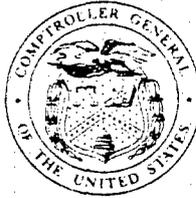


Japikse
Pl 2
14524

DECISION



**THE COMPTROLLER GENERAL
OF THE UNITED STATES**
WASHINGTON, D.C. 20548

FILE: B-191013, B-191013.2 DATE: August 8, 1980

MATTER OF: Information International, Inc.

DIGEST:

1. Once agency denies protest, fact that protester believes agency will reconsider protest does not toll time for filing a protest to GAO since GAO Bid Protest Procedures require protest to be filed within 10 working days of when protester learns of initial adverse agency action.
2. Results of benchmark do not provide proper basis for reconsideration of prior decision dealing with proposed benchmark procedure since benchmark results do not constitute evidence which should have been considered.
3. Protest of methods used to compute costs from benchmark results is untimely where methods used were defined in RFP but protest was not lodged before benchmarking was completed.
4. Since record suggests agency's benchmark-based life-cycle cost approach might not have been sufficiently accurate to support selection of awardee's rather than protester's equipment, and since agency's needs appear to have changed, GAO recommends that agency conduct market survey to determine, before further contract options are exercised, if reliance on awardee's equipment is justified.

Information International, Inc. (III) protests Social Security Administration (SSA) procurement SSA-RFP-78-0001, for multi-font optical scanning equipment

~~011589~~
112979

[Protest Against Equipment Procurement]

designed to machine "read" or "scan" W-2, W-2P, and W-3 forms. The equipment was leased (with an option to purchase) from Recognition Equipment, Inc. (REI), the low evaluated offeror. We sustain a portion of III's protest and dismiss or deny the remainder.

The same procurement was the subject of our decision in Information International, Inc., B-191013, May 31, 1978, 78-1 CPD 406. III also seeks reconsideration of that decision, in which we dismissed as untimely two of III's objections to the benchmark that was required as part of proposal evaluation and found no abuse of discretion by SSA with respect to its use of an "offset printed test deck" to be "read" during the benchmark.

Proposals were evaluated by means of a life-cycle cost method in which various "Bid Equalization Factors" were added to depreciated equipment costs. An estimated cost to the Government was computed for each system configuration evaluated. The amount of each of the bid equalization factors was calculated from data produced during the benchmark.

The systems proposed by III and REI differ in complexity and sophistication. REI offered what is called a "direct paper system," which reads the original documents. III's system reads a microfilm copy and includes an error correction process which permits computer generated images of doubtful characters to be queried, displayed and manually corrected by keyboard operators working in an adjacent terminal room.

Benchmark timing data was used to compute production (throughput) rates which in turn were used to calculate the amount of each vendor's equipment needed to meet SSA's projected workload. Calculation of REI's equipment requirements was relatively straightforward. III, however, qualified proposals for use of two different microfilm cameras (using 16 mm and 35 mm film) and submitted three basic scanning equipment configurations for each. All of III's configurations share a common process but utilize resources differently, providing a range of potential capability at differing costs. Proposals were submitted on lease, full payment lease, lease with option to purchase and purchase terms. While REI was required to offer six units initially and options to furnish two additional units, III's lowest cost evaluated proposal

(35 mm systems on purchase terms) offered three initial so-called dual scanner installations.

III's benchmark performance largely overcame its significantly higher equipment cost and brought its lowest evaluated proposal within approximately \$2.55 million (9 percent) of the REI proposal accepted for award.

Essentially, III says that its proposals should have been evaluated as lowest in cost, that SSA made a number of errors in computing the evaluation factors, that SSA improperly made other "adjustments" to life-cycle costs which improperly penalized III, and that III equipment would have been shown to have been least costly had the benchmark been representative of SSA's actual requirements. III also believes SSA improperly refused to evaluate several III lower cost alternate and "unsolicited" proposals.

I. Reconsideration

III's major objection in its original protest concerned SSA's decision to limit benchmark scanning to the most easily read forms -- approximately 70 percent of the total. Our prior decision concluded that III knew or should have known shortly after December 1, 1977, when it received a letter of that date from SSA, that SSA had rejected the objections which it had raised against this "70 percent limitation." Because III did not file its complaint with us within 10 days thereafter, we viewed the complaint as untimely.

III contends that the December 1 letter should not have been regarded as controlling because it had reason to believe that SSA would still consider the matter in conjunction with consideration of III's other complaint concerning the benchmark. III has now submitted additional evidence regarding a telephone conversation between its General Counsel and SSA's Associate Commissioner for Management and Administration, as well as a letter from its General Counsel to the Associate Commissioner dated February 14, 1978. According to III, this demonstrates that SSA was willing to consider the matter further.

