Geological Survey of Canada

CURRENT RESEARCH
2005-H1

Acquiring reusable location data for stratigraphic studies: insights from the Central Foreland NATMAP Project

Joanna L. Palmer, Robert B. MacNaughton, Larry E. MacDonald, and Larry S. Lane

2005

CURRENT RESEARCH

Natural Resources Canada
Ressources naturelles Canada

Canada

ISSN 1701-4387
Catalogue No. M44-2005/H1E-PDF
ISBN 0-662-39365-1

A copy of this publication is also available for reference by depository libraries across Canada through access to the Depository Services Program's website at http://dsp-psd.pwgsc.gc.ca

A free digital download of this publication is available from the Geological Survey of Canada Bookstore web site:

http://gsc.nrcan.gc.ca/bookstore/

Click on “Free Download”.

Toll-free (Canada and U.S.A.): 1-888-252-4301

Authors' addresses

Joanna L. Palmer*
Department of Geography, University of Calgary
2500 University Drive NW, Calgary, Alberta T2N 1N4

****Current address:***
323 Valencia Place NW
Calgary, Alberta T3A 2C1

Robert B. MacNaughton (romacnau@nrcan.gc.ca)
Larry E. MacDonald (lamacdon@nrcan.gc.ca)
Larry S. Lane (llane@nrcan.gc.ca)
Geological Survey of Canada
3303-33rd Street NW, Calgary, Alberta T2L 2A7

Publication approved by GSC Calgary

Original manuscript submitted: 2005-01-11

Final version approved for publication: 2005-01-28

Correction date:

Acquiring reusable location data for stratigraphic studies: insights from the Central Foreland NATMAP Project

Joanna L. Palmer, Robert B. MacNaughton, Larry E. MacDonald, and Larry S. Lane

Palmer, J.L., MacNaughton, R.B., MacDonald, L.E., and Lane, L.S., 2005: Acquiring reusable location data for stratigraphic studies: insights from the Central Foreland NATMAP Project; Geological Survey of Canada, Current Research 2005-H1, 6 p.

Abstract: Stratigraphic data that can be used by subsequent investigators without the need to revisit a site are reusable data. Accurate location data are a crucial part of ensuring reusability of data, and thus contribute to the reproducibility of scientific studies and to cost savings. A review of some outcrop stratigraphic location data collected during the GSC's Central Foreland NATMAP Project found several categories of data errors that hamper reusability. These include errors in recording map datum, inaccurate co-ordinates, typographic errors introduced into co-ordinates, and errors in map presentation. Project leaders should ensure that all project participants who study stratigraphic sections are aware of the benefits of reusability and that participants follow specific criteria for recording location data; if required, they should provide participants with training in both traditional map craft and emerging technologies. A short time invested in capturing accurate location data on the section translates directly into significant time savings during report preparation.

Résumé : Les données stratigraphiques que des chercheurs peuvent utiliser sans avoir à retourner au site sont des données réutilisables. Il est essentiel que les données géographiques soient exactes pour être réutilisables et permettre ainsi de reproduire les études scientifiques et de réduire les coûts. Un examen de certaines données stratigraphiques issues de l'étude des affleurements réalisée dans le cadre du Projet de l'avant-pays central du CARTNAT de la CGC a mis en évidence plusieurs catégories d'erreurs géographiques qui empêchent la réutilisation des données, notamment les suivantes : enregistrement erroné du système de référence géodésique, inexactitude des coordonnées, présence d'erreurs typographiques dans les coordonnées et présentation cartographique erronée. Les chefs de projet doivent veiller à ce que tous les participants à un projet qui étudient les coupes stratigraphiques connaissent les avantages de la réutilisabilité des données et enregistrent les données géographiques selon des critères précis. En outre, ils doivent, au besoin, leur donner une formation sur les techniques traditionnelles et nouvelles utilisées pour produire les cartes. Le fait de consacrer un peu de temps à la collecte de données géographiques exactes sur les coupes permet de gagner beaucoup de temps à l'étape de la préparation des rapports.

