Appendix: Input Data Format for CSPLIB Instances

CSPLIB instances are described by ASCII files. Two input formats are possible, one for k-dimensional tables with marginals and $1 \leq k \leq 4$, and another for more general tables with entries linked by linear equations. The second format allows one to also describe k-dimensional tables with marginals, for any k. For both formats, an example is given which refers to the following 2-dimensional table with marginals (sensitive cells are in boldface):

1	**3**	**5**	7	16
10	12	**11**	9	42
8	**6**	4	**2**	20
19	**21**	20	**18**	78

The input format for a k-dimensional table ($1 \leq k \leq 4$) starts with the INTEGER k, followed by k INTEGER numbers giving the table size for each dimension. A list of lines (records) follows, each associated with a different table cell and containing the following $k + 8$ items:

- the first k items are INTEGER numbers identifying the cell indices i (0 for marginals);
- a REAL number representing the nominal value a_i of the cell;
- an INTEGER number representing the loss of information w_i incurred if the cell is suppressed;
- a CHARACTER representing the status of the cell ('u' for a sensitive cell; 'z' for enforcing cell publication; 's' for a potential complementary suppression);
- a REAL number representing the lower bound lb_i on the cell value known to the attacker;
- a REAL number representing the upper bound ub_i on the cell value known to the attacker;
- a REAL number representing the lower protection level LPL_i of the cell required by the statistical office;
- a REAL number representing the upper protection level UPL_i of the cell required by the statistical office;
- a REAL number representing the sliding protection level SPL_i of the cell required by the statistical office.

Here is the input file for the sample instance.

```
============================================== top-of-file
2
3 4
0 0 78.0  4 s 0.0 100.0  0.0  0.0  0.0
0 1 19.0  2 s 0.0 100.0  0.0  0.0  0.0
0 2 21.0  0 u 0.0 100.0  2.0  4.0  0.0
0 3 20.0  2 s 0.0 100.0  0.0  0.0  0.0
0 4 18.0  0 u 0.0 100.0  1.0  3.0  0.0
1 0 16.0  2 s 0.0 100.0  0.0  0.0  0.0
1 1  1.0  0 u 0.0 100.0  0.0  4.0  0.0
1 2  3.0  0 u 0.0 100.0  1.0  3.0  0.0
1 3  5.0  0 u 0.0 100.0  0.0  1.0  0.0
1 4  7.0  1 s 0.0 100.0  0.0  0.0  0.0
2 0 42.0  2 s 0.0 100.0  0.0  0.0  0.0
2 1 10.0  0 u 0.0 100.0  1.0  4.0  0.0
2 2 12.0  1 s 0.0 100.0  0.0  0.0  0.0
2 3 11.0  0 u 0.0 100.0  2.0  3.0  0.0
2 4  9.0  1 s 0.0 100.0  0.0  0.0  0.0
3 0 20.0  2 s 0.0 100.0  0.0  0.0  0.0
3 1  8.0  1 s 0.0 100.0  0.0  0.0  0.0
3 2  6.0  0 u 0.0 100.0  2.0  0.0  0.0
3 3  4.0  1 s 0.0 100.0  0.0  0.0  0.0
3 4  2.0  0 u 0.0 100.0  0.0  4.0  0.0
============================================== end-of-file
```

The input format for a generic table is organized in two sections containing the cell description and the equation description, respectively.

In order to discriminate the generic table format from the previous format, the first item of the file must be '0' (i.e., INTEGER zero).

The cell description section starts with the INTEGER number n of cells, followed by n lines containing 9 items each:

- an INTEGER number from 0 to $n-1$ with the cell index i;
- a REAL number representing the nominal value a_i of the cell;
- an INTEGER number representing the loss of information w_i incurred if the cell is suppressed;
- a CHARACTER representing the status of the cell ('u' for a sensitive cell; 'z' for enforcing cell publication; 's' for a potential complementary suppression);
- a REAL number representing the lower bound lb_i on the cell value known to the attacker;

- a REAL number representing the upper bound ub_i on the cell value known to the attacker;
- a REAL number representing the lower protection level LPL_i of the cell required by the statistical office;
- a REAL number representing the upper protection level UPL_i of the cell required by the statistical office;
- a REAL number representing the sliding protection level SPL_i of the cell required by the statistical office.

The equation description section starts with the INTEGER number m of equations, followed by m lines. Each line starts with a REAL number giving the right-hand side value of the correspondent equation, followed by an INTEGER number k giving the number of its nonzero left-hand side coefficients. The line continues with the separator ':', followed by k pairs of the form i (c_i), where the INTEGER number i gives the cell index and the INTEGER number c_i gives the corresponding nonzero left-hand side coefficient in the equation.

Here is the input file for the sample instance.

```
============================================= top-of-file
0
20
 0 78.0  4 s 0.0 100.0  0.0  0.0  0.0
 1 19.0  2 s 0.0 100.0  0.0  0.0  0.0
 2 21.0  0 u 0.0 100.0  2.0  4.0  0.0
 3 20.0  2 s 0.0 100.0  0.0  0.0  0.0
 4 18.0  0 u 0.0 100.0  1.0  3.0  0.0
 5 16.0  2 s 0.0 100.0  0.0  0.0  0.0
 6  1.0  0 u 0.0 100.0  0.0  4.0  0.0
 7  3.0  0 u 0.0 100.0  1.0  3.0  0.0
 8  5.0  0 u 0.0 100.0  0.0  1.0  0.0
 9  7.0  1 s 0.0 100.0  0.0  0.0  0.0
10 42.0  2 s 0.0 100.0  0.0  0.0  0.0
11 10.0  0 u 0.0 100.0  1.0  4.0  0.0
12 12.0  1 s 0.0 100.0  0.0  0.0  0.0
13 11.0  0 u 0.0 100.0  2.0  3.0  0.0
14  9.0  1 s 0.0 100.0  0.0  0.0  0.0
15 20.0  2 s 0.0 100.0  0.0  0.0  0.0
16  8.0  1 s 0.0 100.0  0.0  0.0  0.0
17  6.0  0 u 0.0 100.0  2.0  0.0  0.0
18  4.0  1 s 0.0 100.0  0.0  0.0  0.0
19  2.0  0 u 0.0 100.0  0.0  4.0  0.0
9
```

```
0.0  5 : 0 (-1) 1 (1) 2 (1) 3 (1) 4 (1)
0.0  5 : 5 (-1) 6 (1) 7 (1) 8 (1) 9 (1)
0.0  5 : 10 (-1) 11 (1) 12 (1) 13 (1) 14 (1)
0.0  5 : 15 (-1) 16 (1) 17 (1) 18 (1) 19 (1)
0.0  4 : 0 (-1) 5 (1) 10 (1) 15 (1)
0.0  4 : 1 (-1) 6 (1) 11 (1) 16 (1)
0.0  4 : 2 (-1) 7 (1) 12 (1) 17 (1)
0.0  4 : 3 (-1) 8 (1) 13 (1) 18 (1)
0.0  4 : 4 (-1) 9 (1) 14 (1) 19 (1)
============================================ end-of-file
```