SSA disagrees. It argues that III's evidence is not newly discovered evidence which could not have been considered earlier and denies that the Associate Commissioner

at any time gave III's counsel reason to believe that SSA was contemplating reopening the 70 percent matter for reconsideration.

We find no merit to III's position. The fact that an agency may be willing to further consider a protest which it first rejects does not toll the time by which a protest must be lodged with this Office. Had SSA reexamined III's complaint after December 1, 1977, III's objection to the 70 percent limitation filed here after that reexamination would still have been untimely because our protest procedures require a protest to be filed here within 10 days of initial adverse agency action. 4 C.F.R. § 20.2(a) (1980).

III also argues that its prior protest of the benchmark is inextricably bound to issues raised by its subsequent protest and should be considered at this time because the results achieved by the benchmark are now known.

Our prior decision dealt with a protest filed before benchmarking. The protester sought our review of the proposed benchmark procedure. We were not asked to review other aspects of SSA's evaluation of proposals, or REI's selection, which had not then been made.

We do not believe reconsideration of our decision regarding III pre-benchmark assertions to be appropriate because even if we agreed with III in retrospect that SSA's benchmark assumptions were faulty, we would not agree that its earlier complaints should be considered at this point simply because the test results are now known. The record does not show that our decision dismissing two of the allegations and denying the validity of the other was founded on any error of law or any misunderstanding of the facts existing during the timeframe with which our decision was concerned. See, e.g., Ordnance Research, Inc. -- Reconsideration, B-194043, June 26, 1979, 79-1 CPD 455.

Consequently, III's request for reconsideration is denied.

II. The Protest

A. Preliminary Matters

At the outset, we decline to consider certain aspects of the protest. First, we do not believe it appropriate to review III's objections to SSA's proposed post-scanning error corrections process. SSA's elaborate post-scanning correction process consisted of several steps including manually correcting data. III, however, believes errors experienced during actual operations of the equipment should not be corrected manually through a separate process long after scanning, but that errors due to individual misfilmed documents could be more effectively corrected by refilming documents containing errors during subsequent scanning. The dispute does not involve completely misfilmed batches of documents, which SSA would have refilmed in any event.

We view this issue as outside the purview of the bid protest procedure. It does not concern specification requirements or any other aspect of the procurement for this equipment, but rather the agency's plans for operating the equipment once it is in place. While SSA's application of its corrections process to the benchmark impacted on the life cycle cost evaluation, the validity of the process itself is not, in our view, a proper part of this protest.

Second, we also view as inappropriate for our consideration III's belief that SSA was biased against innovative technology. However wise III may believe SSA should have been in seeking a more technically advanced system, we are aware of no legal requirement that SSA obtain the most technologically sophisticated approach available. On the contrary, SSA's procurement decisions reflect a belief in the importance of proven performance -- a particularly legitimate concern considering the consequences facing SSA if data entry using a scanning process proved unsuccessful. See Pentech Division, Houdaille Industries, Inc., B-192453, June 18, 1980, 80-1 CPD 427; cf. System Development Corporation, 58 Comp. Gen. 475 (1979), 79-1 CPD 303.

We also find that some of III's post-benchmark complaints are untimely.

Section 20.2(b)(1) of our Bid Protest Procedures states that protests based upon alleged improprieties in any type of solicitation which are apparent prior to any closing date for

receipt of initial or amended proposals must be filed by that date. This includes, for example, a date set for submission of additional technical data requested during discussions. In this regard, benchmarking is used for proposal evaluation to produce "descriptive" data which the Government believes is necessary to assess the capabilities and/or cost of equipment proposed. See, e.g., 48 Comp. Gen. 320 (1968); Computer Network Corporation, 56 Comp. Gen. 245, 255-256 (1977), 77-1 CPD 31. We believe, therefore, that a date set for submission of an offeror's benchmark data should be treated as a closing date within the meaning of section 20.2(b)(1). Thus, protests concerning amendments to a solicitation which define how the benchmark and evaluation of benchmark results will be handled must be filed by the benchmark submission date. Cf. Comshare, Inc., B-192927, December 5, 1978, 78-2 CPD 387.

It is in this light that we view III's objection that it should have been charged for costs associated with hiring only one rather than two operators per scanner for its dual scanner configuration. The solicitation provided that offerors would be charged operator costs based upon the number of operators actually used during the benchmark. SSA would have charged III with costs associated with one operator had only one been used for production tasks, but III used two, ostensibly because it did not have sufficient time (four weeks following an amendment deleting a mandatory minimum two operator requirement) to retrain its operators. III was reminded of the solicitation rule during the microfilming portion of the test. Conceding, therefore, that two operators were used, III nevertheless argues that other benchmark parameters were to be determined based on actual measurements of resources required and that operator requirements should be, too. In our view, III should have protested SSA's requirement before benchmarking, or complained then that the time allowed for retraining was not sufficient. Because the objection was raised after evaluation, it will not be considered.)