INTRODUCTION

In stratigraphy, recognizing and interpreting spatial variations in rock successions is of central concern, regardless of whether a study's focus is biostratigraphic, allostratigraphic, lithostratigraphic, chemostratigraphic, or sequence stratigraphic. Thus, concern with the accuracy, reliability, and reusability of spatial data, in particular the location of stratigraphic sections, should be of fundamental importance. Obtaining and providing accurate location data for stratigraphic sections enhance the accessibility and reusability of the resulting stratigraphic data. In this paper, we review the necessity for high-quality spatial data in scientific studies and present a discussion of issues encountered with stratigraphic location data collected as part of the GSC's Central Foreland NATMAP Project. With this background, we suggest ways to improve the quality of stratigraphic location data in future studies. This report is not a critique of the quality of the stratigraphic work conducted during the Central Foreland NATMAP Project, but rather a summary of what we have learned in adapting existing data-collection methods to new, digital data-handling methods.

THE NECESSITY FOR ACCURATE LOCATION DATA

Background

For any field-based research, the necessity for accurate location data is obvious and self-evident. Some older textbooks on field geology do provide detailed discussions of this (e.g. Greenly and Williams, 1930; Himus and Sweeting, 1972). However, newer authoritative works have less to say on the topic (e.g. Compton, 1985; Salvador, 1994). In the stratigraphic literature, location data commonly consist of highly generalized location maps presented at scales too small to indicate accurately where sections were examined. This is not adequate, now that technological developments such as Global Positioning Systems (GPS) have greatly increased the attainable accuracy of spatial data. However, it seems that new methods for obtaining, managing, and publishing spatial data are only slowly being incorporated into stratigraphic studies.

We define 'reusable data' as data that are sufficiently detailed and accurate to allow subsequent workers to use them while referring only to published or archived data sources, without the necessity of revisiting a site. Reusable location data permit subsequent workers to accurately locate a given measured stratigraphic section that is of interest to them, whether for purposes of re-examining the actual section, incorporating the section into a database for analysis or map compilation, or using the derived stratigraphic information for follow-up studies. The accurate recording and clear communication of location data help to ensure high standards of reusability for data and information.

Reproducibility of data

Maintaining high standards of data reusability ensures the reproducibility, and thus the good repute, of scientific studies. In outcrop-based stratigraphy, the need for replication requires that later workers be able, in principle, to revisit the localities documented in a given study. Accurate location data that are clearly communicated are the underpinning of such replication. Inasmuch as the presentation of geological data can be strongly influenced by current theoretical understandings (cf. Harrison, 1963; Miall and Miall, 2001), the ability to revisit studied sites also provides a counterbalance against any tendency to tailor data collection to the template of a prevailing model.

The absence of adequate location data may raise questions about the reliability of a study. An extreme example is provided by V.J. Gupta, who, over a quarter century, reported numerous fraudulent Himalayan fossil occurrences, severely confusing the region's geological history (Talent, 1989). A hallmark of Gupta's publications was utterly inadequate location data: "One of (Gupta's) 'localities'...was 130 kilometres or so of Himalayan valley walls." (Talent, 1989, p. 614).

Cost effectiveness

In general, studies should be conducted and reported in a fashion that provides optimum value. The collection of complete, accurate location information can lead to cost savings over both short and long terms. Particularly in isolated regions where access costs are high, time not spent searching for a poorly recorded site translates directly into money saved. Where costs are prohibitive or funding is unavailable, archival data may be the only recourse. This is also true when a site has been destroyed or otherwise rendered inaccessible (e.g. important stratigraphic locations in British Columbia were inundated following construction of the W.A.C. Bennett dam). Also, in ecologically sensitive areas, accurate location data can give a worker the option of using archival data, rather than revisiting a site and thus increasing a project's environmental impact.

In addition, one worker's data may now be used by subsequent workers at scales other than that of the original work. If location data are collected to the highest accuracy possible, they can be applied to studies across a greater range of scales. At one time a case could be made that "...for any map no greater degree of precision is desirable or necessary than is commensurate with the scale of the map..." and that "... no set of data need be charted in the field by a method more precise than the methods applied in charting the other sets of data..." (Lahee, 1961, p. 445). However, because map data can now be displayed and scaled readily on computer screens, the map scale has ultimately lost its significance (Cheng, 2003). High-quality location data permit projects conducted at a gross scale to contribute to later projects conducted at much finer scales, which can be more cost-effective than re-examining a site.

A CASE STUDY ON THE QUALITY OF STRATIGRAPHIC LOCATION DATA

Reviewing section location data from the Central Foreland NATMAP Project

The Central Foreland NATMAP Project was a collaborative mapping initiative that focused on bedrock and surficial geological mapping in the Foothills of the northern Rocky Mountains and the adjacent southern Franklin Mountains (Lane et al., 1999; Lane, 2004). As part of the project, numerous stratigraphic studies were published as theses or journal papers. Two GSC bulletins compiling detailed stratigraphic sections measured during the project, as well as some existing but previously unpublished data from the project area, are in preparation. In addition, section metadata are tabulated in a project database to facilitate the inclusion of stratigraphic sections on project maps.

