Enhancing the value of public vintage seismic data in the Italian offshore

Paolo Diviacco | Marco Firetto Carlino | Alessandro Busato

Abstract

Vintage geophysical data represent a huge heritage for the whole scientific community (Sopher, 2018; Schaming et al., 2017; Diviacco et al., 2015) because, despite the antiquated acquisition methods and processing tools, they are generally characterized by high penetration and wide regional extension. Nowadays such large-scale projects are in fact very difficult to take place, considering environmental, geopolitical, and funding issues. In addition, reprocessing vintage seismic profiles using up-to-date software and tools, it is now possible to further increase their original signal-to-noise ratio. Italian offshore areas have been widely investigated by a dense network of multichannel seismic reflection profiles, acquired by the Italian Authorities in the years 1960, 1970, and 1980 of the last century in the perspective of mineral prospecting. This asset of data has been recovered, reprocessed, and made accessible within a demanding project undertaken by Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS). The present work aims at describing the data recovery and reprocessing methodologies applied to enhance the quality of these vintage multichannel-seismic reflection profiles, and at detailing how to access this huge dataset, that have been made compliant with the FAIR principles (Findable, Accessible, Interoperable, Reusable) and available through the OGS-SNAP data management web portal.

KEYWORDS
data access, geophysics, Italian offshore, seismic data, vintage data

DATASET

Identifier: https://doi.org/10.6092/SNAP.cbbba251c17b55c737836768c8737589.
Creator: Diviacco,P.
Title: Sismica Riconoscitiva Legge 613/1967 - Zona A.
Publisher: Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS).
License (Tiff data): Creative Commons BY 3.0.
License (SEG-Y data): By negotiation.
Publication year: 2018.

Resource type: Dataset, SEG-Y, Tiff.
Version: 1.0.
Identifier: https://doi.org/10.6092/SNAP.10b2260a0617627002d437ea6a3df022.
Creator: Diviacco,P.
Title: Sismica Riconoscitiva Legge 613/1967 - Zona B.
Publisher: Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS).
License (Tiff data): Creative Commons BY 3.0.
License (SEG-Y data): By negotiation.
Publication year: 2018.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. Geoscience Data Journal published by Royal Meteorological Society and John Wiley & Sons Ltd.
INTRODUCTION

Vintage multichannel seismic reflection (MCS) profiles are data acquired and processed in the past that are not easy to integrate or conform to current practices (Sopher, 2018; Schaming et al., 2017; Diviacco et al., 2015). Such data are mostly stored in paper or PDF formats that are not possible to be read by modern processing and interpretation software.

Hence they need to be heavily revised and converted into usable formats. A demanding work is needed, also in terms of extracting additional information from old reports and/or from the data themselves. In Diviacco et al. (2015) it is possible to find an extensive report on the methods developed at Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS), to rescue a large geophysical dataset, acquired during the 1970s by OGS itself, in the Mediterranean area. There, the issues related to media recovery, data positioning and, in particular, the problems of infilling missing data with others coming from the conversion of paper to digital formats are described in detail.

We will here report on what has been done to recover, convert to standard SEG-Y format (Barry et al., 1975), revise positioning, and enhance signal-to-noise (S/N) ratio another large seismic dataset, acquired by the Italian Authorities for mineral prospecting at the end of the last century, following the Italian law n. 613, 21 July 1967. Differently to the case described in Diviacco et al. (2015), here no data in SEG-Y format and only very little information or documentation were available, which means that everything was reconstructed only from paper or PDF sections and maps.
According to the Italian legislation (law n. 6, 11/1/1957), the documents related to any mineral prospecting project in the Italian area must become public after one year from the termination of the granted license. The documentation must be submitted to the “Italian Ministry of Economic Development” (MISE - Ministero dello Sviluppo Economico), which archives and makes it available for possible further use. To access this large dataset, the MISE, in collaboration with “Assomineraria” (an association that gathers all the companies working on mineral prospecting in Italy), promoted a project named ViDEPI (Visibilità dei Dati afferenti all’attività di Esplorazione Petrolifera in Italia), to provide easy access to seismic lines, geological maps, well-logs, etc., through the following website: https://unmig.sviluppoeconomico.gov.it/videpi/videpi.asp

The great opportunity offered by this initiative has important limitations since MCS profiles are there available only as raster images with gross information on positioning. This, of course, makes it almost impossible to use these data in modern practices and software and, therefore limits heavily their impact on the scientific community and stakeholders.

To address the already mentioned drawbacks, OGS undertook an internal project aimed at converting this large dataset into an easily usable one. The objectives of this project were essentially:

- to convert the dataset from raster images to SEG-Y seismic data;
- to fix the issues related to incorrect positioning;
- where possible, to enhance the S/N ratio using advanced reprocessing methodologies;
- to further valorize this dataset, in terms of easy accessibility.

