

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N氏 (素粒子原子核研究所現場責任者)	M氏 (計算科学センター現場責任者)	Y氏 (素粒子原子核研究所長)	S氏 (Belle collaboration代表者)	H氏 (Belle部内事情に詳しい者)	SS氏 (調査委員会委員)	K氏 (協力業者 N社)	T氏 (協力業者 S社)
◎ 対象者全員に対する確認事項					*			
I 本事案の確認								
1. 貴方は、本事案の兆候もしくは発生を、「いつ」、「どのように」知りましたか。	5月17日にユーザから一部のファイルがないとの連絡をメールで受けた。 大規模なデータ損失があるとわかったのは7月17日で、やはりメールでの連絡で受けた。	5月24日にN氏から一部のデータが移行されていないようであるので、N社に問い合わせしてほしい旨のメールがあったので、問い合わせ内容をN社に伝えた。 この際は、大規模なデータ損失があったとの認識はなかった。 その後7月31日と8月5日にN氏からのメールで、大規模なデータ損失があったとの連絡があったため事態を認識した。	7月終わりあるいは8月はじめにS氏より口頭で連絡を受け、8月6日にN氏による報告文書を受け取った。	5月下旬に新中央計算機にコピーされていないファイルがある可能性を聞いた。 7月下旬に生データ等の損失をN氏の報告を聞いて知った。	KEKCC-users という共通計算機(KEKCC)上のBelle users 全員を対象にしたメーリングリストに6月初旬から、Tau_SkimA に一部読めないファイルがあること、run-dependent Beam Backgorund ファイルが無いこと、などが報告され始めていた。 ただし、これらのファイルは国内外の他機関にコピーされていることもあり、Belleデータ移行に大きな問題がある認識は、この時点では無かった。 その後、N氏から口頭で、一部raw dataが旧Belle計算機からKEKCCにコピーされていないらしいと聞き、最終的に全体像を知ったのは7月のミーティングで報告されたときである。	噂話で聞いたが、時期は覚えていない。	8月6日に貴機構計算科学センターのM氏よりメールにて連絡があり把握いたしました。	8月7日にN社K氏よりメールにて連絡がありました。
2. 貴方は、本事案の兆候もしくは発生を知った際、どのような行動等を行いましたか。	5月17日には当該ファイルを作り直す準備をするように担当者依頼した。 7月17日には、Belle実験グループに報告。被害状況の調査を開始。	5月24日の時点で、N社にB計算機のテープが処分されているか問い合わせたが、すでに処分済みであった。 8月5日の時点では、すでにB計算機にあったデータは復旧不能であることが、5月24日の質問に対する問い合わせで明らかであったため、1) N氏と協力して、つくばB3のBelle所有のHSMシステムから一部のデータを救いだせないか、N氏、S社、N社を含めて検討を始めた。 2) 原因調査報告をN社経由でS者に依頼をした。	報告を受けた当初はコピー元のテープが6000本近くあることからすでにすべて消去されていることはなかろうと考えたが、テープが残っている確認が取れなかったため8月9日にT理事に報告した。	N氏に調査の結果をまとめ報告書を作成するように依頼した。 Y氏に口頭で報告し、報告書を提出した。	情報が不足していたので、事情を知らない人間が今後の作業内容について指示をすることは、逆に混乱を招きかねない経験上判断した。 そのため実際に作業をどのように行ったのかを明らかにする必要は感じたが、担当者がその作業をするために復旧できるものも復旧できなくなる可能性があるため、当事者に一任するのが最適と考え、個人的には行動は起こしていない。	特段になし。	貴機構より連絡をいただき、事実確認のため実際に作業した業者との打ち合わせを行い、その後、貴機構のM氏、N氏を交え復旧方法や対策等について打ち合わせ・検討を行いました。	本件の発生メール連絡を受けた際に、以下の2点の要求について確認・検討する必要があると、KEK様と打合せをしたと相談を受けました。 1. 間違いが起こった経緯の追跡および、破棄されたであろうデータの消去の歴史と現状 2. つくばB3 HSMシステムからの一部データ回復の検討

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N 氏 (素粒子原子核研究所現場責任者)	M 氏 (計算科学センター現場責任者)	Y 氏 (素粒子原子核研究所長)	S 氏 (Belle collaboration代表者)	H 氏 (Belle部内事情に詳しい者)	SS 氏 (調査委員会委員)	K 氏 (協力業者 N 社)	T 氏 (協力業者 S 社)
II 原因					*			
3. 貴方は、本事業が起きたことについて何が原因だと考えていますか。(本事案は、現象的にはデータ移行時にデータを損失したこと、また、データバックアップが完璧でなかったことの二つの事象の複合の結果発生したものと考えられますか。)	<p>現象的には、その通りである。原因は、直接的には会社の移行作業上のミスであるが、それが起こった原因、検出できなかった原因は、複合的なものだと考えられる。</p>	<p>直接理由としては、データ移行時にミスで移行すべきファイルが移行されておらず、また、ミスが判明するのが遅れたために、データがすでに消去されていたためであると考えている。また、データのバックアップがあれば、消失は防がれた。</p>	<p>原因は複数あると考えるが、データ欠損が起こり、すぐに発見できなかった点に絞って言えば、Belle、計算科学センター、協力企業の3者間で責任分担が明確でなく、誰かが何とかするだろう的な相互依存があったことが最も直接的な問題であったと考える。その背後には時間的に切迫していたこと、作業をオーバーサイトする仕組みの欠如、データ(特に生データ)に関する重要性の認識の欠如などがある。個人的にもっとも後悔するのはデータ移行の完全性が確認できるまでコピー元のテープを保存しなかったことである。また、データの分散保管が不完全であったのは、作業と確認をPNNLに完全に依存したことにある。</p>	<p>移行ディレクトリーリストからSEが作成した移行ファイルリストに漏れがあったこと。物理データの最終的な損失は、上記の二つの事象の複合の結果発生したものと考える。</p>	<p>多重な原因によるものと考えられる。raw dataに関しては直接の原因はBelle側で用意したコピーすべきファイルのリストと、それを元に業者が作成したファイルリストに不整合があった事である。一方、mdstと呼ばれる物理解析用データについてはファイルリストの不整合ならびに他機関に依頼していたデータコピーが不完全だったことが挙げられる。</p>	<p>human errorを防ぐ手だてが為されていなかったこと。</p>	<p>データ移行作業およびその準備に十分な時間が割けず、移行データの確認作業がしっかり出来なかったことは原因の1つと考えております。また、確認する体制を弊社として十分に確保できなかったことも要因と考えております。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>
