

# 考古学情報の再現可能性

## ～バージョン管理システム Git を利用した調査データの管理と公開～

石井淳平

### はじめに

発掘調査記録をオープンデータとして活用するための方策として、GitHub に代表されるバージョン管理 + 共有サービスを利用した考古学情報の管理と共有について考える。図書として刊行された発掘調査報告書は、情報の固定という面で優れているが、報告書作成に利用された調査記録や中間成果物へのアクセシビリティが低く、再現性が低いことが問題である。このことは、考古学情報の真正性の問題ともかかわるとともに、考古学情報の社会的利用の根幹に関わる問題である。こうした課題を解決する手法としてバージョン管理システム Git の利用を提案する。

Git を使用し、一次情報である現場図面や写真、最終的な報告書に至る中間成果物すべての改変履歴を記録することで考古学情報の真正性と再現可能性が担保される。また、Git の機能を利用したウェブサービスである GitHub により考古学情報の作成に関わった調査記録の公開が実現する。本報告では、北海道南部の箱館戦争遺跡の調査において Git によるバージョン管理と GitHub を利用したデータ共有の実践例を紹介し、手法やノウハウを報告する。

### 1. ブラックボックス化する考古学情報

考古学情報の保存・公開に用いられる媒体については、紙でも PDF でも情報の再現可能性としては同質である。考古学情報の公開にあたって解決しなければならないのは発掘調査報告のブラックボックス化である。筆者は考古学情報を再利用可能なデータ群として流通させることが考古学における情報化の最大の課題と捉えている。

発掘調査報告書の作成には現地で作成した図面、撮影された写真、測量記録などの一次情報のほかに、それらを整理・清書した中間的な二次情報がある。

例えば、最終的に報告書に掲載される遺物集計表は遺物の整理分類に使用された「遺物台帳」を元に作成されている。しかし、最終的な報告書しか手にすることができない者にはこれらの集計作業が適切に行われたかどうかを判断するすべはない。

また、現地で作成した平面図と断面図には必ず誤差が生じる。多くの場合、「素図」や「第二原図」などと呼ばれる清書図面を作成する際に、原図の修正作業が行われる。どのような修正が行われたのか、その修正は適切だったのかを第三者が知ることはできない。

こうした「修正」作業は限度を超えると「捏造」と変わらない。測量では誤差を全体に散らして、総合的な誤差を「閉合誤差」のように定量化するが、たとえば遺構計測の際にそのような取り扱いがなされているケースを筆者は知らない。修正の履歴を残すことが考古学情報の再現可能性を高めることにつながる。

### 2. 考古学情報の「真正性」

要するに、考古学情報の再現可能性を高めることは、考古学情報の「真正性」を高めることにほかならない。考古学情報の「真正性」を文化財における「真正性 (authenticity)」に引きつけて表現すると次のとおりである。

- 考古学情報の価値は真正性によって評価される
- そのためには考古学情報の作成に使用された情報源が信頼できる度合いが評価されなければならない
- どのようなプロセスで考古学情報が作成されたのかに関する情報が真正性の評価には必須である

つまり考古学情報作成プロセスを明示することが、考古学情報の真正性を評価するための第一歩となる。どのように調査記録が選択され、どのような処理を経て最終的な考古学情報として成立したのかという点に関するメタデータの存在こそが考古学情報の真正性を評価する基本となる。

### 3. データの長期保存とともに考えるべきこと

「真正性」という視点で考古学情報を考えたとき、単なる調査記録の長期保存を考えるだけでは不十分だということになる。真正性を評価するためには、より上位の情報として「考古学情報作成プロセスとしての調査記録の新設・追加・変更の履歴」が必要となる。これは考古学情報の再現にかかるメタ情報であり、第三者が報告書を再構築できる情報であることが求められる。

### 4. Git を利用した発掘調査記録管理

筆者がイメージする発掘調査記録管理のイメージにもっとも近いのは Git を使ったバージョン管理と GitHub によるデータ共有である。Git の適切な使用によって、「どのような変更が」、「いつ」、「誰によって」行われたのかが明示的になる。GitHub のようなデータ共有環境を利用することで、Git によってバージョン管理されたデータ群を、誰でも利用できるようになる。

Git 本体の主な機能は情報の変更単位<sup>1)</sup>ごとに履歴を作成するものである。任意のタイミングで変更を Git に登録することができる。Git による変更履歴の登録には次のようなメリットがある。

いつ、どのような変更が、誰によって行われたのかが明確になる。報告書作成作業では、現場図面をトレースして掲載図版を作成する。このとき、調査者によって不要と判断された記録は掲載図版には転記されない。また、現場図面には記載のない調査担当者の野帳の情報が掲載図版に転記されることもある。上面写真の情報を掲載図版に付記する場合もある。最終的な成果に記載された情報が、どの時点で、誰によって付加されたのかが Git によって明確になる。

過去に遡って記録を回収できる。報告書作成作業では日々の作業で次々と記録が更新されていく。通常、上書きされ、消去された記録を復元することは難しい。仮にバックアップ機能などで記録の復元を行ったとしても、復元された記録が、変更履歴全体

の中で整合性を保つことは難しくなる。Git では変更単位ごとに全ての変更が記録されているため、消去、改変された過去の記録を復元し、取り出すことが可能となる。このことは、記録を「念の為に保存しておく」という行為を不要にし、最終的な考古学情報にノイズが含まれることを防ぐとともに、中間成果も維持されるため、結果的に考古学情報作成にかかわる全ての記録が保存されることとなる。

