://sample.sdrf

It is an error for information in these SDRF columns not to have the corresponding information in the IDF.

**Conclusion:** IDF/SDRF should always be present together, data files and ADF could be referenced by a URL or e.g. FTP, this will allow cases where data/ADF is accessed by a password, or very large to be supported. file:// is not required at the start of the file name.

**Action Item:** Add to MAGETAB 1.1 specification

**Person Responsible:** Don Maier

*\* did we make a decision on case sensitivity and suffixes? My recollection is that file names are not case sensitive and that .txt suffixes are required as .idf and .sdrf have no meaning outside the MAGETAB specification. HP*

## 9. Restriction of # <comment> line to SDRF

**Problem:** The ability to have a comment line beginning with a ‘#’ is defined for the SDRF (only) in Version 1.1. A user would be justifiably surprised to find a line comment in an IDF or ADF file rejected. The restriction of line comments to the SDRF is unnecessary and confusing.

**Proposed Solution:** Line comments are permitted anywhere in all MAGE-TAB files.

**Conclusion:** Agreed as proposed, line comments are permitted anywhere in all MAGETAB files

**Action Item:** Add to MAGETAB 1.1 specification

**Person Responsible:** Don Maier

## 10. Header Column Order and Cardinalities in the ADF Header, Main Table, and Extended Table

### 10.1.1 ADF

**Problem:** For the ADF header, the specification is mute on the (vertical) order and allowed number of each type the (horizontally oriented) columns. Nor does the specification say how many "Value" columns may appear, or whether empty values are permitted.

#### 10.1.2. Main and extended tables

**Problem:** Each type of column is optional and can occur at most once. (This makes the ADF main table optional.) Columns may appear in any order, except that "associated attributes" must immediately follow the "object" that they qualify.

**Proposed Solution: 10.1.1** The ADF header comprises exactly one "Tag" column and exactly one "Value" column. There is no header row for the ADF header containing the entries "Tag" and "Value". The "Value" cell for any (horizontally oriented) row may be empty

**Proposed Solution 10.1.2 .)** Columns may appear in any order, except that "associated attributes" must immediately follow the "object" that they qualify.

**Conclusion: 10.1.1 -** Accepted as proposed. 10.1.2 accepted as proposed, with the proviso that any column that can be made unique using [] e.g. characteristics can occur multiple times egg. Characteristics[OrganismPart]

**Action Item:** Add 10.1.1. and 10.1.2 to the MAGETAB 1.1 specification

**Person Responsible:** Don Maier

## 11. Order and Delimiter for ADF Sections

**Problem:** The ADF is divided into as many as three sections. The specification does not define either the order for these sections or the syntax for their division. Apparently, the division relies on the recognition of the header rows for the main and extended table, distinguishing these, and distinguishing both from any row in the ADF header. This is a problematic syntax design because it relies on the uniqueness of keywords that can occur at the start of each of the three parts.

**Proposed Solution:** If included, the ADF header must precede the main table. If included, the extended table cannot appear without a main table, which it must follow.

A section delimiter, consisting of a form feed ('\f'), must be used to terminate each section but the last. A section delimiter is allowed, but not required for the last section.

An alternative to the form feed would be a line comments (introduced by '#') that starts with a keyword, for example:

```
# end
```

**Conclusion:** Headers, main or mapping –will be used as section delimiters, these will be treated as case insensitive. See also Item 1, blank lines are permitted and can be used to enhance readability, though have no significance in delimiting sections.

**Action Item:** Add to MAGETAB 1.1 specification

**Person Responsible:** Don Maier

## 12. Backward compatibility

**Problem:** As Version 1.1 stands in the current specification; it is not backward compatible with Version 1.0. Some proposals offered in this document (such as section delimiters in the ADF) introduce further incompatibilities. For example,

SDRF files that violate the Version 1.1 column ordering constraints may be valid for Version 1.0.

**Proposed Solution:** Accept lack of backwards compatibility. To fix Version 1.0 well, we cannot afford to accommodate a requirement of backward compatibility. Since we don't yet have substantial use of MAGE-TAB, the penalty for this is small.

**Conclusion:** Version will be added to the IDF from v1.1 on. If there is no version specified version 1.0 is assumed. If a version is specified then both SDRF and IDF must be the same version.