4. 貴方は、本事案は未然に防止できたものと考えますか、あるいは不可避だったと考えますか。	<p>原理的には防止できたと思う。</p>	<p>データ移行時に移行すべきファイルがリストから漏れていた点は人為的なミスであるから、(個人の注意、人員、費用、期間等 いろいろな意味での)コストを増やせば防止が可能であったと思う。データバックアップが不完全であるのは、もともと予算がないということを受容していたリスクであったと考える。たとえば、B計算機のデータは運用中バックアップもなく一か所で保管されていた。特に過去のデータはコンテナハウスで保管されている。天災、火災、雨漏りなどでも失われる。</p>	<p>もっとも基本的なミスは元データのテープを早々に廃棄したことであると考える。これを未然に防ぐことは実は容易であった。</p>	<p>未然に防止できたものと考えられる。</p>	<p>問題がどこで起こったかがほぼ判明した現段階においては、その部分を重点的に検査するシステムを導入することで理論的には回避できないことはない、と考える。ただし、当時の金銭的、時間的、人的資源が不足していた中、優先順位を付けて作業を進めなければならない状況で、それが本当に可能であったかどうかは疑問である。</p>	<p>防止できた。</p>	<p>3の回答に関する部分が事前に準備出来ていれば、未然に防止できた可能性はあると考えております。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N 氏 (素粒子原子核研究所現場責任者)	M 氏 (計算科学センター現場責任者)	Y 氏 (素粒子原子核研究所長)	S 氏 (Belle collaboration代表者)	H 氏 (Belle部内事情に詳しい者)	SS 氏 (調査委員会委員)	K 氏 (協力業者 N 社)	T 氏 (協力業者 S 社)
5. 貴方は、本事実の未然防止を妨げたのは何であると考えますか。	私の担当部分では、ファイルリストの確認をしていれば防げた可能性は高いと考えている。全体的には、予算と時間、役割分担の不足が、妨げになったと考えている。	データ移行時のミスについては、①移行直後にBelle一般ユーザーによるデータアクセスができなかったためにデータが削除されるまでにクロスチェックがされなかった。②当初予定の急な変更により移行スケジュールがタイトになり、スケジュールの遅れを取り戻すことに注力したため、作業側のチェックがおろそかになった。ことがこの順番でより近い要因であったと考える。 移行元システムのデータをすべて長期間保管しておけば、ミスがあってもデータ喪失は防ぎ得たが、これはKEKの強い方針がないと無理であろうと思われる。	データの重要性の認識の欠如が本質的であったと考える。これは直接作業を担当した者に限らず、グループ全体の傾向として物理データさえあればという意識があった。	予算が十分ではなく、十分な時間でより確実に移行が行なえるような契約のもとで作業ができなかったこと。 そのような状況のもとでの作業に対して十分な認識と配慮がなされなかったこと。	* 一人の問題ではなく、携わった全ての人のデータ管理に対する認識の甘さ。	人的資源および予算の枯渇。	回答する立場ではありませんので、回答を控えさせていただきます。	回答する立場ではありませんので、回答を控えさせていただきます。
Ⅲ 調査委員会による報告書								
6. 貴方は、報告書に記載されている内容について承知していますか。	はい。	調査委員会委員長より報告書の写しを頂いた。	承知している。	承知している。	承知している。	はい。	存じ上げておりませんので、回答を控えさせていただきます。	存じ上げておりませんので、回答を控えさせていただきます。
7. 報告書の内容に関して、定性的・定量的な記述、表現、総括のしかたを問わず、意見等がありましたら、忌憚なく申し述べてください。	おおむね正しいと思う。	今回の事故の直接の関係者であるので、報告書に対して意見を言う立場にない。 事実について気づく範囲で誤りは無かった。	最終的に失われたデータが何であるかが明瞭に述べられていないことは気になるが、内容は正しいと考える。	調査委員会の報告書は事実の記述と技術的な原因の性格な記述に重点を置いたので、専門的記述が多くなるのは避けられず、一般人には分かりにくい表現になった。	報告書はBelleデータ損失に関わることなので、特に付加する事項ではないが、raw dataをKEK外の機関にコピーすることが現実的ではないネットワーク環境、地震からの復旧作業など背景に関する記述があっても良いと感じた。	特段にない。	報告書の内容を存じ上げておりませんので、回答を控えさせていただきます。	報告書の内容を存じ上げておりませんので、回答を控えさせていただきます。

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N氏 (素粒子原子核研究所現場責任者)	M氏 (計算科学センター現場責任者)	Y氏 (素粒子原子核研究所長)	S氏 (Belle collaboration代表者)	H氏 (Belle部内事情に詳しい者)	SS氏 (調査委員会委員)	K氏 (協力業者 N社)	T氏 (協力業者 S社)
IV 本事業についての感想等					*			
8. 貴方は、本事業が発生したことについて、どのように感じていますか。	<p>自分の不注意や判断ミスで問題を防げなかった部分があり、責任を感じている。</p> <p>特に、ファイルリストの確認については、もう少し注意を払うべきであった。</p> <p>一方で、根本的な問題は移行の進め方にあったのではないかと考えている。</p> <p>被害が限定的だったのは不幸中の幸いだった。起こってしまったことなので、これを教訓にすべきことが多々あると思う。</p>	<p>作業SEの監督をする立場でありデータが損失したことに責任を感じる。</p> <p>今回の移行作業が、業者責任でなく、KEK責任で行われることについて、N氏に十分説明できていたか？</p> <p>また、どのファイルが重要でどのファイルが不要なのかを選択して移行したが、それはBelle側しか判断できないため、その役割はBelle側に一任したつもりであったが、その作業の役割分担についてBelle側と計算科学センター側で明確になっていたか？</p> <p>そうでないのであればそれを説明する責任があった。</p>	<p>大変残念な事案であり、Belleの代表を長年務めた者として責任を痛感している。</p> <p>物理データ以外のデータを大事にしない雰囲気を作ったことは私の責任によるところが大きい。(例として物理データが確認できれば生データは積極的に廃棄しても良いのではないかとという提案を行っている。)</p>	<p>Belle実験におけるKEK所属の共同代表として、本件が起きたことに責任を感じている。</p>	<p>物理解析に影響は出なかったとはいえ、復旧できないデータがあることは非常に遺憾である。</p>	<p>起きたことは残念だが、今後の教訓として生かしたい。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>
