GeoResJ 6 (2015) 98–105

Contents lists available at ScienceDirect

GeoResJ

journal homepage: www.elsevier.com/locate/GRJ

Curating the web's deep past – Migration strategies for the German Continental Deep Drilling Program web content

Jens Klump [a,*], Damian Ulbricht [b], Ronald Conze [b]

[a] Commonwealth Scientific and Industrial Research Organisation, 26 Dick Perry Avenue, Kensington, WA 6151, Australia
[b] GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany

ARTICLE INFO

Article history:
Received 22 October 2014
Revised 20 February 2015
Accepted 21 February 2015
Available online 21 March 2015

Keywords:
Data rescue
Long-term preservation
Data curation
Metadata
Data publication
Continental Deep Drilling

ABSTRACT

On timescales beyond the life of a research project, a core task in the curation of digital research data is the migration of data and metadata to new storage media, new hardware, and software systems. These migrations are necessitated by ageing software systems, ageing hardware systems, and the rise of new technologies in data management. Using the example of the German Continental Deep Drilling Program (KTB) we outline steps taken to keep the acquired data accessible to researchers and trace the history of data management in KTB from a project platform in the early 1990ies through three migrations up to the current data management platform. The migration steps taken not only preserved the data, but also made data from KTB accessible via internet and citable through Digital Object Identifier (DOI). We also describe measures taken to manage hardware and software obsolescence and minimise the amount of maintenance necessary to keep data accessible beyond the active project phase. At present, data from KTB are stored in an Open Archival Information System (OAIS) compliant repository based on the eSciDoc repository framework. Information packages consist of self-contained packages of binary data files and discovery metadata in Extensible Mark-up Language (XML) format. The binary data files were created from a relational database used for data management in the previous version of the system, and from websites generated from a content management system. Metadata are provided in DataCite, GCMD-DIF, and ISO19139/INSPIRE schema definitions. Access to the KTB data is provided through download pages which are produced by XML transformation from the stored metadata.

© 2015 Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Speaking to people about long-term curation, one thing that is mentioned in almost all conversations is the floppy disk. To many, the floppy disk epitomises what they see as the core challenges in long-term data curation: bit stream preservation and media obsolescence. It is not only this particular medium that is problematic; the general still rapid development of information technology requires regular migrations of content, media, hardware and software. These challenges have been recognised early on and a very well written overview can be found in Rothenberg [21].

Many analytical data in the geosciences can be represented as tables and encoded as character separated value (CSV) files. Accompanied by descriptive metadata, these files pose a relatively minor challenge to format migration and their often small size does not demand large computing resources for migration processes. The challenge lies in the system migrations and describing the contents in metadata for discovery and reuse.

Initially, the web based components of projects in the 1990s were solitary systems, today often termed "silos", run in the context of large projects or as efforts by individual researchers. This is also true for most projects of the International Continental Scientific Drilling Program (ICDP). Notable exceptions are systems like PANGAEA [6] which, from going online in 1995, curates and disseminates data from many different projects in marine environmental research. Project based systems all face the challenge of curating the data long past the end of the project when resources, such as contextual knowledge of the project and funding, may no longer be available. In this sense, this paper does not describe the rescue of data that might have been lost to media obsolescence or had to be digitized from analogue media, but rather the challenges posed by technical obsolescence. In the course of this paper we will discuss the strategies employed in successive projects over 25 years to migrate the data dissemination platform of the German

* Corresponding author. Tel.: +61 8 64368828.
E-mail addresses: jens.klump@csiro.au (J. Klump), damian.ulbricht@gfz-potsdam.de (D. Ulbricht), ronald.conze@gfz-potsdam.de (R. Conze).

http://dx.doi.org/10.1016/j.grj.2015.02.011
2214-2428/© 2015 Elsevier Ltd.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Continental Deep Drilling Program onto new technical platforms. Unlike the ocean drilling programmes, ICDP drilling projects do not have a common structure and differ widely in form and extent of involvement of the ICDP Operational Support Group. This heterogeneity makes it difficult to apply the same approach to all ICDP legacy projects. The scientific review of ICDP in 2014 recommended following the data curation procedures developed for KTB and CONTINENT in future ICDP projects.

2. Pre-web and web KTB

The German Continental Deep Drilling Program (Kontinentale Tiefbohrung, KTB) was a large scale geoscience project conducted from 1987 to 1995 in Windischeschenbach, Germany. Its two super-deep boreholes (4000 m and 9101 m) are worldwide unique masterpieces of drilling engineering. The programme yielded essential insights in the structure and processes of the upper crust of the Earth. For this reason it is one of the most important geoscientific and geotechnical research projects. The great success in geosciences and drilling engineering induced the scientists to establish the International Continental Scientific Drilling Program (ICDP). An overview of the scientific achievements from KTB can be found in Emmermann and Lauterjung [8].