In this last perspective, the restored dataset must be made available through a web-based data system, that allows implementing the FAIR (Findable, Accessible, Interoperable and Reusable) principles, now mandatory for Horizon 2020, which is the European Union Framework Programme aimed at securing Europe’s global competitiveness in Research and Innovation, between 2014 and 2020. The FAIR approach grants the possibility to track data from a research article to the actual repository, using persistent identifiers such as DOIs (Wilkinson et al., 2016).

All the products of data restoration and reprocessing, in order to implement the FAIR principles, being associated with persistent identifiers, were submitted to a web-based Geophysical data system named “SNAP” (Seismic database Network Access Point), (Diviacco, 2005; Diviacco and Busato, 2013), that OGS is hosting and continuously developing, extending, and updating.

The restoration process, iterated for each seismic line, includes an initial procedure, which is “common” for all the MCS sections, and a later processing flow, tailored to
overcome specific problems found on each seismic line. At the very beginning, each paper MCS section needs to be scanned. Many of the seismic lines have been directly scanned by OGS, many others, as stated above, are already available on the ViDEPI website. These can be downloaded as raster PDF files, but since the software that converts scanned seismic sections to SEGY requires files in TIFF format, a PDF-to-TIFF conversion must be performed as first step.

All files have been subsequently inspected to get the main characteristics of the data, i.e., the acquisition time window length, the shot, and the receiver group intervals. While the first tells us about the vertical two-way travel time (TWT) scale length, the last two are used to count the number of traces in every seismic line. The distance between common midpoints is approximately half the receiver group interval, while the number of traces between two consecutive shots is equal to shot interval/half-receiver group interval. This value is then multiplied by the number of shots fired during acquisition, which is normally reported on top of the images of the seismic lines, to calculate the total number of traces in the stacked sections. In case traces related to the first or last shot were present in the vintage stacked sections, these were included as well.

It is very important to consider that very often the original data were available as segmented paper seismic profiles. This means that several PDF files were to be gathered to reconstruct each single MCS line. Unfortunately, not all the segments are consecutive to each other, sometimes they overlap, some of them are missing or were acquired in inconsistent directions.

The reference information that allowed us to understand and fix such anomalies is the seismic line navigation, originally available on the ViDEPI website or as paper maps that needed to be vectorized. Albeit with several limitations and errors, this information was fundamental for a proper data georeferencing. In fact, starting from the available navigations, calculating the overall length of each seismic section, which is given by multiplying the number of shot for the shot interval, and considering also the cross-points

FIGURE 1 Map of the “Sismica riconoscitiva delle zone marine” dataset
with the surrounding lines (reported on top of the original seismic images), it was possible to reconstruct a good approximation of the original seismic surveys, overcoming the often inconsistent segmentation of the scanned profiles.

The reconstructed navigations were subsequently discretized in several shot-points (generally one point every 5 or 10 shots), arranged in ascending order and exported in the standard UKOOA-P1/90 format (https://www.iogp.org/wp-content/uploads/2016/12/P1.pdf). Multiple navigation files for the same seismic line were generated in the event of redundant shot-point values, to avoid problems in data georeferencing; in this case, the correspondent seismic line was split into two or more sections as well.

5.1 TIFF to SEG-Y conversion

Seismic sections were subsequently converted into SEG-Y format, following Miles et al. (2007). The method was developed in the framework of the EC MAST projects “Seiscan” and “SeiscanX,” in which OGS participated as a partner, implementing a raster-to-SEG-Y software conversion, named “SeisTrans.”

For a proper conversion, firstly it was necessary to select a box inside every single image to convert, including only the seismic traces to vectorize; in the frequent case of multiple images to merge in a single SEG-Y file, any eventual redundant portion was included only once. The TWT window length (WL) was then assigned to the selected data area and the sampling interval ($\Delta t$, generally equal to 2 or 4 ms) specified, to allow the software to associate a certain amplitude value to several pixels of the raster files at a fixed time rate, and thus generating a number of samples ($N$) for each trace, such that $N = WL/\Delta t$. To increase $S/N$ ratio and to reduce abrupt (high frequency) variations in pixels values, we often needed to apply high-cut filters. In addition, any vertical and horizontal line, that often cut the seismic data to indicate traces or isochronous horizons, were marked as “parasites” and not included in the raster-to-vector conversion.

Whenever seismic lines were available as multiple sections, often partially overlapping, it was chosen to merge them into a single SEG-Y file, in order to provide a unique georeferenced file for each MCS line. When this was not possible, for example when the line was acquired following different directions, it was necessary to divide the single line into sublines.

The last step of the process was to specify the number of traces, i.e., the number of column vectors that will compose the MCS section; for each of them, the software returns amplitude values spaced at an interval $\Delta t$.