**Action Item:** Add version as a required field in IDF to MAGETAB specification 1.1

**Person Responsible:** Don Maier

### 13. Version column in IDF, producer column in IDF

**Problem:** Absent from 1.0 specification, proposed by Michael Miller

**Proposed Solution:** Add version to IDF, and producer to IDF, where producer is software or organization

**Conclusion:** Version information will be added to v1.1 documents onward (see item 12 above). Addition of producer to the IDF was deferred to a future version as it was not deemed to be urgent by the majority and because person is specified in the IDF not the MAGE contact (that had children person or organization) thus requiring the specification to be changed further. Use comment fields if this needs to be specified for 1.0 and 1.1

**Action Item:** N/A

**Person Responsible:** N/A

### 14. Processing Order for multiple SDRF fields

**Problem:** When there are multiple SDRF files, how does a parser know the order to process them? Michael Miller

**Proposed Solution:** should be ordered with files that reference other files appearing earlier in the list. Otherwise, it would require reading in all the files and figuring out the columns to link the files. This would make the parsing code less complex.

**Conclusion:** No decision was reached for 1.1; this was deferred for discussion on a later version.

**Action Item:** N/A

**Person Responsible:** N/A

### 15. Unambiguous conventions – Identifying objects in MAGE-TAB

**Problem:**

It's very important to Rosetta use of MAGE-TAB that we can identify the same source (and other objects) in files in different formats like SEND, CDISC, MAGE-ML, FuGE, etc. Michael Miller

Requirement to add references to e.g. PowerPoint presentations, posters and pdfs and other supplementary file types to MAGETAB files, Juli Klemm

**Proposed Solution:** Several proposals were made during the discussion including use of LSID, comments etc

**Conclusion:** No decision was made for 1.1; these types of documents can currently be reference by use of comments in local applications.

**Action Item:** Add a comment to 1.1 specifications making clear that comments can be used in this way

**Person Responsible:** Don Maier



## 16. Connection of Factor Value to Experimental Factors

**Problem:** There's a reference that "any column ... also listed in a separate "Factor Value []" column". There's no way to connect what column this was that's repeated as the "Factor Value []" column plus the producer of the document may not have bothered to add the first column since the information is also in the "Factor Value []" column. So I'm not sure this should be mentioned.

**Proposed Solution:** N/A

**Conclusion:** There is no good way to map between sample or protocol parameter columns and factor values where these represent the same information labelled with a different header. After discussion it was decided to retain the current situation where Factor Value columns must be specified separately

**Action Item:** Add a clarification to the 1.0/1.1 specification to make it clear that sample properties and parameters can both constitute factor values, that there is no requirement for these to be internally consistent (e.g. that any protocol parameters specified as Factor Values are also provided as protocol parameters) and that there is no requirement or ability in MAGETAB 1.0/1.1 to link these columns.

**Person Responsible:** Don Maier

## 17. Clarifying the unique "edge" role of "Protocol Ref"

**Problem:** If Protocol ref is the only edge column this should be clear in the documentation

**Proposed Solution:** Clarify 1.0/1.1 specification

**Conclusion:** Agreed

**Action Item:** Add to 1.0/1.1 specification

**Person Responsible:** Don Maier

## 18. Use of manufacturer's name for Array Design Ref

**Problem:** Public repositories and others do not use the Manufacturer's names to identify common array designs

**Proposed Solution:** This is not a MAGETAB issue.

**Conclusion:** Not in MAGETAB scope

**Action Item:** N/A

**Person Responsible:** N/A

## 19. Empty fields

**Problem:** Empty fields should just be empty, that is the usual practice.

**Proposed Solution:** As above

**Conclusion:** Empty fields are interpreted as empty, this includes "". Null does not equal empty

**Action Item:** Add clarification to MAGETAB 1.0/1.1 spec

**Person Responsible:** Don Maier

## 20. Recommended usage for quantitation types

**Problem:** Lack of standardization of quantitation types

**Proposed Solution:** Should standard MAGETAB quantitation types be used?

Michael Miller

**Conclusion:** Outside MAGETAB scope, no universal source of QT

**Action Item:** N/A  
**Person Responsible:** N/A

### **21. Term Accession Number" headers in the IDF**

**Problem:** When we discussed adding the "Term Accession Number" column to SDRF we neglected the IDF and the ADF (see next item). There should be one for each "Term Source REF" header. As for all headers in the IDF, these are optional and can appear in any place and order in the IDF. Tim Rayner

**Proposed Solution:** Add the following to IDF  
Experimental Design Term Accession Number  
Experimental Factor Term Accession Number  
Person Roles Term Accession Number  
Quality Control Term Accession Number  
Normalization Term Accession Number  
Publication Status Term Accession Number  
Protocol Term Accession Number

**Conclusion:** Accepted as proposed

**Action Item:** Add to MAGETAB 1.1 specification

**Person Responsible:** Don Maier

### **22. Term Accession Number" headers in the ADF**

**Problem:** For uniformity with 21 these should be added

**Proposed Solution:** Add:

Technology Type Term Accession Number  
Surface Type Term Accession Number  
Substrate Type Term Accession Number  
Sequence Polymer Type Term Accession Number