The KTB Information System was set up to perform two major functions in the information management in the context of KTB: (1) document and store project data, and (2) support interdisciplinary dissemination of the data. Data were originally stored on tape and at the end of the KTB project migrated from tape storage onto optical storage media. Already in times before the worldwide web data were accessible over the German Research Network (Deutsches Wissenschaftsnetz, WiN), a precursor system to the worldwide web based on the X.25 network protocol. Wächter [26] gives a comprehensive overview of the system at the KTB site during the peak of its operation. After the end of the project, and with the emergence of the modern internet, significant proportions of the data were ported to a web application with a browser based user interface (Fig. 1). The focus of this first migration was on tabulated data. Raster data, such as images of drill cores, and seismic exploration data were deemed as being too large in volume at the time and were stored offline. This study will only discuss the original

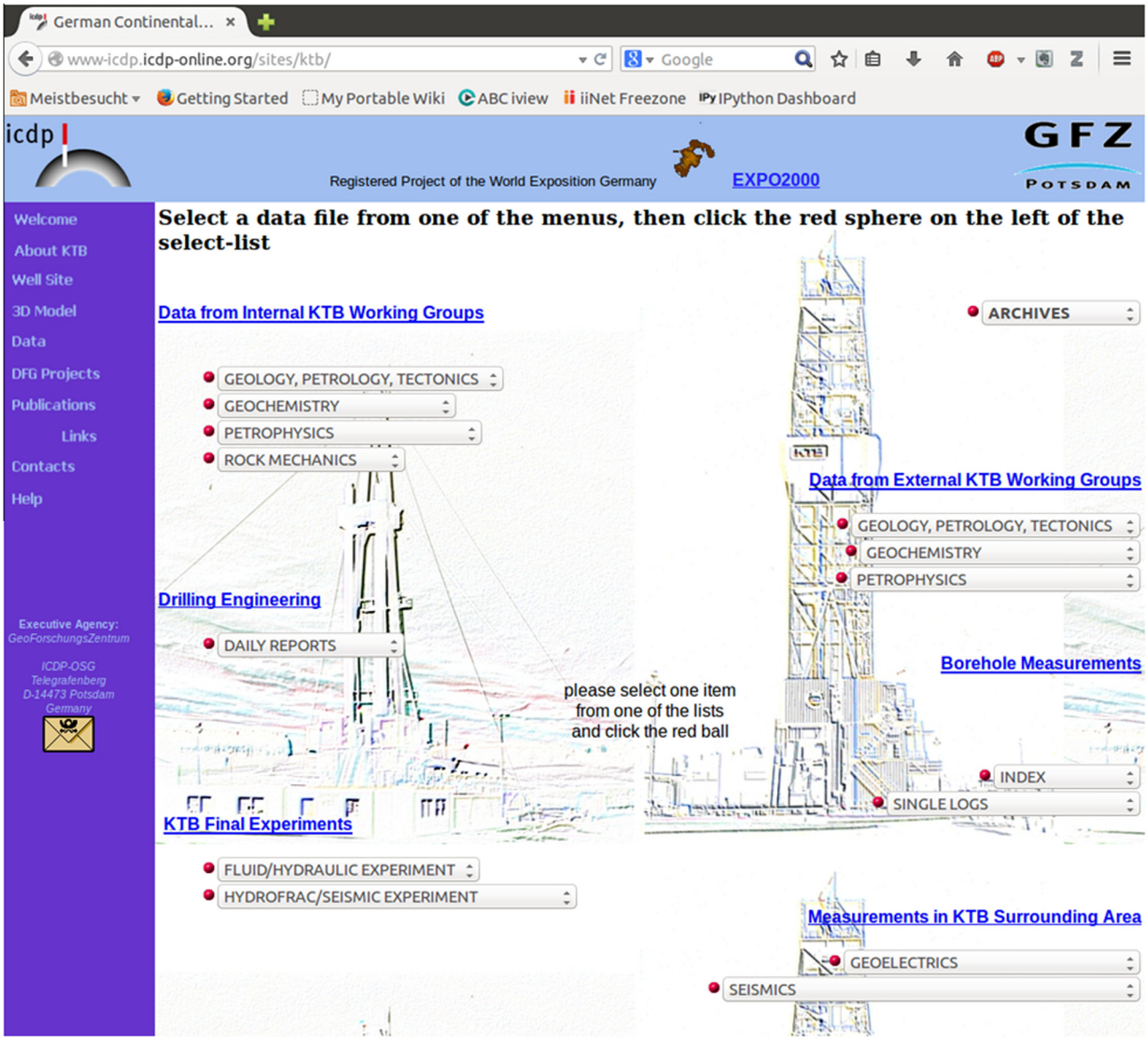

Fig. 1. Screenshot of the first KTB web presence, data access page.

Download English Version:

https://daneshyari.com/en/article/4674449

Download Persian Version:

https://daneshyari.com/article/4674449

Daneshyari.com