III also complains that SSA improperly computed certain so-called residual error rates by arbitrarily assuming that only one percent of manually corrected data would contain errors which would require correction a second time and by making its calculation by counting "fields" rather than individual character errors. According to SSA, the one percent figure by fields is the only data available because it has not kept more detailed statistics in the past.

The one percent residual error rate was included in SSA's solicitation cost tables which indicated how the computation would be performed. Thus, III should have known from the RFP the basis on which the calculation would be made and should have protested this matter also prior to the benchmark.

B. Life Cycle Cost Adjustments and Evaluation

III raises a number of objections to specific adjustments which SSA made, or which III believes should have been made, but were not. III also questions various other aspects of the evaluation. Collectively, SSA's adjustments and other scoring assumptions had a significant impact on its calculation of REI's and III's evaluated life-cycle cost so that III's questions must be addressed before its broader concerns -- attacking the meaningfulness of SSA's cost evaluation as basis for selection -- can be considered.

A number of III's complaints relate to SSA's evaluation of the benchmark results, or to its refusal after benchmarking to consider untested alternate approaches.

For example, after benchmarking III first proposed its so-called "paired singles" configuration, one of the three basic equipment configurations III offered. Apparently, III did not conceive of the the paired singles configuration until benchmarking had been completed.

SSA had placed no limit on what configuration could be benchmarked, and presumably would have allowed the paired singles configuration to be tested had it been proposed earlier. It refused, however, to consider the paired singles approach because it was proposed after benchmarking was completed.

III maintains that the paired singles proposal need not be benchmarked because all of its elements were tested and because the benchmark fully established that the equipment would operate at less than 70 percent of capacity even in the paired singles configuration.

We agree with III that SSA ordinarily could not require that the paired singles proposal be benchmarked without reason, just as a contracting activity cannot require unnecessary descriptive data or reject a proposal which fails to include such data. Dominion Engineering Works, Ltd., B-186543, October 8, 1976, 76-2 CPD 324.

III admits, however, that there is a question as to whether sharing resources as proposed with the paired singles approach would limit performance. It attempts to meet this objection by presenting a worst case mathematical analysis showing that it is the computational capacity of the processor which is the critical limiting factor and that adequate capacity would be available.

Even assuming the reasonableness of III's technical argument, we believe SSA's refusal to consider the paired singles proposal without benchmarking is rationally founded. The financial risks facing SSA were substantial. It chose to use benchmarking because it wanted to base award on proven rather than theoretical performance. In this regard, SSA points out that during the course of this procurement III proposed several different approaches and made numerous performance claims, some of which were not borne out in practice. Not all of III's proposed approaches qualified by passing benchmark minimum performance requirements. In SSA's view, analytical abstract calculations predicting the behavior of interrelated system components may not necessarily correlate with actual results, imposing in effect unacceptable risks. SSA's decision therefore, reflects its best judgment; that judgment has not been shown to be arbitrary.

Moreover, we note that III knew before it ran the benchmark that SSA would insist that each system proposed be benchmarked because III was told specifically that SSA would not agree to limit benchmarking to the double scanner system. None of the factors III cites to support its claim that the paired singles approach did not occur to it earlier because SSA's requirements were constantly changing involve matters which III would not have considered before benchmarking. Obviously, III, by not proposing this approach prior to benchmarking, ran the risk of not having that approach accepted.

III also complains that SSA improperly adjusted certain benchmark data, or failed to adjust all offeror's data equally, giving REI an unfair advantage. For example, SSA found that in some instances character and field substitution and reject rates did not correspond. REI's proved to be especially sensitive to the location of characters within prescribed areas on the test forms. Where characters appeared outside the box assigned to a specific field, the equipment placed them in the

wrong field. SSA says it disregarded such errors if the characters printed on the form were read correctly but were placed in an incorrect field. III believes, however, that the adjustment favored REI and was unfair because edit functions were tested during the benchmark.

SSA admits that REI obtained a larger adjustment in substitution rate than did III but argues that this was proper in the circumstances, noting that:

"Since these types of errors are easily corrected through field editing and because editing of the data was not part of the [benchmark] requirement, it was decided that the properly read but dislocated characters should not be counted as substitutions. III had a smaller amount of adjustment because its scanning parameters were established differently than REI's."

In this regard, we see no basis for protest if an agency adjusts benchmark data in evaluating it, provided the adjustments made are reasonably related to the announced evaluation criteria and provided the basis on which benchmarks were run or evaluation criteria used are not altered. We find nothing improper with the adjustments made which are based on SSA's belief that dislocated characters would be improperly counted if they were treated as substitution characters which appears to be overall consistent with the life-cycle costing evaluation criterion. Cf. AEL Service Corporation, 53 Comp. Gen. 800 (1974), 74-1 CPD 217. Unless III could establish, as it has not, that (1) it set its machine parameters for benchmarking in reliance on the included edit requirement; (2) these parameters were set to minimize the kinds of errors which SSA later corrected; and (3) III otherwise could have adjusted its parameters to better optimize performance, it has no basis for complaint.