We have reviewed the location data from outcrop stratigraphic sections measured by eight stratigraphers involved in the Central Foreland NATMAP Project. The stratigraphers whose location data are considered in this report include GSC staff scientists as well as scientists associated with universities and the private sector. Each stratigrapher submitted data pertaining to a number of measured stratigraphic sections. Location data could be submitted in various formats and the choice of which format(s) to use was left to the stratigrapher. Formats in which data and information were initially supplied included:

- National Topographic System (NTS) map sheets with hand-drawn section traces;
- (aerial) photographs with hand-drawn section traces;
- top and/or base co-ordinates for sections;
- single co-ordinates (for short or nearly vertical sections); and
- section location descriptions in text form.

Most stratigraphers submitted a combination of these formats. Data and information were submitted as unpublished data, as published reports and theses, and as combinations thereof. These raw data were then used to produce digitized section traces and to populate database tables of location information. The conversion of data from hard-copy to digital formats required significant time allocations by part-time and term employees, generally students. Whenever possible, stratigraphers were asked to check the resulting digital plots of sections. When dealing with data submitted by stratigraphers located elsewhere in Canada, the checking of plots as well as the clarification of any problems identified in the submitted data were sources of significant delays, in some cases amounting to several months.

Categories of data errors and their implications

Four main types of error were identified in the location data. Each has implications for the data quality of any resultant study.

1) Datum errors: Six of the eight stratigraphers initially did not provide the horizontal datum (NAD27 or NAD83) in which section location data were recorded. In two cases, the stratigrapher had kept no record of which datum was used, and one stratigrapher apparently was unaware of the existence of different map datums. Because two of the stratigraphers were unavailable for follow-up, determining which datum had been used required a technician to plot the stratigraphers' data points using both map datums. In our project areas, datum error results in locational inaccuracy of approximately 0.25 km. When the map datum is not recorded, it is particularly difficult to rule out inadequate field procedures, such as unfamiliarity with the use of GPS receivers, as a source of error.

2) Co-ordinate errors: This error category includes all cases in which there was apparently a significant error in recording section location data in the field. In some cases, submitted co-ordinates were markedly at odds with written descriptions or photographic documentation of section locations. In other instances, stratigraphers who worked together on a given section provided conflicting co-ordinates for that section, with locations diverging by up to 1.5 km. Unless independent means of assessing location are available, major errors in recording co-ordinates can vitiate the reusability of location data.

3) Typographic errors in co-ordinates: Errors in typing (unpublished data) and typesetting (published data) were a significant source of incorrect co-ordinate and elevation data. They were recognized in five data sets; most were evidently typographic errors in the presentation of SI units and co-ordinates. Such errors may have originated with the original workers or been introduced into database tables during data entry. This type of error generally affected one digit in a set of co-ordinates. Although such errors are seemingly minor, their identification and correction can require hours of effort during preparation of data for publication. Even less desirably, it may be necessary for the end user to try to correct them. Most seriously, such errors, if uncorrected, potentially can render spatial data unusable.

4) Errors in section tracings and maps: In some cases, the maps on which section traces had been recorded were sources of error. Problems included:

- errors in maps that were drafted by the stratigrapher;
- section labelling placed so as to obscure or cover important map features;
- paper maps with section traces shown using inappropriately heavy lines; and
- section traces provided on scanned or photocopied portions of maps, without reference co-ordinates.

In terms of spatial accuracy, map errors generally have less serious effects than typographic errors and are more readily corrected. For example, a submitted report included a recurring overview map that was plotted on a base map on which NTS map designations were numbered incorrectly. However, if such errors are not corrected during the review or editing process, they will be a source of confusion and annoyance to the end-users of published reports.

Discussion

The first two error types noted above have become significant only recently, since GPS positioning, multiple datums, and scale-independent digital manipulation of data sets have become more prevalent. In contrast, the latter two error types are more 'traditional' and require diligent quality control to identify and avoid. However, because digital databases allow such a rapid, broad distribution of data, it is even more important to avoid such errors because they can propagate repeatedly by incorporation into successor databases.

