

データ分散保管技術を用いたがん登録データ 保全のための実証試験

三上 春夫*

1. 目的

長期にわたるがん登録データを安全かつ正確に次世代に継承することは長期統計の基盤的技術である。特に日本は地震国であり、大規模災害や火災等によるデータ消失への対策として広域分散保管の技術的要件について検討する必要がある。今回実施の実証試験においては、宮城県がん登録、千葉県がん登録、神奈川県がん登録の3カ所に個人情報をいっさい含まない試験データを用意し、お互い他の2地域のデータを部分的に保持した。災害を想定して1カ所のデータを削除し、他の2地域に保管したデータより再現した。データ復元の際のセキュリティコントロールはICカードにより使用許諾管理を行ったので報告する。

2. 方法

今回の分散保管実証試験においては遠隔地に位置する地域がん登録室が相互にデータの一部を持ち合う形態を想定した。単純に互いのバックアップを持ち合う場合には、特定がん登録実施県の患者データを他地域のがん登録室に渡してしまうという医療情報保護上の問題がある。そこでバックアップデータを決められたロジックによりビット単位で分割し、システムの機能を用いずにはデータの復元を不可能とし、その上でデータに冗長性を持たせて複数地域に分割保管する仕組みをとった。データ回線については実運用の際は光回線を用いることを想定しているが、今回は

テストデータであり、各所のインターネット回線を使用した。

3. 結果

1. データ保管、復元、整合性確認

想定したいずれのケースでも保管復元データの一致が確認された。

2. データ転送速度

分割・分散保管完了までの所要時間は、実施環境で10MBあたり約1分と、実用に十分耐えうる時間であった。また、ほぼ容量に比例して、分割・分散保管の所要時間が増加した。今後容量が増えた場合の所要時間も予測可能であった。保管データの世代管理においては、任意の順序で任意の世代へ復帰可能であった。

	1回目	2回目	3回目	4回目	5回目
保管	3'9"	2'39"	2'34"	2'51"	2'46"
復元	1'1"	1'1"	1'1"	測定せず	測定せず

100MByte (神奈川県がんセンター)

	1回目
保管	8'55"
復元	5'10"

300MByte (宮城県がんセンター)

	1回目
保管	25'11"秒
復元	19'57"秒

*千葉県がんセンター研究局がん予防センター
〒260-8717 千葉市中央区仁戸名町 666-2

4. 考察

バックアップ量の増大と遠隔地へのデータ搬送手段およびコストは相反する要件であるが、インターネットや高速通信網の発達によりこれらの要件は緩和されつつある。

データの分散保管では1カ所に全データを集中保管しないため、伝送経路と保管先の双方でデータが復号されることがない。復号は標準化された IC カードセキュリティに守られ、許諾権限を有するもののみが自地域以外のサイトで復元可能である。将来複数の地域の地域がん登録が参加し、相互に分散保管に参加した場合、地震などの災害に耐える安全なデータ保全が全国規模で可能となる。保管データのファイル形式に依存しない保存も本法の利点である。

5. 結語

長期統計としての地域がん登録データベースの継続性と耐障害性を高めるためにデータバックアップによるデータ保全の要求はますます高まることが予測される。さらに震災や火災といった大規模災害に対するデータ保全の手段が求められている。今回の実証試験を通じてデータの広域分散保管技術がセキュリティと耐障害性を両立した解決方法となる可能性を示した。