Further, III complains that SSA arbitrarily imposed: (1) a 30 second penalty for job set-up; and (2) a 3 percent penalty for operation of output edit software, which was tested but not required to be used compiling benchmark data.

III has not presented evidence establishing that these

adjustments were unreasonable, per se. Job set-up time was not measured by the benchmark because only one work unit (5,000 documents) was tested. A similar penalty was applied in computing REI throughput. The effect of the penalty was to downgrade throughput somewhat to afford SSA assurance that the quantity of equipment acquired would provide some margin for set-up time. Likewise, the 3 percent was assessed to account for system degradation while editing functions were being performed. A 10 percent degradation was observed during the benchmark, but SSA agreed to the lower figure to account for expected enhanced performance once the higher level language used during the benchmark was converted to the DEC-10 machine language which would have been used if III had received award. Consequently, III's objections to these adjustments are without merit.

We agree with III, however, that several of SSA's adjustments, and its refusal in one instance to consider an adjustment, were inappropriate.

For example, we agree with III that SSA improperly applied the 30 second penalty to magnetic tape changes by arbitrarily adding 30 seconds to the processing time required for each 5,000 document unit. As III points out, a 2,400 foot reel of magnetic tape will hold the information recorded on 35 batches of documents, or approximately 175,000 documents. SSA insists that the requirement is necessary, nevertheless, because the 5,000 page limitation is required to meet subsequent operational steps which SSA would use in annual reporting.

Imposing a 5,000 page limit is not the only way information in 5,000 page blocks could be handled. Data can be "blocked" electronically in 5,000 page units even though many 5,000 page units are recorded on one tape, and we understand such a procedure is possible for use on the equipment SSA planned to use, which would significantly reduce the number of tapes needed. SSA's approach appears to be based on its experience with REI type equipment. III's in-line correction process, however, would have reduced significantly the amount of post-scanning processing which might be required otherwise. SSA's evaluation of III's approach on the basis of what SSA would require by use of REI's system is unreasonable.

In addition to the 30 second and 3 percent penalties, SSA added a 90 second allowance for film changes at the scanner.

III agrees that 90 seconds is a reasonable time for the changes, but questions the manner in which it was applied, pointing out that one scanner in a dual scanner configuration normally continues to operate while film is being changed in the other.

SSA states that it:

"* * * is aware of the situation cited by III * * *. However, SSA could not accept the premise that there would be no situations in which both films would not have to be changed simultaneously. The quality of the scanner input and the number of fonts involved would be among the various factors which will affect the read rate of the scanners. Regardless of attempts taken to preclude such simultaneous changing of the film in both scanners in a dual-scanner systems during full operations, there will be random occasions when both film transports run out of film at, essentially, the same time."

Believing that it would be speculative to attempt to gauge the rate that this would occur, SSA nevertheless estimates that III would require about 1,100 film changes per month using 35 mm film, or 560 changes if 16 mm film were employed.

To the extent that film changes are statistically meaningful, changes in both scanners of a dual scanner system should be counted. Knowing approximately what number of film changes are required, SSA could have estimated statistically the likelihood that two film changes would be required within any 90 second period. Although probably de minimus, the extent to which film changes in a dual scanner system are not statistically independent events could have been evaluated by measuring any changes in throughput of one scanner while the other was not operating.

III also believes that some kind of factor should have been applied in evaluating REI's benchmark results to account for possible paper jams just as III was penalized for costs associated with manually reprocessing individually misfilmed documents. SSA views III's complaint as untimely. We disagree. Although III knew that numerous adjustments were made to its scores, it was entitled to assume that comparable appropriate

adjustments were being made to other proposals. Accordingly, we treat III's complaint as timely since it was filed within 10 days of the date it knew no such adjustments had been made.

SSA argues that both the effect of paper jams on REI's direct paper systems and misfilmed documents on the III process were evaluated. They were evaluated differently, SSA asserts, because the processes differ. REI ran the benchmark as a single continuous process. Paper jams or other technical problems in its system would have showed up directly in reduced throughput, because problems would have to be corrected as they arose. This could not occur with the III process, since until the microfilm used was developed there could be no determination as to misfilmed documents. Since SSA could not measure misfilming errors during the III benchmark it compensated by applying an adjustment based on experience with other microfilm uses.

As a result, III was evaluated using so-called "real world data"; REI was evaluated on this item on the basis of actual benchmark results. SSA does not address whether this was equitable. How much of a difference would have resulted from use of actual forms received by SSA rather than the neatly stacked offset printed forms actually used is unclear but it seems apparent that the difference should have had an effect on REI's evaluation. Thus we believe a "real world" adjustment should also have been applied to REI's test results.

