DIGITISATION OF TELLURIC RECORDINGS OF THE SZÉCHENYI ISTVÁN GEOPHYSICAL OBSERVATORY OF THE HUNGARIAN ACADEMY OF SCIENCES – IAGA DIVISION 5. OBSERVATORY, INSTRUMENTS, SURVEYS AND ANALYSES

Tamás Nagy[1], Orsolya Baillie[2], Árpád Kis[1], Károly Kovács[1], Dávid Koronczay[1], István Lemperger[1], Viktória Németh[1], Attila Novák[1], Ernő Prácser[1], Judit Szendrői[1], Eszter Szűcs[1], Viktor Wesztergom[1]

1 Intoduction

Earth current measurement with high temporal resolution started at the Nagycenk Observatory (also known as Széchenyi István Geophysical Observatory of the Hungarian Academy of Sciences) in late 1957. This is a representative, homogeneous and unique data set for statistical analysis of the long-term variation of the geomagnetic induction effect which represents a real space weather related risk. In the frame of the EURISGIC FP-7 Space project a special semi-automatic digitizing method and workflow had been proposed and elaborated (Nagy Telluric Digitiser – NTD method) to digitize the historical telluric recordings. Altogether 4539 film rolls (with a total length of approximately 8 km) has been digitized, covering data from the year 1957 to 1997. During the process a best practice has been developed about how to split the project into multiple stages allowing collaborative work of a group of people, how to store data in a secure and easily accessible way, how to control the quality of the resulting data product and how to make it available to the public.

2 Discussion

There are essentially five different layers of the digitization process. The first layer holds the analogue data stored on a transparent material comparable to film rolls. This serves as the initial input to the entire process. The second layer holds static (or rarely changing) files which are the digitized (scanned) versions of the film rolls in the first layer. The result of vectorising the files on the second layer gives the files on the third layer. During the process of vectorization these files get updated very frequently, therefore the files associated to this layer are under version control. Additional to the vectorised data, on this layer an another type of file collection, the so-called reference time catalogue exists also in the form of version controlled files. On top of that, there is also a film roll catalogue here which holds textual data collected from the film rolls. All information contained on this layer is then loaded into a relational database, which serves as our fourth layer. The fifth layer consists of collection of server side programs and SQL views which serves data to the clients from the relational database (fourth layer) or from the image repositories (second layer). Analogue film rolls were digitized using an ordinary long-format paper scanner in the resolution of 118.18pixels/cm, which corresponds to 300dpi. 5cm on the film roll covers 2 hours of registration. A typical length of a film roll varied somewhere between 0.5 and 1m, but some of them where as long as 4.5m. These latter ones were too large to scan them in a single shot due to the limitations of our device, leading to the need of further post-processing on the resulted digital images. Altogether 4539 files were created by a group of 5 people, but to reduce their dimensions we cut them into smaller segments, therefore the total number of files in our final image repository was 22055. The film roll catalogue (which is a single plain text file containing tabular information) of layer3 were created during the scanning process and the reference time catalogue (which is a hierarchical collection of plain text files containing tabular information) was created during the cutting process.

Engauge Digitizer (ED) to vectorise the curves which were present on our digital images. The vectorization involved 2 steps: first we defined a coordinate system on each image by specifying the

[1] *MTA CSFK Geodetic and Geophysical Institute*
E-mail: nagy.tamas@csfk.mta.hu
[2] *British Geological Survey, Geomagnetism Group, Edinburgh*

actual coordinates of three reference points (this step defines the transformation between pixel coordinates and actual coordinates), then we traced the point of the curve with a digitizing tablet using either the semi-automatic or the manual sampling mode offered by ED. This step also resulted in a hierarchical collection of plain text files containing information about the vectorised curves in a tabular format.

So as to facilitate the data access all of the information contained in the files of layer 3 is loaded into a PostgreSQL database. In this database there is a separate table for the film roll catalogue and another one for the reference time catalogue. In order to speed up queries, the vectorised point sets are sorted into smaller tables each of which containing nearly one year of registration. The end users communicate only with this relational database, either directly via a database view or via a web service wrapped around a certain view. A graphical user interface for exploring our data is available at: http://geodata.ggki.hu/tellurics

Acknowledgements. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7 / 2007–2013) under grant agreement 260330.
This study was supported by the TAMOP-4.2.2.C-11/1/KONV-2012-0015 (Earth System) project sponsored by the EU and European Social Foundation.

Development of the research infrastructure was also sponsored by the Infrastructure Development Fund of the Hungarian Academy of Sciences.

References

Baillie O, Thomson A, Nagy T, Dawson E, Kelly G, Wesztergom V, Clarke E (2013): Extreme value statistics applied to geoelectric activity in Europe. European Space Weather Week 10, Antwerpen, 18-22nd November 2013 (S1: Poster #7)

Nagy T, Wesztergom V, Lemperger I, Kis Á, Prácser E, Kovács K (2014): The digitization of historical telluric recordings. In: Kovács J, Szabó Gy M (Eds.): New challenges in astro- and environmental informatics in the Big Data era, Proceedings of the workshop, Szombathely, ELTE Gothard Asztrofizikai Obszervatórium, 2014. 91-95.