9. 本事業と同じことを繰り返さないようにするためには、何が考えられますか。	<p>報告書にかかっている通りだと思うが、実務としては、役割分担と責任の所在をはっきりすること、第三者のチェックが必要だと思う。</p> <p>特に、Belleと計算科学センターの間の役割分担をはっきりさせる必要があると思う。</p> <p>あとは十分な時間をかけて予定をたてることが重要だと思う。</p>	<p>今回と同じことを行うとした場合には、いろいろな対処が考えられるが、いずれも後知恵ともいえる。</p> <p>調査報告書にあるように、かなり危うい事とわかっていながら、何らかの理由でどうしても行わなければならない場合は、実際に現場でかかわっている者たちとは、離れた客観的な目で忠告をする者の存在が不可欠であるように思われる。</p> <p>しかしながら、その者はある程度現場を知っていなければならず、かつ当該作業について自分のこととしてかかわる必要があると言えらると思う。</p> <p>そのような人員を確保できるであろうか？</p>	<p>本事業と全く同じ過ちを繰り返さないということでは、データとは貴重なものであるという当然のことをよく認識したうえで作業の方法、オーバーサイトの在り方を見直すということになるが、もっと一般化して、深刻な事態に至りうるミスを犯さないという点では、時間的、人的余裕をもった作業計画を立案することが必要と考える。</p>	<p>報告書に書かれているとおりであると考えます。</p>	<p>基本的には報告書4「今後の対策の指針」に記載されている通りである。</p> <p>強いて強調するならば、データ管理は重要である、という認識を皆が持つこと。</p> <p>作業レベルでデータ移行の重要性とそれに必要な資源や手順を説いても、それが適切にBelleグループ、素核研、計算科学センター、機構に受け入れられなければこの問題は再び起こる。</p> <p>この認識があることが前提で、</p> <ul style="list-style-type: none"> ・ Belle側としてはファイル移行に関する手順の第三者確認システムの構築 ・ 新規計算システムの柔軟な導入(システム切り替えで使えない期間をなくす。たとえば計算機を半分ずつ入れ替えるなど) <p>のシステムや慣例に捕らわれない柔軟な対応策が必要である。</p>	<p>人的資源の十分な配分を要し、精神論からの脱却を目指す。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N氏 (素粒子原子核研究所現場責任者)	M氏 (計算科学センター現場責任者)	Y氏 (素粒子原子核研究所長)	S氏 (Belle collaboration代表者)	H氏 (Belle部内事情に詳しい者)	SS氏 (調査委員会委員)	K氏 (協力業者 N社)	T氏 (協力業者 S社)
◎ 該当する対象者に対する確認事項					*			
V 個別具体の確認 (データの損失に関して)								
10. バックアップ等の復旧作業の結果も含めて、最終結果として損失した物理解析用データの内訳及び総量はどの位になりますか。	生データでは全体の5%くらいが失われた。 物理データは、ypipi skimと呼ばれるハドロン共鳴の解析用のデータが2.7%失われた。	該当しない。	生データのうち約5%、50TB、および物理データのうち0.09%、0.14TB。(但し、この数字は自ら確認したものではない。)	生データは、全データ1050TBのうちの50TB(5%)相当が損失した。 物理解析用データは、エネルギースキャンデータのypipi mdstが0.12TBのうちの0.009TB(7%)およびY(5S)でのデータのypipi mdstが5.3TBのうちの13TB(2.5%)が損失した。これは物理データ全て(およそ150TB)の0.09%である。	報告書に記載されているとおりraw dataについて復旧できないものは全体の5%で約50TB。 物理解析用データの内、Ypipiと呼ばれる一部のデータセットのみであり、物理解析用データ内での比率は0.9%、0.14TBが損失。 なお、物理解析には影響はない。	該当しない。	データ内容を存じ上げておりませんので、回答を控えさせていただきます。	データ内容を存じ上げておりませんので、回答を控えさせていただきます。
11. Monte-Carlo 疑似データ及びBackgroundデータについては、それぞれの位の損失量になりますか。	失われた分については復旧しているので、損失はない。	該当しない。	MCデータについては調査委員会報告書にある通り。 Backgroundについては全量復元できた。(但し、自ら確認はしていない。)	Background データは他の機関の計算機にコピーされており、Monte-Carlo疑似データ復元可能なので実質的な損失はない。	Monte-Carlo疑似データが意味するところが分からないが、仮にMonte-Carloで作った物理解析用データに関する損失はない。もしMC generator fileを意味するならば全て損失。約2TB。 Backgroundデータについても全て損失。約1TB。 ただし、損失したMC generatorについてはMC物理解析用データに組み込まれているので実害はないし、新しく作る際にはより新しいバージョンのMC generatorを使う必要があるため、今回の損失による影響はない。 Backgroundデータに関しては容量として小さいので、既に他機関にコピーされており、これらをKEKに再度コピーすることで対処済みである。	該当しない。	データ内容を存じ上げておりませんので、回答を控えさせていただきます。	データ内容を存じ上げておりませんので、回答を控えさせていただきます。
12. users、subdirsのデータについては、それぞれの位の損失量になりますか。	容量で46%。	該当しない。	調査委員会報告書にある通り。(但し、自ら確認はしていない。)	461.4TBのうちの217.4TB(47%)が損失した。 これらは、過去に解析で使用したデータなどが主であり、必要なら同じ解析を行なうことにより復元可能である。	報告書に書かれている以上のことは把握していない。 報告書によると uses、subdirsそれぞれ206TB(46%)、11TB(85%)となっている。	該当しない。	データ内容を存じ上げておりませんので、回答を控えさせていただきます。	データ内容を存じ上げておりませんので、回答を控えさせていただきます。

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N 氏 (素粒子原子核研究所現場責任者)	M 氏 (計算科学センター現場責任者)	Y 氏 (素粒子原子核研究所長)	S 氏 (Belle collaboration代表者)	H 氏 (Belle部内事情に詳しい者)	SS 氏 (調査委員会委員)	K 氏 (協力業者 N 社)	T 氏 (協力業者 S 社)
13. 上記の「10」「11」「12」の質問項目について、損害状況が未確認の箇所がありましたら、その認識概況、実態確認に要する時間の目途を申し述べてください。	特になし。	該当しない。	ユーザーデータについては、かつて作成してそのままになっているNtupleやスキムデータなどが多くあったと考えられるが、作成したユーザー本人も確認できないものが多くあると想像している。 これらは未確認かつ実態把握は不可能であるが、実際上損害とは言えない部分が大部分である。	損失量の未確認の箇所はない。	* 現在の所、更なる損失は認識していない。	該当しない。	データ内容を存じ上げておりませんので、回答を控えさせていただきます。	データ内容を存じ上げておりませんので、回答を控えさせていただきます。