III raises several complaints whose relationship to SSA benchmark is less direct, but which are important nevertheless in laying a foundation on which we may review SSA's life-cycle cost evaluation.

For example, III maintains that microfilm purchase and processing costs were evaluated improperly. Except for III's full service plan offering to set up contractor-owned contractor-operated (COCO) processing facilities, III's proposals assumed that the Government would furnish film and processing. III maintains, however, that by considering only Eastman Kodak Film and processing, SSA overestimated III's life-cycle cost by at least \$500,000. In III's view, SSA should have surveyed potential film and processing suppliers to determine pricing, using the Federal Supply Schedule (FSS) as a starting point.

SSA states that it used Kodak pricing because III specified Kodak film in its proposal and used it during the benchmark. SSA admits that III proposed Rochester Film products as an alternate source, but states that the Rochester Film proposal was received after the closing date for receipt of best and final offers.

III admits that it identified two types of Kodak film by number and used Kodak film in the benchmark. We believe that by failing to indicate that other film might be acceptable, III ran the risk that SSA would believe the film chosen might be critical to contract performance and evaluate on that basis. Consequently, we do not object to this aspect of the evaluation.

III also objects to SSA's rejection of its so-called best and final "zero preventive maintenance" proposal, in which it entered "zero" for scheduled daily preventive maintenance time. Arguing that the concept of preventive maintenance is made obsolete by modern semiconductor technology, III maintains that maintenance should be scheduled only as required. III says daily scheduled maintenance down time is unnecessary, because modern equipment, e.g., memory, is constantly monitored and shows gradual degeneration, thereby permitting planned replacement. In III's view, it is unrealistic to deduct one or two hours from available working time to allow time for maintenance which will not be performed.

SSA asserts that III's complaint is untimely, and rejects III's view that preventive maintenance time and down time are conceptually interchangeable. Noting that offerors were required from the outset to state how much time was to be set aside for preventive maintenance, SSA asserts that III's protest on this issue should have been filed before the closing date for receipt of initial proposals. Arguing that "pre-scheduled maintenance" is preventive maintenance regardless of what III wishes to call it, SSA questions how III could achieve a 92 percent (or for that matter, a 90 percent) availability rate without scheduling some kind of maintenance.

III counters that the issue is timely and that, moreover, its equipment achieves a 97 to 98 percent availability rate.

We do not view III's complaint as untimely because nothing in the RFP prevented consideration of an offer proposing zero

time for preventive maintenance. SSA may not have meant to allow a zero time proposal. It did not preclude one, however, and III protested as soon as it learned its zero time offer had been rejected. The rule SSA relies upon applies only to defects which are apparent on the face of a solicitation. 4 C.F.R. § 20.2(b)(1).

With respect to the merits of the issue, we point out that the burden is on the offeror in submitting his best and final offer to affirmatively demonstrate its merits. The contracting officer need not reopen negotiations, or speculate as to whether an unsubstantiated proposal could be supported with adequate technical data, but may downgrade or reject the proposal as the circumstances warrant. Here, III's zero preventive maintenance proposal marked a significant departure from its earlier approach, and SSA did not find anything convincing in III's proposal with respect to the validity of this approach. We find no basis for taking exception to SSA's rejection of this approach.

From SSA's answer to this allegation, however, we have some doubt regarding SSA's consistency in applying its life cycle costing technique. In explaining its position, SSA argues that downtime and time allocated for preventive maintenance:

"were never envisioned as, nor should they be considered, interchangeable elements. The 90 percent up time [10 percent possible downtime] requirement is only a threshold factor. SSA does not plan for the OCR system to be down for 10 percent of the time, in addition to preventive maintenance. The 90 percent criterion is the level below which the contractor must pay penalties." (Emphasis added.)

Interchangeable or not, III correctly notes that downtime and time for preventive maintenance were treated as cumulative. SSA added their separate contribution together, in effect, by deducting both from total available time to compute monthly operating time and production figures. Moreover, an offeror naturally is induced to trade preventive maintenance for availability to maximize the calculated cost effectiveness of its equipment. We note that REI submitted an alternate proposal, also rejected, offering 75 minutes daily preventive maintenance (down from 2 1/2 hours) plus 6 hours of weekly [weekend] maintenance.

We agree with SSA that unscheduled downtime and time during which the equipment will be shut down for scheduled servicing are different. We see no basis either for objecting to SSA's use of a threshold figure for determining when liquidated damages should apply, or for that matter, to its use of a 90 percent figure for computing throughput and equipment requirements, provided the contracting activity reasonably believes that a 10 percent downtime rate will be experienced.