To allow a proper data georeferencing, UKOOA-P1/90 navigation and SEG-Y files needed to be linked; this was done using the shot-point numbers as a reference value. They were loaded in the SEG-Y trace headers, using the “SeisTrans” software or other commercial programmes or utilities developed by OGS, and associated with those reported on the UKOOA-P1/90 navigation files. It was possible then to derive any shot-point position by interpolation. Following this procedure, for every vintage seismic section, a SEG-Y, and UKOOA-P1/90 navigation files were generated.

5.2 Reprocessing of vintage seismic data

As stated above, the quality of the vintage SEG-Y files depends largely on the quality of the original raster files, which in turn comes from the scanning of old papers, and thus it cannot be compared to that of the “modern” MCS lines.

However, similar to what Diviacco et al. (2015) did for other datasets, we tried to improve $S/N$ ratio applying several post-stack processing algorithms to a subset of all the available converted MCS lines. Processing of vintage data was mainly aimed at attenuating the artifacts introduced during the scanning of the old paper documents and the process of conversion of the raster files. Moreover, post-stack time migration was applied to enhance horizontal resolution collapsing diffractions and to put reflected events in their correct position, laterally as well as vertically.

Every single SEG-Y file was then imported in dedicated processing softwares; the first step was to mute the noise recorded during the time span between the airgun shot and the primary bottom reflection arrival. Subsequently, a band-pass filter was applied to attenuate the frequency components not related to seismic signals, but only to noise introduced by the dematerialization of the original paper documents. Frequency filtering, depending also on the original sampling rate of the data (generally 2 or 4 ms) was generally set to: low cut – 4/8 Hz; high cut – 120/240 Hz). Since, as stated above, most of the vintage MCS lines loaded on the SNAP database come from the combination of several converted raster files, often showing considerable differences in image quality, amplitude values along each composed MCS section needed to be balanced. For this reason, we applied two different algorithms. The first is an ensemble scaling, which calculates the average absolute amplitude value $A$ of all traces, and then multiplies each sample for a scale factor (mostly equal to 1) divided by $A$. The second is a weighted traces mixing, which generates a new section where each seismic trace is given by the sum of the same trace and the contiguous ones, each multiplied by a scale factor (often we mixed five traces assigning the following scale factors: 0.33 - 0.66 - 1 - 0.66 - 0.33), (Figure 2a). To attenuate random noise uncorrelated from trace to trace, enhancing coherent signals, we often applied spatial prediction filters designed and applied in the
Amongst the many processes applied to seismic data, seismic migration is the most directly associated with the concept of seismic imaging. Regardless of the approach implemented, since migration involves retropropagation of the recorded wavefield back to the reflecting interfaces, a precise estimation of the actual subsurface seismic velocity is essential. To adopt a suitable velocity field, we used all the available information, derived from well data or from the stacking velocities reported on top of the vintage seismic images. Figures 3 and 4 show that migration successfully imploded diffraction hyperbolas and migrated the dipping reflectors to different space-time locations, making easier the interpretation of the fault planes. In particular, the migration procedure applied is based on finite-
difference solutions to the scalar wave equation (Doherty and Claerbout, 1974; Claerbout, 1985).

The processing sequence significantly improved the S/N ratio on MCS vintage lines. Additional steps were applied to further increase S/N ratio, keeping in mind, however, that since processing can be possible only on post-stack data the data quality will very unlikely improve from the results shown in this paper.
DATA ACCESS

The world of scientific research is increasingly interconnected. New ideas and visions can emerge and be corroborated only providing researchers with the possibility to confront and discuss at all levels. In this perspective, the possibility to access other researchers and institutions data is a key factor.

The importance of this vision is well understood by funding agencies but is becoming more and more important also in the world of scientific publications. In fact, the possibility to replicate the experiments (an additional R to the FAIR acronym), the need to access data that often can be difficult to represent graphically and the possibility to connect and interact directly with the authors of a research, can, potentially, change completely the paradigm in this sector.

Often, scientists themselves are responsible for the reluctance in sharing. This mental attitude, most of the times, is related to the need to protect their work and investments. In fact, when data are not acquired within a framework that mandates full data openness, a completely open approach is sometimes perceived as a threat to intellectual property. The need for interconnectivity and will to impact the scientific community, and the need to protect the intellectual property, should, therefore, be balanced carefully.

To address such problems, OGS developed a specific web-based data system called SNAP (Seismic data Network Access Point) (Diviacco, 2005; Diviacco and Busato, 2013) where all geophysical data OGS exposes are uploaded. SNAP holds data mainly in the Mediterranean Sea and the Black Sea and can be reached at the following web address: https://snap.ogs.trieste.it.

Other similar initiatives, based on the same framework, but that will not be described here, are available for other areas, such as the Antarctic Seismic Data Library (SDLS) in Antarctica, or the ARCA initiative in the polar areas.