14. 上記の「10」「11」「12」の質問項目に関するBelleグループ内での感想について、どのように認識・把握していますか。また、そのためにBelleグループ内でのインパクト把握について、誰がどのように図りましたか又は図ろうとしていますか。	生データが失われたことについては、残念に思う人は多いと思う。 ただ、実際のところ、生データを使って何かをしたかった、という人はいても、実際に全部を使おうとする人はほとんどいないので、5%失われたことについて実害があるようには思えない。 データ損失については、Belleグループの全体会議で報告した。 また問12のユーザーデータについてはユーザにメールで被害状況の報告を求めたが、反応はほとんどなかった。 ユーザーデータは、そもそも半年間アクセスできない予定だったので、重要なものは別の場所に待避されていたものが多かったのでは、と推測している。	該当しない。	グループミーティングで損失状況が報告され、回復の可能性について議論が行われた。 現グループ代表者のS氏がその把握を行った。	Belle コンピューティングマネージャーのN氏より、Belleグループにデータ損失の報告および質問等の依頼がなされた。 少数の問い合わせがあり可能な対処がなされ、大きなトラブルや苦情は起きていないとの報告を受けている。	グループ全体には今夏のグループミーティングで報告された。 データが損失したことについて、何故起こったのか、再発を防止するための案、などの意見が交わされたと認識している。	該当しない。	データ内容を存じ上げておりませんので、回答を控えさせていただきます。	データ内容を存じ上げておりませんので、回答を控えさせていただきます。

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N 氏 (素粒子原子核研究所現場責任者)	M 氏 (計算科学センター現場責任者)	Y 氏 (素粒子原子核研究所長)	S 氏 (Belle collaboration代表者)	H 氏 (Belle部内事情に詳しい者)	SS 氏 (調査委員会委員)	K 氏 (協力業者 N 社)	T 氏 (協力業者 S 社)
(データの移行に関して)					*			
15. データ移行作業を行う際のBelle グループ内外の責任・対応・報告体制について、認識されているところを改めて申し述べてください。	<p>データ移行は計算科学センターが中心に行うと認識していた。</p> <p>私はBelleグループの計算機担当として、その作業に参加するとともに、Belleグループへの意見の集約や、Belleグループへの報告などを行った。</p>	<p>Belle側の責任・対応・報告については該当しない。</p> <p>作業自体の責任に対しては、作業前に明確に役割分担をして作業していたわけではない。これは、今回の事だけでなくB計算機の運用全般にもいえる。</p> <p>移行すべきデータの選別、データ移行の際にどの順番に行うか、どのサーバの運用を停止し移行するか等についてはBelle側が行い、計算科学センターは、作業が順調におこなわれるか、移行選択したファイルが確実に転送されているかの確認などを分担した。作業については当初はメールで、スケジュールが遅れてからは定期的な会合で情報を共有し対処を決定した。</p>	<p>Belle側の直接の作業担当者であったN氏が計算科学センター・M氏や関係会社と連絡を取りつつ作業を行っていた。</p> <p>その報告はグループ代表者であるS氏が受けていた。</p> <p>作業に関わる予算については私(当時素核研副所長)やN氏(当時素核研所長)が相談を受けていた。</p>	<p>Belle コンピューティングマネージャーのN氏がBelle側から移行作業に参加した。</p> <p>移行の計画や進行状況は、適宜Belleグループおよび共同代表者に報告され、必要に応じて議論がなされた。</p> <p>Belle グループ内外への最終対応および責任は、共同代表者にある。</p>	<p>Belle spokespersonの判断により、Belle側から一名(N氏)がこのデータ移行に従事するために割り当てられた。</p> <p>計算機センター側から一名(M氏)が割り当てられたことは後に情報を得た。</p> <p>その他、業者側の人数については関知していない。</p> <p>責任・対応についてはこれら割り当てられた者間で決められると認識していた。</p> <p>報告体制については、情報として漏れ聞くことはあったが、基本的にはデータ移行のコアメンバーではなく、報告を受け立場にはない。</p>	該当しない。	回答する立場ではありませんので、回答を控えさせていただきます。	回答する立場ではありませんので、回答を控えさせていただきます。

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N 氏 (素粒子原子核研究所現場責任者)	M 氏 (計算科学センター現場責任者)	Y 氏 (素粒子原子核研究所長)	S 氏 (Belle collaboration代表者)	H 氏 (Belle部内事情に詳しい者)	SS 氏 (調査委員会委員)	K 氏 (協力業者 N 社)	T 氏 (協力業者 S 社)
16. 貴方は、1年の作業予定を7か月に短縮したことにより、当然作業工程に無理が生じることを予想されましたか。それに対して貴方は、どのような注意を払いましたか(例えば、特別な工程管理を考えるなど)。	<p>予想した。</p> <p>そのために、一部のファイルだけを選んでコピーする方法に変更になった。</p> <p>必要なファイルの選定には十分に注意し、ディレクトリリストを会社側にわたした。</p> <p>しかし、作業時間が短縮されたことよりも、B計算機のリース期間中に移行することになったのが作業工程に大きく影響した。</p>	<p>質問意図と異なるかもしれないが、まず1年の作業予定が7か月に短縮された事「だけ」が問題ではない。</p> <p>当初計画では、B計算機と新中央計算機を同時に動作しながら移行する予定であったが、計画変更後は、同時に動作することがなく、また両システムの運転が同時に停止し数カ月のデータにアクセスできない時期があった。</p> <p>同時期に両システムが動作しなかったために、B計算機運用中にBelleユーザーがデータ解析を続けながら移行を行わなければならなかったため、実際に移行に利用できるI/O資源(テープドライブ、サーバー)に制限が生じ、当初予定より大幅に時間がかかった。また、エンドユーザーによるデータの損失チェックが出来なかった。</p> <p>質問の趣旨に戻ると、これらのことから工程に無理が生じ、移行作業がB計算機撤去までに間に合うかどうか、ぎりぎりの状態で推移した。このため、計算科学センターではN氏と協力しSEを含めて10月以降ほぼ毎週打ち合わせで進捗管理を行い、またデータ転送に抜けが無いよう、データ転送中にファイルが変質しないように業者を指導し、チェックした。これについては、役割を果たせたと思う(ミスはデータ転送にはなく、データファイルリスト作成にあった)。</p>	<p>7か月でやることは聞いていたが、期間の短縮によって特に無理が生じるとは考えておらず、特段の注意は払っていない。</p>	<p>作業工程が厳しいことは予想されたが、Belleグループの責任者として十分な配慮が足りなかった。</p>	<p>* 通常1年かかると考えられていた作業を7ヶ月間でやること自体に無理が生じることは十分考えられた。</p> <p>ただし、我々Belle側としては金銭的な制限からこれ以外に選択肢はあり得なかった。</p> <p>またHSMからコピーするファイルを取捨選択することで、専門家が7ヶ月でできる、と判断したことに、外部から「不可能」と指摘することは出来ない。</p> <p>指摘する場合はそれに代わる代替案を許された予算と期間、人的資源の中で提示しない以上、意味をなさない。</p> <p>一方注意として、一般的にデータの移行自体に問題が起こりえることは、2010年夏のグループミーティング等で個人的に指摘しており、その際に物理解析用データのコピーだけでもKEK外の他機関で持ってもらえないか、と提案している。その際にはBelleグループの内部でさえ、真摯に必要性を認識してもらえなかった。</p> <p>その後2011年3月に地震があり、KEKの計算機システムが使えなくなった際、全ての解析作業が停止したのを受け、データコピーの重要性を各機関も認識し、限られた電力事情の中で計算機の復旧共々データコピーの作業を行った。実際、このコピーを行ったお陰で、復旧できたデータも多くあると考えている。</p>	<p>該当しない。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N氏 (素粒子原子核研究所現場責任者)	M氏 (計算科学センター現場責任者)	Y氏 (素粒子原子核研究所長)	S氏 (Belle collaboration代表者)	H氏 (Belle部内事情に詳しい者)	SS氏 (調査委員会委員)	K氏 (協力業者 N社)	T氏 (協力業者 S社)
