

# 日本周辺海域の反射断面データベースの復旧と代替システムの構築

## Recovery of database of seismic profile and backup system

佐藤智之<sup>1\*</sup>・岡村行信<sup>2</sup>・荒井晃作<sup>1</sup>

Tomoyuki Sato<sup>1\*</sup>, Yukinobu Okamura<sup>2</sup> and Kohsaku Arai<sup>1</sup>

**Abstract:** Database of seismic reflection profiles was damaged by the Great East Japan Earthquake on March 11, 2011. The database had been already recovered with replacing of hard disk drives and re-installing of operation system and database software. After the recovery, we compress the raw dataset to ease of backup. As a result, compressed law dataset become a substitute data-search system which is executable without a proprietary software on Linux, Mac OS X, and Cygwin on Windows.

**Keywords:** database, seismic profile, data compression

### 1. はじめに

産業技術総合研究所では 30 年以上にわたり日本周辺海域での反射法音波探査を実施し、その成果を海底地質図として公表してきたが、その膨大なデータセットの持つ可能性を活かすべくデータベースを作成して一元化を目指してきた(岡村ほか, 2009)。データベースはほぼ完成に近づき、シームレスなデータ閲覧、検索も容易になり、外部からのデータ提供依頼への対応も充分こなせるようになっていた(佐藤ほか, 2011)。しかし、2011 年 3 月 11 日の東日本大震災によりシステムが損傷してしまった。本報告では震災時の損傷状況とその後の復旧、その反省に立った代替システムの構築状況について報告する。

### 2. 東日本大震災によるデータベース被害状況

2011 年 3 月 11 日に発生した東北地方太平洋沖地震の際、職員は即座に屋外に避難したが、その時データベースは無停電装置に接続して通常起動していたものの作業はさせておらず、待機状態にあった。後日電源復旧後に被害状況を確認するとシステムの筐体が倒れたり、落下して物理的に大きく破損していることは

なかったが、おそらく停電によりシステムは停止していた。設置された無停電装置によって電力供給停止直後に正しい動作停止手続きが踏まれたかどうかは不明である。状況確認のためシステムを起動させると通常通りに動作し、いくつかの動作を試してみても特にエラーが発生しなかった。そのため、損傷無しと判断してそのままシステムを起動させたのだが、時間の経過とともにデータシステムの物理的母体であるワークステーションのオペレーションシステム(OS)である Red Hat Linux の起動すらしなくなった。

保守契約によるメンテナンスを受けた結果、ハード面では記録装置である内蔵ハードディスクドライブ(HDD)が損傷していることがわかった。内蔵 HDD は HDD4 台による RAID 5 (Redundant Arrays of Independent Disks 5) の構成をとっており、HDD1 台の損傷に耐える冗長性を持っていたが、今回は 2 台以上が同時に損傷したために復元不能に陥ってしまった。一旦起動した後に動作しなくなった経緯や原因は不明だが、起動状態で大きな余震を何度か被ったことによる HDD のヘッドかプラッタの損傷、あるいは本震による損傷が記録を徐々に壊したのだろうと想像される。他に RAID 5 構成の外部 HDD も接続してあったが、こちらは一年経過した現在も問題なく動作し

\* Correspondence

<sup>1</sup> 産業技術総合研究所 地質調査総合センター 地質情報研究部門 (AIST, Geological Survey of Japan, Institute of Geology and Geoinformation)

<sup>2</sup> 産業技術総合研究所 地質調査総合センター 活断層・地震研究センター (AIST, Geological Survey of Japan, Active Fault and Earthquake Research Center)

ている。

ソフト面では、OS およびデータベースソフトウェアが内蔵 HDD に記録されていたため失われた。整理して登録されていたデータセットは外付け HDD に記録されていたため失われなかったが、ソフトウェアの専用形式で複雑に関連付けされた形で保存されていたため、ソフトウェアが失われることで実質使用不可能となってしまった。