SNAP grants access to the data contained upon permissions assigned to the end user after a phase of data licensing. Once users find the data they are interested in, they can submit a request for a license. OGS evaluates the request, issues the license and sets how the end user can access the data on the web portal, which can be either seeing or positioning only, previewing data with watermarking, full previewing or even downloading of the data. This can be set dataset by dataset and user by user.

At the same time, SNAP allows to mint and handle DOIs.
These are generated and assigned to the data upon their upload, thanks to the fact that OGS is a member of DataCite via the Conferenza dei Rettori delle Università Italiane (CRUI), an organization that gathers all Italian Universities and Research centers.

A DOI example produced by SNAP is: 10.6092/SNAP.cbbba251c17b55c737836768e8737589

This identifies a subset (Zona A) of the whole dataset mentioned in this paper (the full list of DOIs related to the dataset described in this paper can be found in the references), and can be searched for, using the search page of the DataCite web site or directly including the DOI in a URL which uses a resolver (https://doi.org/) which support HTTPS, such as in the following example: https://doi.org/10.6092/SNAP.cbbba251c17b55c737836768e8737589

For each loaded dataset, SNAP, automatically generates a landing page after Leadbetter et al., (2013) which URL is associated with the relative DOI. It is, therefore, possible upon resolution, to be sent immediately to the SNAP landing page where all information about the requested data can be immediately seen (Figure 5). From here it is possible to access directly the data (Figures 6 and 7).

7 | CONCLUSIONS

This work provides to the geoscience community the possibility to access a large (over 65,000 km) and important asset of vintage seismic lines in the area of the Italian peninsula. This has been obtained through the recovery, digitalization, and conversion of a very large quantity of paper seismic sections that, so far, were only partially usable. In fact, through the process applied, the data eventually obtained can now be used with modern data processing and interpretation software. Following a FAIR approach, the whole dataset was made available through the web data system named SNAP that allows the use of DOIs to identify, cite, and follow links to preview and download the data.

OPEN PRACTICES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://doi.org/10.6092/SNAP.cbbba251c17b55c737836768e8737589. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.
CONFLICT OF INTEREST

The authors declare that they have no potential source of conflict of interests.

ORCID

PaoloDiviacco http://orcid.org/0000-0001-9404-9611

REFERENCES

Barry, K.M., Cavers, D.A. and Kneale, C.W. (1975) Recommended standards for digital tape formats. Geophysics, 40(2), 344–352. Available at: https://doi.org/10.1190/1.144053.

Claerbout, J.F. (1985) Imaging the Earth’s Interior. Hoboken, NJ: Blackwell Scientific Publications.

Diviacco, P., Wardell, N., Forlin, E., Sauli, C., Burca, M., Busato, A et al. (2015) Data rescue to extend the value of vintage seismic data: the OGS-SNAP experience. GeoResJ, 6, 44–52. Available at: https://doi.org/10.1016/j.grj.2015.01.006.

Diviacco, P. and Busato, A. (2013) The Geo-Seas Seismic data viewer: a tool to facilitate the control of data access. Bollettino di Geofisica Teorica Applicata, 54(3), 257–270

Diviacco, P. (2005) An open source, web based, simple solution for seismic data dissemination and collaborative research. Computers and Geosciences, 31(5), 599–605.

Doherty, S.M. and Claerbout, J.F. (1974) Velocity analysis based on the wave equation, Stanford Expl. Proj., Rep. No. 1, Stanford University.

Leadbetter, A., Raymond, L., Chandler, C., Pikula, L., Pissierssens, P. and Urban, E. (2013) Ocean Data Publication Cookbook, IOC Manuals and guides, No. 64.

Miles, P., Schaming, M. and Lovera, R. (2007) Resurrecting vintage paper seismic records. Marine Geophysical Researches, 28, 319–329. Available at: https://doi.org/10.1007/s11001-007-9034-5.

Schaming, M., Rise, L., Chand, S., Reiduly, B., Osmundsen, P.T. and Redfield, T. (2017) The SPARDIG project – transforming analogue sparker records from the Norwegian continental shelf into SEG-Y format, first results, Poster at EGU2017.

Sopher, D. (2018) Converting scanned images of seismic reflection data into SEG-Y format. Earth Science Informatics, 11, 241. Available at: https://doi.org/10.1007/s12145-017-0329-z.

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, 3, 160018. Available at: https://doi.org/10.1038/sdata.2016.18.

Yilmaz, O. (2001) Seismic data analysis: processing, inversion, and interpretation of seismic data. Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560801580.

How to cite this article: Diviacco P, Firetto Carlino M, Busato A. Enhancing the value of public vintage seismic data in the Italian offshore. Geosci Data J. 2019;6:6–15. https://doi.org/10.1002/gdj3.58