17. プロジェクトの規模と比較して、作業人数が少なくなったとの指摘が報告書にあります。そのようなのは何故ですか。また、誰がその作業人数で行うことをどのように判断し、了承しましたか。	従来の移行でも人数は同じ程度(計算機の運用に関わっていたメンバー)だったので、誰の決定というわけではなく、既存のメンバーで作業を行うことになったのだと思う。 しかし、急に複雑な移行方法になったのに、作業の人数は増やさなかった(ただ、計算センターから作業が複雑になるから入念な確認をしてほしい、というような要求もなく、私の方からBelle側に人を出してほしいという希望はだしていません)。	計算機作業側に限って述べると、SE作業人数は少なかった。 これは、「協力すること」というB計算機の仕様の事項だけで作業を依頼したために、常駐SEの範囲で行う必要があったためである。 B計算機の通常運用は継続していたため、作業に従事したSEは通常の業務の一部も行っていった。	該当せず。 (Belleグループの現在の陣容、Belle-IIIに向けての作業状況、データの重要性の認識の度合いからしてデータコピーに一名しか割り当てられないという判断になったと推測する。判断したのはS氏で、私には了解が求められたわけではないが、求められたとしても了解していたと考える。)	Belle実験に関連するBelleメンバーの作業の体制および人選は、共同代表者が責任を持って行なう。 データ移行作業には、計算科学センターおよび業者の人も従事するのでBelleから一人でもよいと判断したが認識が甘かったといえる。	* そもそもデータ管理に対する認識の現れだと考えている。計算機運営、データ管理等は加速器、検出器と並んで実験の重要な要素であるにもかかわらず、検出器に関わる人数に比べて圧倒的に計算機に関わる人数がKEKのBelleグループ内で少ない。 その状況の下、Belle実験、Belle II実験の計算機・ソフトウェアに関わる全てを取り仕切らなければならない、データ移行に専念して出来る人材を確保することは難しい。 この状況を鑑みて、最終的にはspokespersonがBelleデータ移行に一人割り当てると判断したと認識している。	該当しない。	回答する立場ではありませんので、回答を控えさせていただきます。	回答する立場ではありませんので、回答を控えさせていただきます。
18. 移行ファイルリストの作成段階において、クロスチェックが行われなかったのは何故ですか。	ディレクトリリストからファイルリストを作ることは簡単な作業で間違えることはないと思っていた。 また、ファイルリストのファイルを実際に期限内に移行するところが大変だったので、そちらにばかり注意が行っていた。 そもそもBelleからは移行ディレクトリリストを渡した時点で、Belle側の役割は基本的には終わっていたという認識だった。	該当しない。	該当せず。 (正しいの思い込みが先行したためと推測する。)	移行ファイルリストの作成作業は比較的簡単であり、BelleのN氏と計算科学センター側のチェックで十分であると判断し、Belle側でさらにクロスチェックを行なう人員を手配しなかった。	Belle側で用意すべきは移行すべきファイルのリストを作成することである。 そしてこのリストを元に、コピーの能率を上げるため、どのテープにどのようなデータが入っているかを選別し、業者がリストを焼き直し、と考えていたので、リストのチェックは業者もしくは計算科学センターで行われるものである、と認識していた。	該当しない。	回答する立場ではありませんので、回答を控えさせていただきます。	回答する立場ではありませんので、回答を控えさせていただきます。
19. 契約会社との責任分担が曖昧であるとのこと報告書で指摘されていますが、そのようなのは何故ですか。また、誰が契約内容をどのように判断し、了承しましたか。	該当しない。 (契約関係は計算科学センターが担当で、移行している当時はどういう契約で行われているのかは知らなかった。)	契約会社との責任分担は曖昧でなく明確。 作業はKEKが主体となりおこない、B計算機業者、新中央計算機業者は協力を行った。 ただし、実作業で両業者間での意思疎通や作業分担については調整が必要であったのは事実。 予算圧縮の観点から、当初計画どおり各業者に移行作業を別に発注することをせず行うことについて最終的に判断したのは、計算科学センター長。	必要経費を削減するために、責任分担が曖昧になることを承知の上で「協力を求める」としたものと考える。 判断、了承は契約のKEK側の主体であった計算科学センターで行われた。	「データ移行作業」の契約による作業ではなく、「協力する」という態勢で作業を行なったためと考える。 契約内容の判断・了承はセンターが行なったので「該当しない」。	旧Belle計算機について誰がどういう契約を結んだのかは関知しないので該当しない。	該当しない。	回答する立場ではありませんので、回答を控えさせていただきます。	回答する立場ではありませんので、回答を控えさせていただきます。

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N 氏 (素粒子原子核研究所現場責任者)	M 氏 (計算科学センター現場責任者)	Y 氏 (素粒子原子核研究所長)	S 氏 (Belle collaboration代表者)	H 氏 (Belle部内事情に詳しい者)	SS 氏 (調査委員会委員)	K 氏 (協力業者 N 社)	T 氏 (協力業者 S 社)
20. 報告書において、「2011年11月4日のデータ転送ミーティングで、真鍋氏が「総転送容量が1.27PBとなっているが、テープライブラリの総容量の半分以下である。本当に間違っていないか?」と打合せの資料[10]で疑問を呈している。ここでより深くチェックされていれば、問題を発見できた可能性がある。」という記述がありますが、その際に、総転送容量が非常に少ないにもかかわらず諒とされた理由・事情は何ですか。	急に一部ファイルだけ移行するという方式になった関係で、移行データの総容量は不明だったため、単に移行データの総容量が少ないだけで正しいかもしれないと思った。 また、B計算機の運用中にデータ移行を行うことになったため、テープライブラリの容量などの情報で正確な値が取得できず、報告された値が信頼できるものではなかった。そもそも、この時点で深くチェックするには時間が足りなかった。	経緯を説明させて下さい。 B計算機のテープライブラリの容量は3.5PB 磁気ディスクの容量は1PBあった。どこにどのようなデータを置くかは全てBelleが差配していた。これは、複数のグループがリソースを共有している共通計算機や中央計算機とは異なる状況。 