データベース登録元の音波探査デジタルデータおよび位置情報も外付け HDD に記録されていたため、幸い既存調査デジタルデータの散逸は免れた。

### 3. 復旧状況

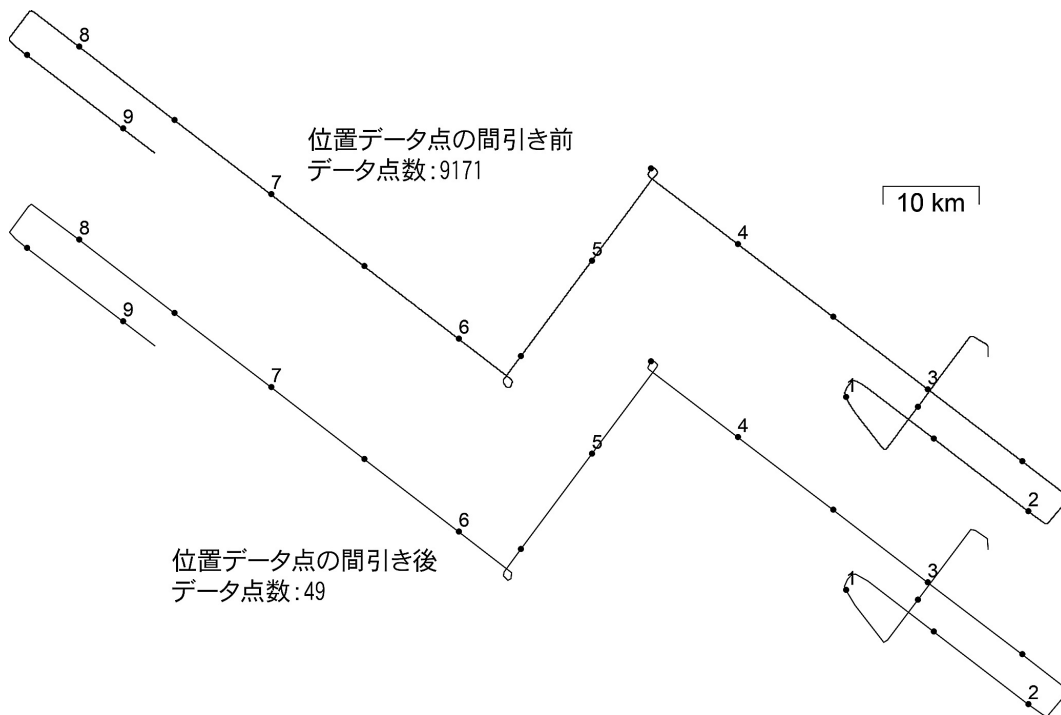
震災後の損傷状況確認の後、内蔵 HDD 4 台の交換と OS、データベースソフトウェアの再導入を行った。震災前までの作業でソフトウェアの不具合を複数発見しており、最新版ではそのいくつかが修正されているため、OS とソフトウェアは震災前のものではなく最新版を導入した。加えてこれまでの反省に立って生データを再整理してからデータベースシステムへ登録しなおした。2011 年度中に再登録まで全て完了し、震

災前と同等の状態へと復帰した。

### 4. 代替システムの構築

今回のような大震災は極めてまれではあるものの、災害への備えは必要である。むしろ災害時にこそ、過去の地下構造データを使用し現象の地質学的把握が必要となり、災害に強い、あるいは復旧しやすいシステムが望まれる。今回システムが失われたのは、母体であるワークステーションが損傷したからである。システムに採用されていた RAID 構成による冗長性は HDD 単体ごとの突発的な故障には有効であるが、RAID 全体に同時に影響を及ぼす地震、火災や停電に対してはそれほど有効ではないことがわかった。また、データセットがデータベースソフトウェアの専用フォーマットにて保存されており、それに依存していたことも復旧の妨げとなった。