**Conclusion:** Accepted as proposed

**Action Item:** Add to MAGETAB 1.1 specification

**Person Responsible:** Don Maier

### **23. Mapping MAGETAB to MAGE**

**Problem:** Mage-tab to mage-ml mapping-great start. It would also be nice to have a mage template document (for developers) where the column headers surrounded by something like '{ }' are the replaceable attribute values. Rosetta has used this to communicate with customers how their data can go in a mage document. Michael Miller

**Proposed Solution:** Generate Documentation

**Conclusion:** Agreed, docs are useful

**Action Item:** Documentation will be generated

**Person Responsible:** Junmin Liu, Michael Miller, and EBI

### **24. Process for MAGETAB specification e.g. HUPO**

**Problem:** Do we want a formal process e.g. HUPO like for MAGETAB

**Proposed Solution:** N/A

**Conclusion:** No formal process was agreed for MAGETAB. There was a note that we would like to follow through the ISATAB process (Whatever that will be)

**Action Item:** Provide feedback on ISATAB

**Person Responsible:** N/A

## 25. Semi colon delimited fields

**Problem:** Semicolon-separated fields into a tabular format (e.g. Person Role) may result in formatting issues: Problem when sources are two different CVs/Ontology. Phil Jones.

**Proposed Solution:** introduce and document restrictions for the ontology resources to use

**Conclusion:** All role terms must be from one ontology only, will be documented as a known restriction for 1.0 and 1.1. Semi colons are restricted to the following fields: IDF person role, protocol parameters, ADF Map2Reporters, SDRF provider and performer. (5 in total)

**Action Item:** Document restriction in MAGETAB 1.0/1.1 specification

**Person Responsible:** Don Maier

## 26. Characteristic and Parameter column labels not specified as being derived from an ontology

**Problem:** The type of Characteristics and (less importantly) Parameters are not declared and are not anchored to any ontological resource. For example when declaring Characteristics[*strain*], '*strain*' cannot be annotated in any way to ontological resources. Susanna Sansone, Marco Brandizi

**Proposed Solution:** Just as for factor, enforce declaration of Characteristics and possibly assigned to a type. It addresses the problem and makes declaration more consistent.

**Conclusion:** This solution is acceptable as long as it is optional for v1.1

**Action Item:** Add to MAGETAB 1.1 specification, provide an example

**Person Responsible:** Don Maier

## 27. Documenting technology/type dependencies

**Problem:** need to document the dependencies between technology types and headers: e.g. assays and technology types. This is a general issue for ISA-TAB proposal but MAGE-TAB might have to face this situation to deal with high throughput sequencing application to transcriptomics. For example: Array Data Files (does not make much sense in absence of Array). Susanna Sansone, Marco Brandizi

**Proposed Solution:** create generic data field header (raw data files, processed data files) + documentation of the specific dependencies depending on applications. Terms to be included in OBI.

**Conclusion:** Was dealt with by adding Assay and Technology Type to v1.1 Issues with data files was discussed and no conclusion was for 1.1. Proposals will be accepted for 1.2.

**Action Item:** N/A

**Person Responsible:** N/A

## 28. ParameterValue[] is uncontrolled

**Problem:** Parameter value can be free text or numeric values – Susanna Sansone, Marco Brandizi

**Proposed Solution:** Add a term source ref to parameter or allow the solution proposed in item 29 below

**Conclusion:** Parameter Value can have a term source ref

**Action Item:** Add to MAGETAB specification 1.1

**Person responsible:** Don Maier

## 29. Dealing with Characteristics Columns of the same type

**Problem:** E.g. Characteristics[OrganismPart] can be duplicate but contain more or less specific information e.g. liver, left lobe in different columns. Was also a MAGE-ML issue. Philippe Rocca-Serra

**Proposed Solution:** Allow for ordering or twinning of columns more formally and document usage in order to facilitate use of format to support such cases (these are real use cases already faced by ArrayExpress).

**Conclusion:** Multiple columns are permitted but making relationships between the columns make the sheets less human readable and attempt to provide ontology relationships in MAGETAB and this is outside the scope of the specification

**Action Item:** Clarify MAGETAB 1.1. documentation to state that multiple columns of characteristics are allowed

**Person responsible:** Don Maier

## 30. Cross over designs

**Problem:** Somehow related to previous comment. Cross over designs aren't simply represented in MAGE-TAB -lack of documentation and structures Philippe Rocca-Serra

**Proposed Solution:** None at present

**Conclusion:** When a solution is proposed it will be considered for the appropriate version

**Action Item:** N/A

**Person responsible:** N/A

## 31. Representing non-hybridization experiments

**Problem:** not all transcriptomics measurement are carried out using hybridization, not all DNA microarray application are not meant to survey gene expression (i.e. ChIP, Genotyping).