計算科学センターでは最大テープライブラリの3.5PBを移行する必要があると考えていたが、当初6月20日には2.8PBとされ、その後磁気ディスクにあるファイルは別途移行をおこなうことで7月28日時点には1.8PBに低減され、さらに重複のファイルをできるだけ転送しないこと等のBelleグループの努力で9月26日には概算1.5TBとなった。 転送量はBelle側のHSMデータの利用終了まで確定しなかったが、それが11月4日の時点で1.3PBと確定したと記憶している。 このように転送する量をできるだけ圧縮してスケジュールの遅れを取り戻そうとした。 したがって、3.5PBのデータ量があるのに、いきなり1.3PBの転送量とされて変に思わないのか?という単純なものではない。 また、どのディレクトリのデータを転送するのをやめて切り詰められるかの判断はBelle側にしかできないため、念のため大丈夫かもう一度確認してほしいという意味で11月4日(実際には以降回数)に記述した。	一般にテープに書かれるデータ量はレコードギャップなどのためにテープの最大容量に比べて少ないのが普通で、半分であったとしても疑問を抱く理由にはならなかったものと考ええる。 転送すべきデータ量が1830TBであったことをあらかじめ把握する手段があったはずであるが、時間的制約のためにそれを割愛したものと推測する。	調査の結果、テープにある全てのデータではなく、一部の必要なデータのみを移行したが、Belleデータのデータベースが不十分だったことなどで移行すべきデータの総量が簡単に把握できなかったためと考える。	* これについても、実際にこのミーティングの現場にいたわけではないので該当しない。 しかし、全てを移行するのではなく、ファイルを取捨選択して移行するのであるからテープライブラリの総容量と食い違いがあっても不自然ではない。 一方、総容量の内、どれくらいが実際に使われており、それと比較して極端に転送容量が低い、と分かっていた場合には、より詳しいチェックを要請することが出来たのではないかと考える。	該当しない。	回答する立場ではありませんので、回答を控えさせていただきます。	回答する立場ではありませんので、回答を控えさせていただきます。

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N 氏 (素粒子原子核研究所現場責任者)	M 氏 (計算科学センター現場責任者)	Y 氏 (素粒子原子核研究所長)	S 氏 (Belle collaboration代表者)	H 氏 (Belle部内事情に詳しい者)	SS 氏 (調査委員会委員)	K 氏 (協力業者 N 社)	T 氏 (協力業者 S 社)
(問題発生要因等に関して)					*			
21. Directory丸ごとコピーでなく、ファイル名指定でのコピーなど、結果として作業を複雑にしてリスクを増やすことを選択しなければならなかったとの指摘が報告書にあります。そのようなことになったのは何故ですか。また、誰がどのように判断し、了承しましたか。	<p>必要なデータだけコピーしようとする、これしか方法がない。</p> <p>移行の打合せ(私を含むKEK担当者との打合せ)で必然的にそのように決まったと思う。この部分は特にBelleグループに報告していない。</p> <p>必要なデータだけコピーする方針は、計算科学センター(所長)から移行の方針の決定事項として伝えられたという認識である。一部のデータのみコピーするというのはBelleに報告して了承されている。</p>	<p>「Directory」丸ごとコピーでなく、ファイル名指定でのコピーとなったため、リスクが増えたというは誤認。</p> <p>パーティション単位の丸ごとコピーでなく、ディレクトリを選択してのコピーとなったため、大枠の転送量の確認ができなくなったためミスの発見が遅れた(運用をつづけながらの転送であることから転送する全部のファイルが確定するのが遅れた。また多量のファイル数であるからファイルリストを取るだけでも時間がかかった。一方パーティション単位でコピーする範囲を決めていけばdfコマンドだけですぐに判明する)。</p> <p>また、HSMの仕組み上どちらの方法でも最終的には、通常の様例えばcp -rでdirectory指定でサブディレクトリまで含めて簡単に転送することは不可能であり、ひとつひとつのファイルがどこのテープに収納されているかを確認しながらファイル毎の転送を行わなければならない。</p> <p>パーティション丸ごとのコピーができなかったのは、転送進捗が遅れたため、このため期限までに転送を終了するためには内容が同じファイル等を転送しないなどが必要だった。この方針については、転送の打ち合わせで話し合いながら決めていった。(20を参照)</p>	<p>時間的制約からテープの丸ごとコピーではなく、必要に応じた選択的コピーが選ばれた。</p> <p>判断は計算科学センターが行い、Belleグループの了承のもとに行われた。</p>	<p>予算を圧縮する必要があると判断されたため、リース契約を移行作業に十分な期間にわたり延長することができなかったため。</p> <p>Belle内で事前に議論することではなく、決定後の事後承諾となった。</p>	<p>我々としては旧B計算機と新共通計算機にデータ移行のためのオーバーラップ期間が必要であることは、常に主張していたが、それは旧B計算機の延長にかかる費用が高く(月単位の延長を会社が認めなかったため)、認められなかった。</p> <p>実際、データ移行が本当に重要だと考えているのなら、誰がそのような判断をしたのか、逆に知りたい。</p> <p>我々に与えられた選択肢は旧B計算機の契約延長をすることなく、かつデータをオーバーラップ期間がない新共通計算機に移行することであり、それを実現するためには報告書にあるような複雑な手順(旧B計算機のデータをまず旧共通計算機へ移行する。次に旧共通計算機のデータを新共通計算機に移行する、という手順である。)が必要となった。</p> <p>なお、この時には新共通計算機をどこの会社が落札するかも分かっておらず、この手順が許された金銭的、時間的制約内で最適であると考えた。</p> <p>しかし、この手順を決めた段階で、既に残された時間は限られており、ファイルを取捨選択してコピーする以外に、期限内に終わらせることはかなり難しいことになったと考える。</p>	該当しない。	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>
22. データ移行に掛ける予算が不十分であったと報告書で指摘されていますが、そのようなことになったのは何故ですか。また、誰が予算をどのように判断し、了承しましたか。	該当しない。	<p>Belle2建設の予算が不足であったため他の予算を削減するためと聞いている。</p> <p>B計算機データ移行に確保してあった予算を使わない事に決定判断したのは計算科学センター長。</p>	<p>計算機のレンタル契約を延長してオーバーラップさせ、無理のない方法でデータ転送を行うには1億円の費用が必要とされ、この拠出についてBelleから相談を受け、私(当時素核研副所長)からNK氏(当時所長)に相談の上、費用が掛からない方法を選択するように指示した。</p>	<p>予算を圧縮する必要があると判断されたため。</p> <p>後者の質問は「該当しない」。</p>	<p>該当しない。</p> <p>21にも述べたが誰がどういう基準でこれを決定したのか知りたい。</p>	該当しない。	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N 氏 (素粒子原子核研究所現場責任者)	M 氏 (計算科学センター現場責任者)	Y 氏 (素粒子原子核研究所長)	S 氏 (Belle collaboration代表者)	H 氏 (Belle部内事情に詳しい者)	SS 氏 (調査委員会委員)	K 氏 (協力業者 N 社)	T 氏 (協力業者 S 社)