Our concern is that SSA did not believe that a 10 percent downtime figure will occur. In its own words, it did "not plan for the * * * system to be down 10 percent of the time." Based on the experience of other users of III equipment, SSA seems to accept a 92 percent rate as achievable. SSA, III says, included planned downtime for preventive maintenance in computing the 92 percent figure from III experience data at other installations in effect discounting it twice. Correctly calculated, the data shows that 97 to 98 percent is achievable. Thus, we think there is some question as to the reliability of SSA's costing approach as it impacts on the overall cost evaluation.

At this point, we turn to examine the impact which errors in SSA's adjustments could have had on its evaluation of the REI and III proposals, and in this light, to examine III's complaint that SSA's methodology did not in fact provide a valid analytical measure of the relative cost of its and REI's probable life cycle costs. Our review of the accuracy of SSA's methodology will focus on III's 35 mm purchase proposal, because that proposal was evaluated as lowest in cost and because our analysis indicates that no evaluating error made by SSA would permit any other III proposal to displace the 35 mm purchase as least costly.

SSA did not attempt to differentiate through the benchmarking between the vendors' ability to read a particular portion of SSA's projected workload. If we understand its intended methodology correctly, it assumed that REI and III would be able to process the projected 70 percent workload. Contrary apparently to III's view of what should have been done, SSA instead attempted to compute the cost impact which could be expected due to random sources of error, in effect introducing a source of "noise" (randomly formed characters) and measuring system response to it.

REI produced significantly higher figures, and received correspondingly greater cost penalties. The difference in magnitude of the REI and III's scores, however, is of little importance, because the accuracy of SSA's cost evaluation depends upon the precision with which these projected figures were measured in calculating the cost difference.

It is difficult in this regard to understand how the benchmark results could bear any rational relationship to projected comparative costs--at least to within an accuracy of better than \$2.55 million, the original difference separating the III and REI proposals. Our calculations indicate that costs were extremely sensitive to numeric balance field substitution errors while the cost of processing rejected balance fields or reinstatement items was of comparatively negligible importance. One single numeric (balance field) substitution "error" has a \$24,000 cost impact. A single "error" in SSA's construction of test set reproduced ten fold would have had an impact of almost a quarter of a million dollars. A minimum of one hundred seven individual SSA character "errors" in a 5,000 page test deck could account for the entire difference SSA calculated between REI and III life-cycle costs.

By attempting to use a single test deck to measure throughput and accuracy, rather than multiple test decks and appropriately adjusted weighting factors, SSA had to assure that the deck would be meaningful for two quite different purposes. To the extent that processing time might reflect the difficulty of the material read, the test deck had to be representative of the total projected annual reporting workload, leaving only a very small portion of the benchmark to have any effect on accuracy.

SSA's methodology, however, included virtually no checks in the benchmark process. Apparently, SSA lacked any quality controls save visual inspection of portions of the test materials and the standards imposed on the offset printing process. It would have us believe that it nevertheless could distinguish test results to an accuracy approaching one part in ten thousand, the accuracy required to measure the impact of balance field substitution errors with the precision needed to support SSA's cost justification for selecting REI. SSA evaluated test results by comparing each vendor's data with what SSA "knew" had been typed on the original 500 page test set. So far as we are aware, SSA did not attempt to manually keyboard the

test deck contents to determine what if any minimum (residual) error rate it had introduced. SSA, moreover, found that offset printing produced changes, eliminating fine lines.

Forced by III to admit that it did not budget for post-scanning processing costs at rates determined by the benchmark, SSA argues that its representation in III's prior protest that the benchmark would produce meaningful results was not meant to convey the notion that there would be an exact correspondence between benchmark and actual performance--only that the results would be representative. It did order the quantity of equipment which its benchmark based throughput calculations indicated would be needed.

This, however, amounts to applying a double standard -- one to compare accuracy and a second for calculating equipment requirements. Meaningful results consistent with a life cycle costing approach are obtained regardless of the scale chosen to measure costs. Scoring will not be meaningful or rationally founded, however, if disproportionate weights are assessed different elements making up the total. A rational relationship is not maintained where any one significant contribution to total costs -- here, out-of-pocket equipment cost -- is keyed to a fixed unit of measure while other equally significant costs are not. Nor is it enough that SSA believes the accuracy portion of the benchmark produced relative performance data. III won that round, but lost because its equipment was more expensive than REI's. A weighting factor error of two in the accuracy, i.e., if SSA's test should have been twice as difficult, makes a \$12 million difference (\$7 million for balancing alone) in the parties relative standing.

