

Appendix 1 Summary of the validation committee report

1. Committee Membership

The validation committee consists of the following six members:

Ken-ichi Uehara (Ken-ichi Uehara, Tsukuba Industrial Liaison and Cooperative Research Center, Professor, Chair)

Kuniaki Takigawa (Certified accountant, system audit)

Minako Takizawa (Science journalist)

Toshio Fukushima (National Astronomical Observatory of Japan, Public Relations Center, Director, Professor)

Ken-ichi Miura (Research Organization of Information and Systems, National Institute of Informatics, Professor emeritus)

Takuo Yamaguchi (Tsukuba Academia Law Office, lawyer)

2. Outline of the data loss

The data transfer in question was carried out from June 2011 to January 2012. In this process, some of the files that should have been copied were not properly identified and as a result, some data was lost in the transfer process. Without knowledge of this fact, by April 20th 2012 the contractor had erased the data on the original tapes in accordance with the contract, rendering the data files that had not been copied unrecoverable. Thus, a partial data loss resulted. The first indication of missing data was noted in mid-May 2012, and as the data storage in the new computer system became fully accessible, the data loss of a major scale became evident in July 2012. The extent of the data loss amounted to 29% in terms of data size, or 62% in terms of file count, of what should have been copied. However, a considerable fraction of missing data was recovered from distributed backup sources. Considering the sequence of events that led to this incident, the committee considers that it was due to sheer luck that the final magnitude of data loss was contained to this level.

3. Cause of the data loss

Clearly, this incident has been caused by human error and not by an inevitable accident or equipment failure. While the final magnitude of the data loss was very limited, the initial loss during the data transfer process was very large. The incident was caused by a combination of errors in the lists of files for copying, which were prepared by a contract worker who cooperated in this work, the lack of rigorous checks of the file lists, which should have been performed by KEK staff members, and the vulnerability in the data transfer procedure, including the schedule, and particularly, that of the erasure of data on the original tapes.

4. Effect of the data loss

The data that was unrecoverable consisted of a fraction of the raw data from the Belle experiment and a small fraction of the physics data for the analysis of phenomena

*This summary was prepared by KEK by quoting the contents of the validation committee report.

related to hadron physics, which was outside the original scope of the Belle experiment. The physics data pertaining to the original research objective of the Belle experiment have been fully recovered. The outcome of the Belle experiment is represented by papers published in scientific journals. Since all the physics data related to the original scope of the Belle experiment have been recovered, no adverse effect is foreseen on the relevant physics papers that may be published in the future. Naturally, the papers that have already been published will also not suffer from any effect of the loss of raw data. Data for subjects (hadron physics) outside the original scope of the Belle experiment are of interest to Ph.D. students as topics for their thesis work. The papers already published in this category are obviously not affected. However, one paper, for which the analysis work is in progress, as of March 2013, is affected in that the statistical accuracy of the data reported in this paper will be slightly compromised. However, the supervisor of this Ph.D. student has confirmed that there should be no problem in publishing the work. Collaborators in the Belle group have reported that the impact of the data loss is small and negligible in light of carrying out their research activities.

5. Conclusion

After reviewing the materials provided by KEK and through interviews of individuals involved in and potentially affected by this incident, the committee could presently not observe any practical damage caused by this data loss. Members of the concerned research groups have not brought to KEK's attention any specific or practical damage caused by the data loss. The committee concludes that this data loss incident has not caused any damage to the scientific accomplishments of the Belle experiment reported in the past and neither will it cause any such damage to them in the future. However, the committee has to point out that the procedure adopted for the data transfer had numerous problems. This incident raised concerns among researchers whose work is related to the Belle and other experiments at KEK, although no practical damage is reported. If in the future, actual damage occurs due to another similar incident, KEK is likely to lose the confidence of the international community formed by researchers from related scientific fields. Therefore, KEK should reflect seriously on its practice surrounding the matters related to this incident.

6. Recommendation

The committee concurs with the proposal given by the investigation committee concerning the actions to be taken by KEK. In addition, the committee recommends that KEK should reinforce its organizational policy on information security and impel discussions on the present and long-term preservation of scientific data in the future. The direct cause of this data loss was human error. However, humans are prone to errors and computer programs and operating systems are prone to having bugs. These facts must be considered as a part of the underlying assumptions when the operation and management of organizations are being planned and executed. In this sense, the root cause of this incident should not be trivialized as the error of individuals. Instead, this incident should be taken as an important opportunity to identify the hidden risks and organizational issues involved in carrying out operations at KEK. Further, it should be noted that this incident imposed a certain extra workload on the concerned researchers

and staff members. KEK should take these aspects very seriously and strive hard to make improvements in this regard.

¹ After recovery, the final data loss amounted to 5% of raw data and 2.6% of the special physics data which is related to one analysis outside the scope of the original research objective (a subject that emerged later during the experiment).