利便性の高いデータベースを利用する以上、要求される専用フォーマットを利用するのは不可避としても、その一台に 30 年以上にわたる調査結果全てを委ねるのは危険である。将来的には専用フォーマットが古くなって利用できなくなる可能性も考えねばならな



第 1 図 ダグラス・ポーカーアルゴリズム適用前後の位置情報精度の比較。測線 gh1002 を例に間引き前後の測線を上下にずらして示す。位置情報のデータ容量は 1% 以下に圧縮されているが、直線部分はもちろん、方向転換時の円軌跡もデータ間引き後でも残っている。計算上の最大誤差は数 m である。測線上の黒点は反射点 500 毎の位置、数字は反射点番号の千の桁を示す。

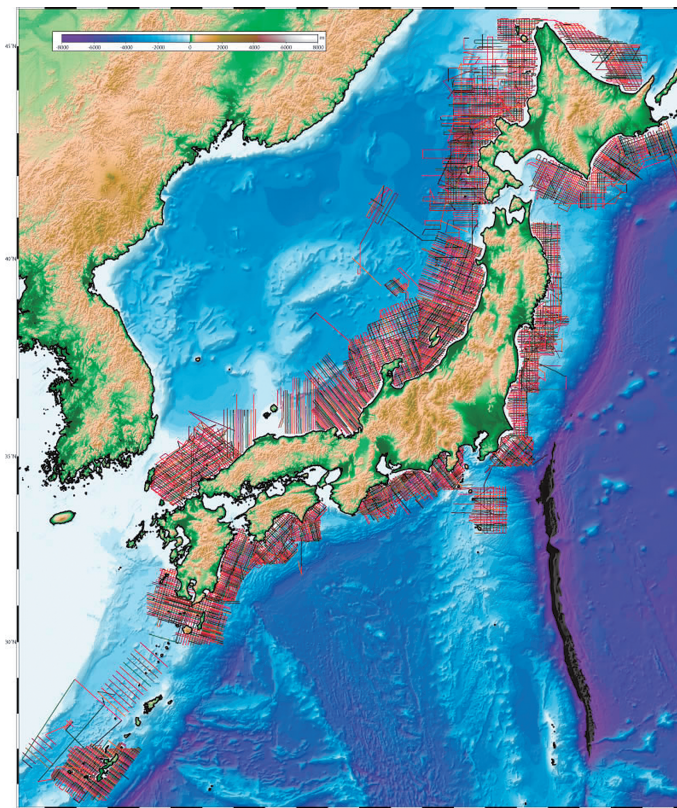
Fig.1 The line maps of seismic profiles before (upper) and after (lower) the compression of the position data with douglas peucker algorithm. Example line is gh1002.

らず、現状では一般的かつ規格が公開されたフォーマットである SEG-Y 形式およびテキスト形式の状態でも整理されたデータセットを保持している必要がある。この状態ならば外付け HDD や別の計算機にバックアップするだけでなく、一般的なソフトウェアを使えばバックアップ先にてデータを利用することも可能である。とはいえ、日本周辺海域を覆う膨大なデータであるため、データ量の圧縮と簡易な検索システムの作成が必要であると考えた。

まず、データ量圧縮のため位置情報の間引きを行った。従来生データは共通反射点ごとの緯度経度を全てテキスト形式で羅列していたが、データ量が大きくて取り回しも悪く、検索時やソフトウェア（例えば GMT など）を利用して地図上に測線を描画させる際にも不便であった。簡易な作業用に 100 反射点ごとに間引いたデータを利用することもあったが、精度の維持とデータ量圧縮のバランスが悪かった。そこで折れ線データ点の間引き法であるダグラス・ポーカーアルゴリズム (Douglas and Peucker, 1973) を用いてデータ点の間引きを行った。アルゴリズムの詳細は省略するが、間引き前の全点と間引き後の点列を結んだ