23. 新中央計算機の運用とB計算機の運用期間にオーバーラップがなかったことが報告書で指摘されていますが、そのようになったのは何故ですか。また、誰がオーバーラップのない状態をどのように判断し、了承しましたか。	<p>オーバーラップがとれなかったのは、予算のせいだと認識している。</p> <p>また、B計算機の契約延長が月単位でできなかったのも関係していると認識している(結果論だが、B計算機の仕様書に月単位の契約延長を入れなかったのが問題だった)。</p> <p>予算の条件などを考えると、これしか解はなかったと思われる。以前から相談はあったが、基本的には計算科学センターでの決定事項として、私のところに伝えられた。この方法は、B計算機の運用に影響があり、またデータにアクセスできない時間も長くなるので、Belle側担当者の立場では好ましくないと考えた。しかし、データ損失につながることは予想しては、この方式で移行できそうに思えたので、予算の関係で他に方法がない事情を勘案して、了承した。</p> <p>Belleには報告して了承を得た。</p>	<p>理由は当初の予定を変更して予算を圧縮したため。オーバーラップがあるかどうかについての判断は計算科学センター長。</p>	<p>この理由はひとえに予算上の問題であり、前述22の回答がそのまま該当する。</p>	<p>予算を圧縮する必要があると判断されたため、リース契約を延長することができなかったため。</p> <p>後者の質問は「該当しない」。</p>	<p>21に回答しているのでそちらを参照。</p>	<p>該当しない。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>
24. 上記の「16」の質問事項に関係しますが、データ移行に与えられた期間が約7か月と短かったことが報告書で指摘されていますが、そのようになったのは何故ですか。また、誰がデータ移行の期間をどのように判断し、了承しましたか。	<p>予算の都合と思う。</p> <p>ぎりぎり移行が完了することなので、了承して、Belleにも報告した。</p>	<p>理由は予算削減のため。</p> <p>事前に転送試験を行い、その結果をもって、B計算機担当者、共通・中央計算機担当者の意見を聞いて計算科学センター長が判断した。</p> <p>しかしながら、実際には16に述べたように運用上 計算機資源が制限されたためテスト性能を大幅に下回る性能でしか転送できなかった。</p>	<p>震災からの復旧後、旧B計算機が撤去されるまでの期間でデータ移行を完了させるという境界条件から移行作業の期間が7か月とされた。</p> <p>この判断、了承は計算科学センターが行ったと考えるが、これは前述の予算の問題を除けば動かしがたい条件であり、事実上選択の余地がなかった。</p>	<p>予算を圧縮する必要があると判断されたため、リース契約を延長することができず、B計算機の契約期間中に旧共通計算機にデータを行こうすることになったため。</p> <p>後者の質問は「該当しない」。</p>	<p>21に回答しているのでそちらを参照。</p> <p>新中央計算機の運用とB計算機の運用期間にオーバーラップがなかったこと、データ移行に与えられた期間が約7か月しかなかったこと、は全て旧B計算機の延長を選択肢から外したことに起因している。</p>	<p>該当しない。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>
25. 計算機関係のエキスパートを十分確保できなかったのは何故ですか。また、誰が人数をどのように判断し、了承しましたか。	<p>わからないが、Belleの人手が不足しているのは関係していると思う。</p>	<p>SEが十分確保できなかったのは、今まで述べてきたとおり。</p>	<p>質問17の回答を参照されたい。</p>	<p>Belle実験に関連するBelleメンバーの作業の体制および人選は、共同代表者が責任を持って行なう。</p> <p>データ移行作業には、計算科学センターおよび業者の人も従事するのでBelle側からはコンピューティングマネージャーのN氏一人でエキスパートとして十分と判断した。</p>	<p>Belleグループ内の事を質問されているなら、17の回答を参照。</p> <p>計算科学センター内については該当しないので回答できない。</p> <p>業者に関しても該当しないので回答できない。</p>	<p>該当しない。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N 氏 (素粒子原子核研究所現場責任者)	M 氏 (計算科学センター現場責任者)	Y 氏 (素粒子原子核研究所長)	S 氏 (Belle collaboration代表者)	H 氏 (Belle部内事情に詳しい者)	SS 氏 (調査委員会委員)	K 氏 (協力業者 N 社)	T 氏 (協力業者 S 社)
(作業の管理責任に関して)					*			
26. 全体の進捗状況の管理、作業上での問題の調整等について、報告書の記載への補足等がありましたら申し述べてください。	特になし。	10月以降作業の進捗状況や作業上での問題の調整については、毎週の会合で調整していたことは前述していた通りであり、必要であれば会合の全資料を提出する。 しかしながら、この部分については、困難であったが完遂された。 問題があったのはファイルリスト作成過程。	なし。	Belle実験共同代表者として、結果として全体の進捗状況の把握や調整に関して認識や配慮が十分でなかったと反省している。	特筆すべき事はない。	該当しない。	回答する立場ではありませんので、回答を控えさせていただきます。	回答する立場ではありませんので、回答を控えさせていただきます。
27. 旧計算機にコピーした時に、データが全てコピーされているか確認しなかったと報告書に記載されていますが、確認しなかったのは何故ですか。また、誰がそのような作業形態をどのように判断し、了承しましたか。	移行作業の上では、ファイルリストにのっていたファイルについては旧計算機で確認されていた。この方針は、移行作業の打合せで決定された。それとは別の次元で、旧計算機で実際にユーザ等が利用する環境を準備すれば、この時点で問題が発見された可能性があるが、そもそも旧計算機にはBelleの利用者はいなかった。	手続的には、転送すべきディレクトリリストがBelle側より出され、作業SEからこれに対するファイルリストが提出され、Belle側が了承したと考えている。 この手続は一回だけでなく、変更などにより何回か繰り返され修正された。 ファイルリスト自体は膨大であるので、実際には指定のディレクトリが入っていることのみ確認され、ディレクトリ内のファイル全てが網羅されているかはチェックされなかったと、喪失発見後に聞いた。 ファイルリスト確認以降はこのファイルリストを基に作業が行われ、確認も行われた。	伝聞であるが、ファイルリストにあるファイルが正しくコピーされていることは複数回チェックされているものの、ファイルリストが正しいかどうかのチェックが行われていなかった。 ファイルリストを作成することはほぼ自明で正しく行われているに違いないとする思い込みが先行したものと思われる。 確認方法に関する判断、了承はBelleグループにおいて行われた。	調査の結果では、ファイルリストにあるデータが全てコピーされたことは確認されたが、旧共通計算機にBelleのソフトウェア環境がないため、実際にデータを読んで確認することができなかった。 作業形態の判断・了承は計算機センターの責任のもとにされたと理解している。	旧共通計算機にデータをコピーした際、旧B計算機で行われていたような解析できる環境を旧共通計算機上に構築し、実際にデータを読むなど努力をするべきだったと思う。 また解析環境を整えないまでも、ファイル数のチェックなど、コピーの正当性を確かめる手段はあった。 これをしなかったのは、こちらで用意したリストと業者の作成したリストの整合性は、業者もしくは計算科学センターでチェックされている、思いこんでいたことである。 またもしKEK内で人材が確保できないなら、他大学、機関に協力を依頼することも考えるべき事だったかも知れない。	該当しない。	回答する立場ではありませんので、回答を控えさせていただきます。	回答する立場ではありませんので、回答を控えさせていただきます。