### 5. Git 利用の実際

Git はコマンドラインで操作されることが多く、筆者も基本的にはコマンドラインで Git を操作している。原稿執筆環境がテキストエディタ中心であることから、別途 Git 用のアプリケーションを起動することなく執筆作業に集中できること、日常的に使用する Git コマンドがシンプルで暗記しやすいためである。

ただし、Git には秀逸な GUI クライアントソフトが数多く開発されており (GitKraken、GitCola、SourceTree など)、これらのソフトウェアにより Git を操作することが可能である。筆者はシンプルな GitCola をインストールしている。GUI クライアントソフトを利用することでコマンド入力から解放されること、変更履歴 (commit という) や変更の分岐 (branch) や結合 (merge) が視覚的に一覧できるメリットがある。

Git での基本的な作業は「commit」と呼ばれる変更履歴の登録である。特定のディレクトリ (例えば発掘調査報告書作成用のディレクトリ) に対して行われた変更の全てが「commit」を行うことで Git のリポジトリ (データベースのようなもの) に登録される。「commit」の際に適切なコメント (「F2 塹壕図のグリッドラインを描画し、断面を修正」など) を付すことで変更の内容が第三者にもわかるようになる。

GitHub などのウェブでの共有サービスを利用している場合は「Push」という操作が必要になる。これは最新の変更を GitHub に反映させるためのものである。

別の編集者が GitHub にあるデータに対して変更を加えた場合には「Pull」という操作で変更点をダウンロードし、自分のコンピュータと GitHub のデータを同期する。

以上をまとめると Git 及び GitHub を利用したデータ共有は「Pull」コマンドで GitHub と自分の

## 6. 北海道北斗市二股台場調査での Git 利用

筆者らが取り組んでいる北海道南西部北斗市二股台場の測量調査では、GitHub による調査記録の公開(<https://github.com/IshiiJunpei/Futamata>)を行っている。

本調査は「箱館戦争遺跡プロジェクト」という有志が主体となって実施しているものであり、関係者が日常的に顔をあわせて作業する環境にはない。そのため、調査情報をリモートで共有する必要があったこと、複数の調査者が調査記録の編集に関わることから、作業重複を避ける工夫が必要であった。調査では Git と Git の仕組みを利用した GitHub というウェブサービスを利用して調査記録更新にかかる公正性の確保と作業重複の回避を行った。

複数の調査者で調査記録を共有する際に重要なことは適切なディレクトリ構造とファイル名付与である。考古学記録のアーカイブ方法については愛知県埋蔵文化財センターの電子納品ガイドラインが先行的な取り組みといえる（堀木 2018）。二股台場調査では筆者の扱いやすいディレクトリ構造やファイル名付与を行ったが、標準化されるべきことがらの一つであろう。

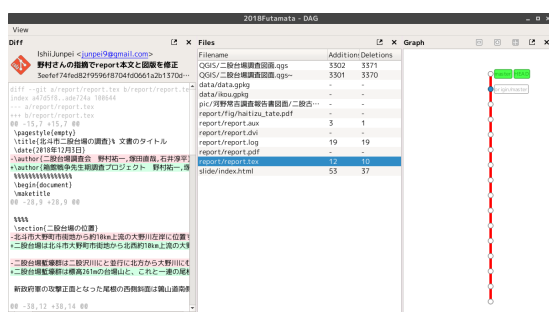


図 1 二股台場調査の Git 情報

図1は二股台場調査記録のログである。右のペインはGitの更新履歴を示す。中央ペインは変更が行われたファイルの一覧である。変更が行われたファイルを選択すると左ペインにファイルの内容と変更箇所が示される。共同調査者からの指摘で報告文

本文の修正と図版データの修正を行った変更ログである。このように全ての変更履歴が変更後の差分とあわせて確認できることが Git を利用する大きなメリットとなる。

## 7. コンパイル状態の考古学情報

図書として刊行された考古学情報は一種のコンパイル状態といえる。「コンパイル」とは、プログラミングコードをコンピュータが読むための2進法の記号に変換することであり、コンパイルされたプログラムは人間が読むことはできない。コンパイル状態の情報は真偽を問うことも、ソースを再利用することも不可能である。

紙媒体やPDF保存された発掘調査報告書も中身の情報の真偽を問うことやデータを再利用することが不可能という点でコンパイルされた状態といえる。コンパイル状態の報告書からは最終的な結果を知ることができても、元になった調査記録や中間成果物などの再利用可能なデータを取り出すことは不可能である。

## 8. オープンデータとしての考古学情報

考古学情報が適切なデータを使用して作成されているか、適切な手順を踏んで作成されているかを判断するためには、ソースコードに相当する調査記録や中間成果物にアクセスできる環境が必要となる。考古学情報の真正性は「情報の固定化（コンパイル化）」ではなく、すべての情報に市民がアクセスできる環境の構築によって担保される、というのが筆者の考え方だ。

オープンソースのソフトウェアのソースコードから必要な部分を切り取って自由に利用できるようになったことで、コンピュータソフトウェアの開発が大きく進んだのと同じように、考古学情報にかかわる調査記録や中間成果物を自由に利用できるようになる学問上のメリットははかりしれない。

全国遺跡報告総覧への報告書アップロードも十分とは言えない状態であり、筆者のアイディアは時期尚早といえるかもしれないが、自らの研究環境を向上させるための工夫を続けるという地道な取り組み以外に変化を起こす方法はないと信じている。

### 引用参考文献

堀木真美子 2018「発掘調査データの整理と公開システム」『文化財の壺』Vol.6, 文化財方法論研究会, pp.8-11