**Proposed Solution:** introduce the concept of Assay with 2 qualifiers: 'application/endpoint' and 'technology type'. This is how ISA-TAB deals with the problem and defines the various Assay tabs.

**Conclusion:** Assay/Technology Type have been added, 1.1 will allow these. End Point is not accepted and can be addressed in the ISATAB specification

**Action Item:** N/A

**Person responsible:** N/A

## 32. Three fields to add when specifying an ontology term

**Problem:** with the introduction of Term Accession header, there are now 3 fields to be used for providing a value from an Ontology or CV. The increased number of header reduces legibility and makes document harder to read

**Proposed Solution:** REF + Term Accession columns into one field {value|accession|source}: as in {lymphocyte|CL:0000236|CL}. We can agree on having pipe | as separator between value, accession and source.

In case multiple entries are required, as for item such as role in the IDF, semi-colon; can be used as separator between entries, (using this separator is consistent with its use in IDF and ADF). We realize that this point maybe controversial but one of the first reasons for using spreadsheet tab2mage, MAGE-TAB) was to facilitate data presentation. Having 3 fields for every now seems overly complicated and adds weights.

**Conclusion:** This was unpopular. Not accepted

**Action Item:** N/A

**Person responsible:** N/A

### **33. Factors that are neither Protocol Parameters nor Properties of Sample**

**Problem:** Can Factor be declared so that they are neither a Characteristics of a Biological Material or a Parameter of a Protocol ? Another way to put it: is there ever a case where FactorValue is neither Characteristics nor Parameter?  
Note that both Performer and Date can be treated as equivalent to Protocol Parameter values since (if we think in MAGE-OM terms) Performer and Date are attributes of ProtocolApplications (as would be the values specified for Parameters during that same protocol application). Susanna Sansone, Michael Miller

**Proposed Solution:** make it mandatory that Factor should either be Parameter of a protocol or characteristics of a sample

**Conclusion:** Proposal was rejected.

**Action Item:** N/A

**Person responsible:** N/A

### **34. Names as globally/locally unique identifiers**

**Problem:** Should 'Names' be treated as globally unique identifiers or should they locally resolvable? Michael Miller, Weida Tong

**Proposed Solution:** We need to work on defining when Name can also be REF to external entities (i.e. external to the TAB being considered). It seems to clear for things like protocols and ArrayDesigns but less so for other objects which use Name as identifiers

**Conclusion:** Names are locally resolvable only, this doesn't preclude using a globally resolvable identifier.

**Action Item:** Modify MAGETAB 1.0/1.1 specification

**Person responsible:** Don Maier

### **35. Differentially dimensioned arrays in same SDRF file**

**Problem:** Are differentially dimensioned arrays allowing the same SDRF file ? GenePattern

**Proposed Solution:** N/A

**Conclusion:** Differentially dimensioned arrays, and multi technology investigations are allowed in the same SDRF

**Action Item:** Add to MAGETAB 1.0/1.1 specification

**Person responsible:** Don Maier

### **36. Packaging SDRF/IDF and data files**

**Problem:** Are all files present in the same directory/archive? GenePattern

**Proposed Solution:** N/A

**Conclusion:** SDRF/IDF should be together in same dir or archive, no restrictions on anything else including data and ADF. If these are present then these should be in the same dir or archive with no sub directories

**Action Item:** Add to MAGETAB 1.0/1.1 specification

**Person responsible:** Don Maier

### **37. Annotation Examples in public domain**

**Problem:** Limited ArrayExpress derived annotation examples are available. Helen Parkinson

**Proposed Solution:** Many groups provide such examples and these are cross validated

**Conclusion:** Ari Kahn, caArray and Chris Stoeckert agreed to provide examples

**Action Item:** Send to Helen Parkinson for addition to MGED MAGETAB site  
**Person Responsible:** Helen Parkinson

### **38. Improving MAGETAB take up**

**Problem:** Many MAGETAB applications available, no central list

**Proposed Solution:** Provide a GO tools like list of tools, location and usage. Tools do not need to be completed to be present, planned work is allowed.

**Action Item:** Send to Helen Parkinson for addition to MGED MAGETAB site

**Person Responsible:** Helen Parkinson, GenePattern, Bioconductor, caArray, SMD, Rosetta, U. Penn, Dana Farber etc