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N氏 (素粒子原子核研究所現場責任者)	M氏 (計算科学センター現場責任者)	Y氏 (素粒子原子核研究所長)	S氏 (Belle collaboration代表者)	H氏 (Belle部内事情に詳しい者)	SS氏 (調査委員会委員)	K氏 (協力業者 N社)	T氏 (協力業者 S社)
(分散バックアップ体制の問題等に関して)					*			
28. raw data、all-dst、skim-dst等のデータバックアップを保持するためのBelleグループ内の責任・対応・報告体制について、その形成経緯を含めて申し述べてください。	<p>データのバックアップには以前から議論があったが、震災でKEKの計算機が長期間停止したことを契機として、物理データ(skim mdst)を他の研究所に分散させる活動が具体化した。</p> <p>Belle内では、DSTグループと呼ばれる計算機関係のグループが中心となって、各研究所の人にデータのコピーを依頼した。</p>	該当しない。	<p>すべてのデータについてバックアップを持つことは予算措置を含めて意図されていなかった。</p> <p>物理データ(Skim-mdst)については極めて重要であるとの認識があり、量が少ないこともあって国内外にコピーをおいていた。</p> <p>All-dstは一次的に作ったファイルであるとの考えで重要視していなかった。</p>	<p>2011年3月の大震災によりKEKのB計算機の稼働が長期にわたり制限されることになったため、物理解析に必要な物理データ、疑似データ、バックグラウンドデータを名古屋大学とPNNLの計算機にコピーすることを決め、共同代表者の指揮のもと体制を整えコピーを行った。</p> <p>作業は、物理コーディネータ、コンピューティングマネージャを作業責任者として数人のグループで行なった。</p> <p>作業は、共同代表者主催の執行委員会やグループに報告された。</p> <p>生データは、量が多いのでコピーする計画はなかった。</p> <p>それまでは、必要に応じて一部の物理データ等がコピーされていた(2010年に物理データのコピーについて執行委員会で示唆・議論がなされたが、機が熟さず実行には移されなかった)。</p>	<p>Belleグループ内でデータバックアップに対する責任の所在は決まっていないと認識している。</p> <p>ただし、all-dstやskim-dstを作成するDST productionグループおよびMC productionグループがあり(私がこのグループを統括していた)、このグループがデータを管理すると暗黙の内に思われているが、データバックアップの話など2010年の夏に私がグループミーティングで指摘するまで、考えられていなかったのではないかと思う。</p> <p>現在は地震の後に行った他機関へのデータのコピーにより、以前に比べて状況は改善されていると思うが、raw dataはデータ量が非常に大きく、他機関でのバックアップは、実際のネットワークスピード、記憶媒体を誰が調達するのか、などクリアしなくてはならない問題点が多く現実的ではない。</p> <p>また報告体制としては、仮に今回のような問題が発生したときには、まずBelle計算機責任者のN氏が中心となり行動を起こし、必要ならBelle II計算機責任者である私に連絡が来るであろう。</p>	該当しない。	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>	<p>回答する立場ではありませんので、回答を控えさせていただきます。</p>

〈別表〉 Belle実験データの一部損失に関するヒアリング調査シートの取りまとめ(2012年12月14日、同年12月19日にヒアリング実施)

	N 氏 (素粒子原子核研究所現場責任者)	M 氏 (計算科学センター現場責任者)	Y 氏 (素粒子原子核研究所長)	S 氏 (Belle collaboration代表者)	H 氏 (Belle部内事情に詳しい者)	SS 氏 (調査委員会委員)	K 氏 (協力業者 N 社)	T 氏 (協力業者 S 社)
29. 結果的に本事案における最大の痛手の一つとなった skim-dst のデータ一部損失に関して、そのバックアップ保持が完璧でなかった原因は何であると考えますか。	<p>問28の活動が始まった時には、データ移行が始まっており、データ移行に関連する部分が一ユーザから隔離されてしまい、一部のファイルはバックアップがとれなかった(実際には運用終了前にコピーするチャンスがあったが逃してしまった)。</p> <p>また、データのバックアップについてはコピー先の研究所の人に任せきりだったので、一部のバックアップが不完全だった。</p>	該当しない。	<p>PNNLにypipiというデータをコピーし終わる前に本件のデータ移行が始まってypipiのコピーが中断された。</p> <p>また、コピーし終わったデータにもファイル構造が壊れていたものがあって読み出し不可能となったものがあった。</p> <p>これらの原因には、PNNLにデータコピーを完全に任せてしまったこと、またPNNL側でも責任を持ってデータを受け取って管理する体制がなかったことがあげられる。</p>	<p>ypipi mdst は、他のmdstに比べて新しく作られたもので、HSMIに保存されたが一部が磁気ディスク上にコピーされなかった(データ移行作業のためHSMIへのアクセスが制限されていたためと思われる)。</p> <p>PNNL計算機へのコピーは磁気ディスクからなされたため、一部がコピーされなかったものと思われる。</p>	<p>* 物理解析用データである Ypipiデータが一部損失したことについて、データ移行前にお願いしていた他機関への Ypipiデータのコピーも不完全であったことが問題である。</p> <p>こちらは他機関の物理解析を行った経験のある研究者がデータコピーをしていたので、正確にコピーしてくれていると思っていた。例えばMC物理解析用データを各機関で作成してもらい、それをKEKに転送するときには、ファイル名や大きさだけでなく、md5というファイルの整合性を調べるツールを用いてファイルが壊れていないかどうか調べていた。</p> <p>こういう手順が標準だったので、他機関の研究者も同等の手順を踏んでいると思っていたが、認識が甘かったようである。</p> <p>なお、失われた Ypipiデータは、実際の解析には影響はほとんどないことを明記しておく。</p>	該当しない。	回答する立場ではありませんので、回答を控えさせていただきます。	回答する立場ではありませんので、回答を控えさせていただきます。