The first principle in the publication of quality location data is the verification of source observations, which must be followed by maintenance of error-free procedures during the data-management process. This can only be ensured if the original data collector participates in the verification process and if assisting geologists and technicians are diligent in their handling of data. Not only must co-ordinate values plot consistently with section traces, but the section information must be evaluated with respect to how well the position of the section trace agrees with other geological data for the area. (For example, it is not sufficient to resolve typographical errors in co-ordinate values so that they fall within a bold, inked line on a small-scale map.) This evaluation provides an overall confidence in the quality of the stratigraphic section location data, which determines whether the data are reusable or not.

As noted above, the scientific value and reputation of a report can be severely reduced by poor location data. We have found that a small time investment in accurately recording and verifying location data while on the section, combined with careful recording of the lineage of a data set, dramatically reduces the time required to validate location accuracy after the fact.

RECOMMENDATIONS

Recommendations for project planning

Our experience convinces us that it cannot be assumed that geologists are aware of what constitutes adequate location data, the necessity for it, or the correct methods by which to obtain it. Project leaders should ensure that all project staff who may be called upon to measure outcrop stratigraphic sections comprehend these issues. Additionally, project leaders should ensure that all participants are trained in map craft and in the use of any digital technology used for collecting location data (e.g. GPS, computer applicatons).

Basic protocols for section measuring, especially for recording of location data, should be agreed upon and established for a project before any field work is undertaken. Excessive regimentation of section-measuring procedure is undesirable, in view of the highly varied scales at which sections may be measured and the variety of recording methods commonly used by stratigraphers (e.g. written section descriptions, coded section-logging forms, graphic logging). However, protocols for recording section location data should be stringent. It may be helpful to provide project staff with standardized forms (hard copy or digital) upon which key section location data can be recorded. Copies of the completed forms, together with supporting maps and aerial photographs, should be submitted to the project leaders for archiving at the end of each field season.

Recommendations for data recording

The impact of Web communications, corporate databases, and data sharing among scientists requires a conscientious effort to collect data adequate for these purposes. Staff responsible for field data collection should be aware of and use modern as well as traditional methods for obtaining locations and accurately identifying them. Traditional methods of pin-pricking aerial photographs and using a hard lead pencil to place a dot on a map can be augmented through the use of GPS receivers and small computers. Portable computers, adapted to provide forms for database population, interface with a GPS receiver, and display digital map bases and orthorectified aerial photographs, offer potential to increase the completeness and accuracy of field data. If such tools are to be applied successfully, they will require some modernization of geologists' map-craft skills. However, fully adequate results can be obtained by the conscientious use of traditional methods, augmented by a GPS receiver, and the following comments focus on this approach. Whatever the methodology, any time spent in the field ensuring that location data are accurate and thorough will translate directly into time saved during the preparation of data for publication, and will improve the usefulness of the final product.

In our experience, GPS data are required at the base and top of the section, as well as at the start and end points of each section segment if the section is discontinuous. The minimum recorded data should consist of x, y, and z co-ordinates (i.e. map co-ordinates and elevation) along with the horizontal and vertical datums of the reference spheroids. It is also recommended, especially for longer sections, that co-ordinates be acquired at major changes in lithology, possible unconformities, significant covered intervals, and formation, member, or group contacts. At the same time, these points should be plotted on a topographic map. Because co-ordinates derived from GPS may not always agree with those derived from a map (e.g. Bednarski, 2001), it is desirable, whenever feasible, to sketch the section trace independently on a map or aerial photograph. At the very least, GPS co-ordinates should always be plotted and checked critically against a map and aerial photograph while still on the section. These observations should also be supplemented with written descriptions of site locations, as well as photographs of the section and surrounding topography. Oblique aerial photographs can be especially helpful in this regard.

Recommendations for data submission

Requirements for data submission (format, standards, timetable, etc.) should be communicated by project leaders to participants before field work is undertaken. Ideally, stratigraphic section location data should be submitted to project managers as either:

a. spatially referenced digital traces with tagged attributes; or

b. spatially referenced digital traces accompanied by a database of the attributes (e.g. province/territory, map number, collector, elevation and UTM co-ordinates of base of section, elevation and UTM co-ordinates of top of section, lithological unit, contacts, etc.).

If digital-format data cannot be provided or submission of hard copies is preferred, then staff resources will need to be made available to convert hard-copy data to a digital format. In addition, the following raw elements should be included in submissions of 'tombstone data' for stratigraphic sections:

- total section thickness;
- date section was measured;
- stratigraphers, paleontologists, etc., present during measuring; and
- citation of any previous workers whose work has been used as a guide.