折れ線との距離が指定した値以内に収まるようにデータ点を間引くことができるアルゴリズムである。今回はこれを元に 500 反射点ごとのデータ点が必ず残るようにアルゴリズムを改造して用いた。これにより断面観察時に有用な要所ごとの座標が明確に残り、解釈しやすくなる。全測線の位置データに対し、まず測地系を全て WGS84 に変換し、最大誤差距離 0.001 度 (数 m 程度、緯度と経度で単位長が異なるため厳密ではない) で改造したダグラス・ポーカーアルゴリズムを適用した。その結果、全位置データ容量が 319MB から 1.8MB となり、1% 未満まで圧縮できた。本システムで対象としている反射法音波探査の場合ほとんどの区間が直線であるために非常に有効に作用したようだ。位置精度についてもメートルオーダーで確保されているため (第 1 図)、測量誤差を考えれば実用的と考える。また、圧縮前データも合わせて保存してあるため、より厳密性を期したい場合はそちらを使えばよい。

この位置データ圧縮により、検索や作図が比較的容易となった。位置データには測線名と航海名も入っているため、1.8MB の位置情報ファイルをテキストエ



第 2 図 データベース管理されている反射法音波探査デジタルデータの 4000 本を越える全測線図。1.8MB まで圧縮した位置情報のテキストデータを元に作図ソフトウェア (GMT) で作成した。測線ごとに色を変えてある。断面の表示はできないが、データベースシステムを介さずに測線の検索までは可能である。

Fig.2 Line map of the all seismic lines. This map was made based on compressed navigation data without database software as a substitute data-search system.

ディタや表計算ソフト（MS-Excel など）で読み込むことで、日本周辺全海域のデータセットから任意の測線やある区画内の緯度経度を持つデータを検索することも実用的になり、測線図を描画するのも容易となった（第 2 図）。実務用にシェルスクリプト（Linux や Mac OS X に加え、Windows に導入可能な Cygwin で標準搭載されている Bash 向け）を作成し、2 点の座標を与えればメルカトル図法上でそれを対角線とする長方形に含まれる測線を自動的にリストアップすると同時に GMT によって測線図が作図されるようにし、データセットに付属してある。専属のデータベースシステムには劣るものの、数 MB のデータとフリーソフトウェア上で動作するシェルスクリプトのみで目的海域のデータセットがすぐに把握できるようになった。これに容量数百 GB の音波探査データを加えれば、測線図から断面図を直接参照できないといった制約があるものの、既存データ全てを網羅して利用することがさしあたり可能である。これは物理面、ライセンス面から特定の機材、計算機環境に依存しないため、バックアップや代替システムの立ち上げが容易であり、万が一データベースが破損してもデータ参照が可能である。

## 5. まとめ

2011 年 3 月 11 日の東日本大震災により日本周辺海域の反射法音波探査データベースは損傷したものの、生の調査デジタルデータセットは残っていたためにシステムの入替えと再登録により復旧することができた。さらにこれまでの反省に立ち、位置情報の測地系を WGS84 に統一するなど、生データの整理を行った。位置情報のデータ量圧縮を行い簡易検索スクリプトを付属させることでバックアップを容易にし、さらにバックアップ先でも計算機環境を選ばずにある程度のデータ検索ができるようになった。

## 文献

- Douglas, D.H. and Peucker, T.K. (1973) Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. *The Canadian Cartographer*, 10, 112-122.
- 岡村行信・辻野 匠・荒井晃作・井上卓彦 (2009) 海域反射断面データベースの構築. 平成 20 年

度沿岸域の地質・活断層調査研究報告, 地質調査総合センター速報, 49, 141-145.

- 佐藤智之・岡村行信・井上卓彦・荒井晃作 (2011) 日本周辺海域の反射断面データベース構築. 平成 22 年度沿岸域の地質・活断層調査研究報告, 地質情報総合センター速報, 56, 205-208.