Direct Processing of Field Data

George M. Angleton, Charles D. Bonham, and Larry L. Shannon

Associate Professor of Biostatistics, Department of Radiation Biology, Graduate Research Assistant, Department of Botany and Plant Pathology, and Instructor, Department of Mathematics, Colorado State University, Fort Collins.

Ecological, agricultural, and range data frequently traverse a processing route which subjects the data to an unacceptable risk pattern of erroneous modification. When the processing of such data becomes routine and involves the utilization of a computer facility, most of the detectable errors arise in the phase of the processing route which lies between the actual initial data recording and its submission to the computer for data reduction. This includes re-transcription of the data and subsequent key-punching with verification.

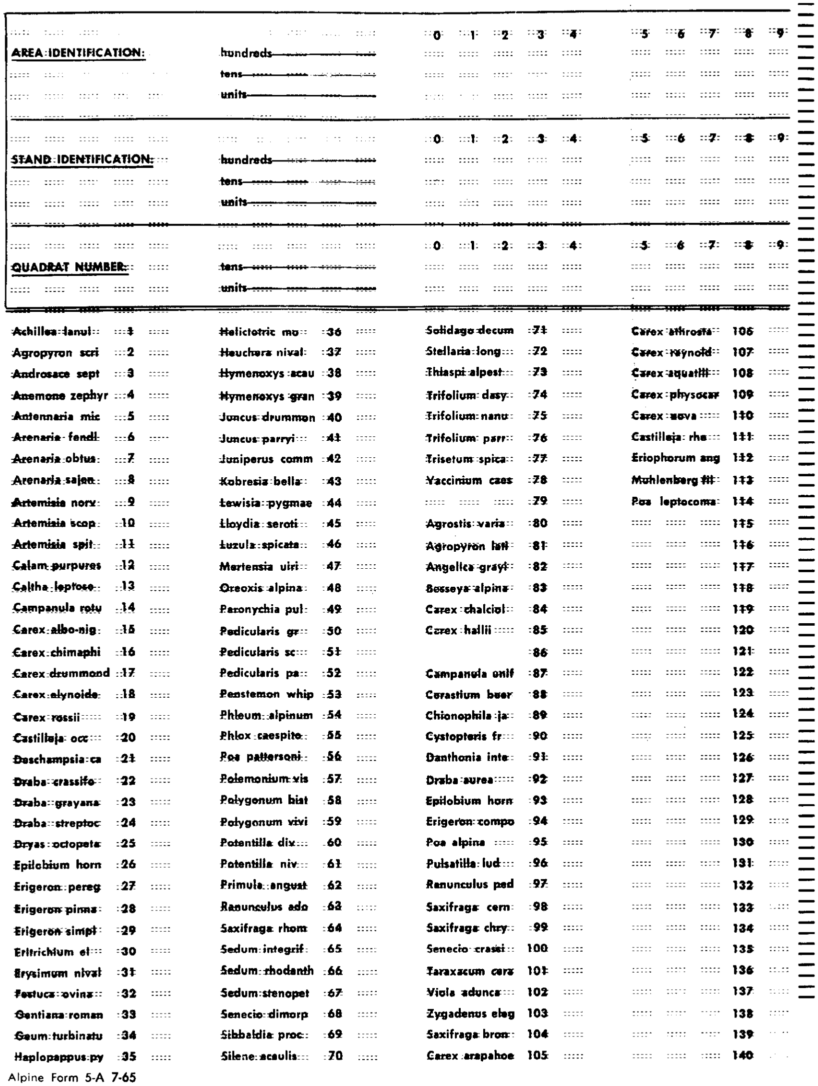
In an operation of small to moderate scope and involving a normal turn-over of support personnel, the frequency of errors arising from re-recording of data can be expected to be moderately high. These errors may be both difficult and expensive to detect and to correct. The experiences of the authors have shown that such errors are primarily the result of either random mistakes or misunderstandings of instructions. While both types of errors are costly, the latter becomes extremely so when large volumes of data are involved.

In addition to the errors arising from data handling there exists the inconvenience that the usual elapsed time from the initial recording of data in the field to the storing of it in a suitable punched card format runs from a minimum time of one week up to several weeks. This means that visual scanning of new data in a meaningful and useful display is delayed for at least an equivalent period of time. In turn, then, the feasibility of in-the-field verification of doubtful results is significantly reduced; whereas, with a rapid data-display return the opposite could be true.

While there may be several ways of overcoming most of these objectionable characteristics of this route of data acquisition, there exists one which may prove to be superior to others from both the practical and economical points of view. This procedure involves the use of special data acquisition forms for use with an optical reader attached on-line to a small computer. The actual facility utilized by the authors for this purpose is located in the Statistical Laboratory of Colorado State University; it includes an IBM 1231 Optical Mark Page Reader attached on-line to an IBM 1401 computer with card, tape and printer input-output devices.