Further, we question the adequacy of SSA's throughput results. SSA did not use the entire test deck to determine III scanner throughput rates. Evaluation was based on a 1,000 page microfilm sample for the single scanner system. A 4,000 page sample (divided into four 1,000 page subsamples) was used to time throughput for the III dual scanner system. III's scanners read the material in 12 (single scanner) to 18 (dual scanner) minutes per 1,000 page subsample. Using SSA's 435.6 usable hours per month estimate, the 8 year evaluation period consists of 1,672,704 usable minutes--a magnification of more than one hundred thousand times the (15 minutes) average scanner test period. Timings of the four subsamples included in the dual scanner test varied in extreme by more than one half

percent. Moreover, to test a hypothesis that no deterioration would be experienced, some disk and core were removed, and one 16 mm single scanner configuration was tested, twice. Oddly, performance using less equipment improved by 2.28 percent.

The effect of a variation of but a few percent can have disproportionate impact where as here offeror's cost proposals are evaluated based on an integral number of units, by rounding any calculated fractional equipment to the next whole number. Use of a quantified evaluation procedure can, and indeed in this case did, skew the relationship between a change in scanner throughput rates and equipment costs.

Calculating the potential impact of scanning throughput error is somewhat complex because III was free to propose any combination of single or dual scanning systems it wished, provided that SSA's peak monthly workload figures were met.

Raw and adjusted throughput rates for III's 35 mm equipment are as follows in pages per minute:

<u>Configuration</u>	<u>Raw</u>	<u>SSA Adjusted</u>	<u>GAO Adjusted</u>
Single Scanner	78.6	72.46	72.97
Dual Scanner	115.1	103.62	108.08

In this connection, SSA adjusted III's throughput rate downward by including time for tape and film changes as well as the three percent edit program degradation factor. (The GAO adjusted figure includes a 30 second tape change factor only once every 175,000 pages and does not assume that the dual scanner stops each time a film change is needed for either scanner.)

Our review indicates that SSA's minimum throughput requirement would have been met by a III proposal offering less equipment, if the corrected SSA throughput rates are too low by as little as three percent. A comparable increase in REI's computed throughput rates has no such effect, however, because REI's equipment requirements are altered downward only by a 10 percent variation and then only in years 1 through 3.

If, instead, SSA's benchmark procedure produced faulty throughput rates so that the calculated throughput rates are too high, the difference is more easily absorbed by the equipment proposed by III than that offered by REI. A decrease of only 1.5 percent in REI throughput would require that an additional machine be added during years 4 and 5 because REI met SSA's requirement by furnishing only 8 units because SSA did not require that it furnish document numbering during those two years. (REI is required to furnish document numbering during years 1 through 3 and after year 5 while III was required to propose microfilming equipment with numbering throughput the evaluation period. Evidently, SSA viewed automatic numbering as merely a matter of convenience.)

C. Incidental Issues

The protester has stated several complaints which we consider to be only incidental to our decision.

For example, III complains that SSA required that microfilming throughput rates be computed using a so-called "document flipper" because the III TDC (planetary) cameras otherwise would invert the sequence in which the documents are arranged. SSA says it expects to hire poorly qualified temporary personnel to retrieve and process documents for correction during reinstatement processing, and consequently, needs to assure that a consistent document sequence is maintained. The cost of the flipper itself is insignificant. We find this issue to be inconsequential.

Throughput rates for 35 mm microfilming equipment were determined by running two 500 page samples. Times of 13 minutes, 22 seconds and 11 minutes, 42 seconds were produced for the III proposed 35 mm camera using the document flipper. Although equipment requirements were based on the total time required for the two runs, the 6.6 percent difference between each of them and their average alone is enough to account for one camera of the eight SSA required that III offer to meet year 1 through 3 requirements.

SSA's computations, in this regard, showed that its initial workload could be processed if 7.3 cameras, and supporting staff and facilities, were provided. This 7.3 camera figure, however, includes a five percent degradation SSA estimated would result from use of the document flipper. Adjusting SSA's

figures to eliminate this factor, but considering the 6.6 percent spread, shows a requirement for between 6.5 and 7.4 cameras with slightly less than 7 cameras the most likely estimate. The years 4 and 5 figures would be 9.8 and 11.4 with 10.5 cameras most likely. The effect of a one camera savings for years 1 through 3 thus has a limited effect on SSA's calculation of III's evaluated cost reducing the difference in III and REI pricing by less than \$100,000, to almost inconsequential proportions.

III further complains of SSA's refusal to consider its so-called "full service proposal" in which III offered to establish a COCO facility to process all SSA machine readable forms. Conceding evidently that the full service proposal could not be considered under the RFP, III states that it was offered as an unsolicited proposal and that SSA's failure to consider it "is simply another piece of evidence that SSA's purpose was to obtain familiar technology, not to save cost." Although SSA tenders a multitude of reasons supporting its decision to reject this proposal, it is enough to point out that SSA could not accept it because it was clearly outside the scope of the procurement and because the Federal Procurement Regulations (FPR) do not anticipate acceptance of an unsolicited proposal to furnish supplies or services which would normally be procured by competitive methods. FPR § 1-4.901. To have accepted this proposal would amount to a sole-source award without justification.