A brief description of the methodology used in measuring the section is also useful.

When section locations are provided as hand-drawn tracings on a paper map, section traces should be drawn using a sharp pencil or fine-tipped pen. Section traces drawn in digital format should use a line no wider than is necessary to be clearly legible. Section labelling should not cover or obscure any important map references. Data should be presented in a consistent manner, unless a clear explanation is provided for variations in technique.

If co-ordinates have been converted to an alternate datum or format (e.g. NAD 27 to NAD 83, UTM to latitude and longitude), the converted co-ordinates should be double checked for consistency against the original location co-ordinates. This should be done by the geologist before location data are submitted.

Recommendations for data handling

In verifying and validating location data, we have found that it is important to record the lineage of any data set. This should comprise a text file listing the process of how the data have been manipulated, and when. This provides a concise record that is especially useful if the data are being used by a group of scientists and technicians, eliminating potential confusion as to which version of a file is current. For example, if sections are digitally traced and co-ordinates are then taken from the tracing, co-ordinates in the database may no longer match the co-ordinates provided by the stratigrapher. If no record has been made of this, confusion may arise when data are passed from one technician to the next.

Recommendations for report preparation

Because spatial data may be overlooked or treated cursorily in the final editing stage of published reports and theses, it is up to the author to double check and ensure that co-ordinates are presented correctly. In preparing any publication, incorrect location co-ordinates should be guarded against with the same thoroughness as errors in spelling or grammar that hide the intended meaning of a text. Indeed, incorrect location data should be treated as a problem every bit as serious as incorrect scientific data.

ACKNOWLEDGMENTS

Stratigraphic location data analyzed during preparation of this report were collected as part of the Central Foreland NATMAP Project. We thank the stratigraphers who worked on the project. The review of stratigraphic location data was carried out by Palmer as part of a co-op undergraduate student placement through the Department of Geography, University of Calgary, funded by GSC Project NR4600. We thank the following persons for their comments and assistance: Andrew Couch, Peter Davenport, Dave Morrow, Barry Richards, Sandra Rosenthal, James White, and John-Paul Zonneveld. Tony Hamblin critically reviewed the manuscript.

REFERENCES

Bednarski, J.
2001: Serious mapping problems when using NTS digital bases and GPS locations; *in* Central Foreland NATMAP Project, R.B. MacNaughton and L. MacDonald (ed.); 2001 Fall Workshop, Program and Abstracts, p. 5.

Cheng, Q.
2003: From digital mapping to spatial information extraction in geographic information systems; Geolog, v. 32, no. 1, p. 8.

Compton, R.R.
1985: Geology in the Field; John Wiley & Sons, New York, New York, 398 p.

Greenly, E. and Williams, H.
1930: Methods in Geological Surveying; Thomas Murby & Co., London, United Kingdom, 420 p.

Harrison, J.M.
1963: Nature and significance of geological maps; *in* The Fabric of Geology, C.C. Albritton (ed.); Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, p. 225–232.

Himus, G.W. and Sweeting, G.S.
1972: Elements of Field Geology (third edition); University Tutorial Press Ltd., London, United Kingdom, 280 p.

Lahee, F.H.
1961: Field Geology; McGraw-Hill Book Company, Inc., New York, New York, 926 p.

Lane, L.S.
2004: Foreword to Central Foreland NATMAP: stratigraphic and structural evolution of the Cordilleran Foreland, Part 1; Special Issue, Bulletin of Canadian Petroleum Geology, v. 52, no. 4., p. 273–275.

Lane, L.S., Cecile, M.P., Currie, L.D., and Stockmal, G.S.
1999: Summary of 1998 Fieldwork in Trutch and Toad River map areas, Central Foreland NATMAP Project, northeastern British Columbia; *in* Current Research 1999-E, Geological Survey of Canada, p. 1–8.

Miall, A.D. and Miall, C.E.
2001: Sequence stratigraphy as a scientific enterprise: the evolution and persistence of conflicting paradigms; Earth-Science Reviews, v. 54, p. 321–348.

Salvador, A. (ed.)
1994: International Stratigraphic Guide (second edition); The International Union of Geological Sciences, IUGS Secretariat (Trondheim) and The Geological Society of America, Inc. (Boulder). 214 + xix p.

Talent, J.A.
1989: The case of the peripatetic fossils; Nature, v. 338, p. 613–615.

Geological Survey of Canada Project NR4600