The 1231 is a sensing device capable of detecting and translating markings from an 8½ x 11 inch sheet of paper into records in the memory of the computer. The interpretation and manipulation of the records can be accomplished through programs in such a fashion that the data from the 1231 sheets are simultaneously punched on cards, stored on tape and listed on a print out.

Fig. 1 shows the form we are using in an ecological study of alpine vegetation. Area identification is marked with a black felt pen marker by the observer directly on these sheets; the presence of a species in a given sample plot is then indicated by similar marks in the tracks following the code number of that species. These data sheets are processed without additional error providing the investigator with a data listing and a deck of punched cards. Such processing of data sheets can be performed at the rate of 1,500 to 2,000 per hour, thus providing the investigator with an immediate display



		0	1	2	3	4	5	6	7	8	9
AREA IDENTIFICATION:	hundreds										
	tens										
	units										
		0	1	2	3	4	5	6	7	8	9
STAND IDENTIFICATION:	hundreds										
	tens										
	units										
		0	1	2	3	4	5	6	7	8	9
QUADRAT NUMBER:	tens										
	units										

Achillea lanul	1	Helictotric mo	36	Solidago decum	71	Carex athrosta	106
Agropyron scri	2	Heuchera nival	37	Stellaria long	72	Carex reynold	107
Androsace sept	3	Hymenoxys acau	38	Thlaspi alpest	73	Carex aquatili	108
Anemone zephyr	4	Hymenoxys gran	39	Trifolium dasy	74	Carex physocar	109
Antennaria mic	5	Juncus drummon	40	Trifolium nanu	75	Carex nova	110
Arenaria fendl	6	Juncus parryi	41	Trifolium parr	76	Castilleja rhe	111
Arenaria obtus	7	Juniperus comm	42	Trisetum spica	77	Eriophorum ang	112
Arenaria sajen	8	Kobresia bella	43	Vaccinium caes	78	Muhlenberg fil	113
Artemisia norv	9	Lewisia pygmae	44		79	Poa leptocoma	114
Artemisia scop	10	Lloydia seroti	45	Agrostis varia	80		115
Artemisia spit	11	Luzula spicata	46	Agropyron lati	81		116
Calam purpures	12	Mertensia uiri	47	Angelica grayi	82		117
Caltha leptose	13	Oreoxis alpina	48	Besseya alpina	83		118
Campanula rotu	14	Paronychia pul	49	Carex chalciol	84		119
Carex albo-nig	15	Pedicularis gr	50	Carex hallii	85		120
Carex chimaphi	16	Pedicularis sc	51		86		121
Carex drummond	17	Pedicularis pa	52	Campanula unif	87		122
Carex elynoide	18	Penstemon whip	53	Cerastium beer	88		123
Carex rossii	19	Phleum alpinum	54	Chionophila ja	89		124
Castilleja occ	20	Phlox caespito	55	Cystopteris fr	90		125
Deschampsia ca	21	Poa pattersoni	56	Danthonia inte	91		126
Draba crassifo	22	Polemonium vis	57	Draba aurea	92		127
Draba grayana	23	Polygonum bist	58	Epilobium horn	93		128
Draba streptoc	24	Polygonum vivi	59	Erigeron compo	94		129
Dryas octopeta	25	Potentilla div	60	Poa alpina	95		130
Epilobium horn	26	Potentilla niv	61	Pulsatilla lud	96		131
Erigeron pereg	27	Primula angust	62	Ranunculus ped	97		132
Erigeron pinna	28	Ranunculus ado	63	Saxifraga cern	98		133
Erigeron simpl	29	Saxifraga rhom	64	Saxifraga chry	99		134
Eritrichium el	30	Sedum integrif	65	Senecio crassi	100		135
Erysimum nival	31	Sedum rhodanth	66	Taraxacum cera	101		136
Festuca ovina	32	Sedum stenopet	67	Viola adunca	102		137
Gentiana roman	33	Senecio dimorp	68	Zygadenus eleg	103		138
Geum turbinatu	34	Sibbaldia proc	69	Saxifraga bron	104		139
Haplopappus py	35	Silene acaulis	70	Carex arapahoe	105		140

IBM 551

Alpine Form 5-A 7-65

FIGURE 1. Data acquisition form for in-the-field use in conjunction with an ecological study.

of recently acquired data. Utilizing the described computing facilities we find that the processing cost per document exclusive of card cost is about 2¢ each; utilizing the equivalent facilities of the same institution in the manner described at the beginning of this report, the processing cost per document is about 6¢ each.

Briefly, then, utilizing this type of data acquisition, significant increases in efficiency are obtained with regard to;

1. Accuracy of data procurement.
2. Processing time.
3. Direct costs of data acquisition.

Sample forms will be sent to readers upon request.