III also believes SSA counted existing SSA-owned or leased REI equipment in computing REI equipment requirements, but not in determining how much equipment III would have to furnish. Nothing in the record supports III's concern. At the outset of the procurement SSA owned one REI scanner and leased another, using them to process Quarterly Report requirements. The second machine was leased in part to meet excess requirements until the current procurement was completed. The lease since has been terminated, while the first REI machine continues to be used to meet other SSA scanning requirements. There would have been no proper basis for III to have objected, however, if SSA had advised offerors that existing surplus capacity would be considered or if SSA were to use any surplus capacity for annual reporting purposes.

Finally, SSA raises various questions regarding actual performance under the REI contract. The possibility, of course,

that a contract does not work out as expected is not material to a bid protest, because the reasonableness of the assumptions made during the procurement process must be judged by examining the circumstances as they were then believed to exist. Performance, moreover, is ordinarily a matter of contract administration which is not considered by this Office.

D. Conclusion and Recommendations

Although we believe the benchmark methodology used may have provided a reasonable basis for determining the competitive range, we agree with III that the validity of the benchmark as an analytical tool for distinguishing between its and REI's proposals is questionable. Great precision is required of the accuracy portion of the benchmark if it is to serve as a basis for drawing a rational distinction between the III and REI proposals based on differences in indirect costs. Moreover, because indirect costs were added to direct (equipment and equipment related manpower, space and facilities) costs, it would not have been sufficient had SSA managed, as it believes it did, to measure a "relative difference" in III and REI performance. The two types of costs must be measured using a standard which permits them to be compared. SSA's use of validation methods-- specification of off-set printing standards coupled with visual inspection of but a fraction of the test materials-- appear on their face inadequate to assure that anything close to the necessary degree of accuracy was maintained. Its evident lack of concern that all factors (e.g., actual anticipated availability and the effect of 200 rather than 26 typefonts, as discussed in the SSA technical evaluation) be considered and its belief that relative accuracy data was sufficient leave us without any foundation from which to conclude that the comparatively close cost data computed for III and REI necessarily reflects a measured difference in the life cycle cost of either the REI or III system. The record thus suggests to us that the life cycle cost evaluation was inconclusive with respect to measuring the costs the Government could reasonably assume it would incur.

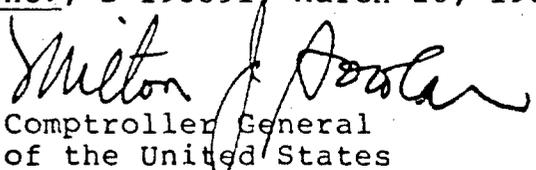
However, we are aware of no evidence indicating SSA acted other than in good faith, or that SSA would not have awarded III a contract had III been able to establish that its approach would be more cost effective than REI's. We believe this result occurred because the procurement was structured for direct paper systems using a benchmark that

was conceived to discriminate between relatively similar equipment. SSA restructured its solicitation, allowing microfilm based processes, once it became clear that III might be able to compete. The benchmark was adapted to permit comparison of microfilm and direct paper systems, but was pressed beyond its limits not as a result of any SSA desire to preclude III but because III's performance and higher equipment costs focused the competition along lines which the benchmark methodology was incapable of handling.

[In view of the uncertainty and difficulty SSA faced, and the inconclusive nature of the evaluation scores it computed, we believe it would be appropriate for SSA to validate its initial procurement decision. Thus, we believe SSA should conduct a market survey to determine whether continued reliance on REI equipment actually serves the Government's best interest, providing in connection with the survey an opportunity for further testing, using statistically representative samples selected from actual annual reporting data. In this connection we note that while SSA professes to be satisfied with the REI equipment, SSA's anticipated working environment has changed since the procurement was conducted. For example, although SSA had microfilmed earnings documents for more than 20 years, it viewed microfilming as an unnecessary expense with scanning and as a unique cost to be charged to III. Since making award, SSA has entered into an agreement with the IRS which requires that SSA microfilm all forms for the IRS. We also understand that original documents are being retained on site until the microfilm copy is developed, altering in part the assumptions on which SSA states its approach to post-scanning processing was based. Since microfilming costs alone contributed more than five million dollars to the cost of the III proposal, a quite different result might be achieved were the procurement conducted today. Moreover, a test conducted today should suffer few of the difficulties SSA faced, because SSA knows what its experience has been and has actual forms from which a statistically representative workload sample could be selected. A much larger sample could be used, reducing the statistical significance of any one individual anomaly.]

By separate letter to the Secretary of Health and Human Services, we are recommending such action be considered before any decision is made to exercise further REI contract options. See B&W Stat Laboratory, Inc., B-195391, March 10, 1980, 80-1 CPD 184.

For the


Comptroller General
of the United States