PROCEEDINGS OF SCIENCE

E-Science Activities in MAS/ Mongolia

OTGONSUVD Badrakh *a,**, ***UUGANBAATAR Dulamragchaa*** *a*

a Institute of Mathematics and Digital Technology, Mongolian Academy of Sciences
Enkhtaivan avenue 54b, Ulaanbaatar, Mongolia

E-mail: otgonsuvd@mas.ac.mn

In this article, we discussed the data at the Mongolian Academy of Sciences and the data created from the projects carried out in the last 3 years in the Department of information technology of the Institute of Mathematics and Digital Technology. Also, there we were discussed the Computing Center of the Mongolian Academy of Sciences, the Artificial intelligence and machine learning laboratory that implements E-science in Mongolia. As a developing country, it is very important to have a Computing Center in the institutes and universities in our country.

International Symposium on Grids and Clouds (ISGC) 2023,
19 - 31 March 2023
Academia Sinica Taipei, Taiwan

*Speaker

PoS(ISGC&HEPiX2023)030

1. Introduction

There are 14 institutes in Mongolian Academy of Sciences in Mongolia. Depending on their research field, these institutes collect the following types of data which has been collected due to the research activities and field investigation performed for the last 60 years.

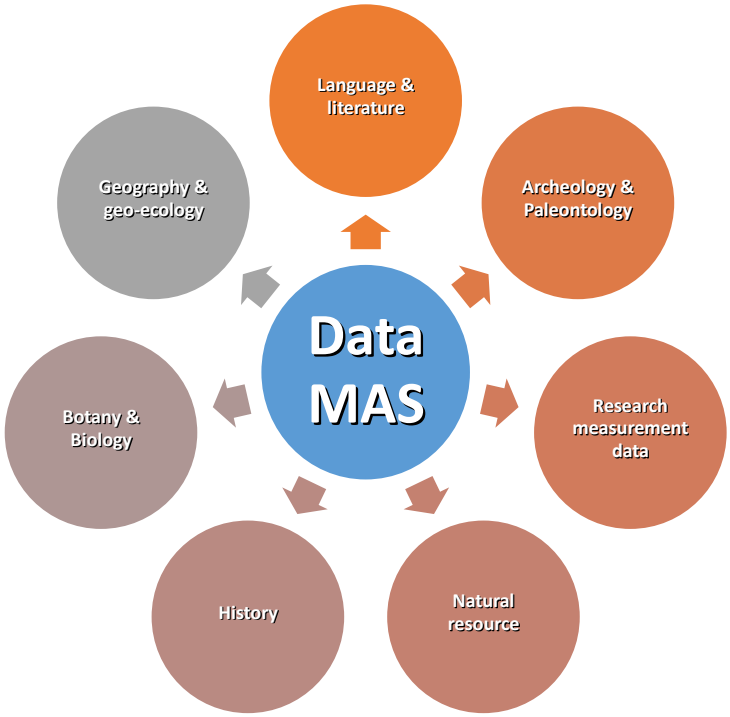

Figure 1. Data of Mongolian Academy of Sciences

2. Research data of Mongolian Academy of sciences

2.1 Language & literature archives

There are 6889 Tibetan manuscripts stored in the Institute of Language and Literature. The Mongolian people had a rich long history of writing culture and a specific technology of printing books and sutras. Today, the monuments of script that have been traditionally preserved for us, the petroglyph birch-bark paper printed sutras, to unique artworks by craftspeople embroiderers, are being kept in the National Library, Gandan Tegchlen monastery, the Institute of Language and Literature of the MAS and museums.

Most of these materials are stored in the Clear script and have been digitalized by our center for Digital Culture heritage. Books and manuscripts are on the verge of destruction as the paper dries out and burns due to malpractice of preservation procedures such as desiccation and humidity and in numerous mechanical damages. Therefore, restoration of the Mongolian cultural heritage and creation of its digital archives require the use of delicate technology and introduction of advanced methods [1].

Figure 2. Language & literature archives

2.2 History archives

MAS cooperates with The Central Asian Archives in the Department of Mongolian and Tibetan Studies at the University of Bonn, Germany. They collected digital heritages over 100 European museums. Now Mongolian heritages are also digitalized in this archive. 22.000 objects belonging to the cultural and household culture of Mongolia and Central Asia. There are 55.000 filing cards, including photographs, descriptions and data of the museums.

Figure 3. History archives

2.3 Natural resources

The Institute of Astronomy and Geophysics is responsible for processing information for the Mongolian territory network. The network consists of 6 mini arrays, 9 Broadband, 9 GPS, 12 accelerometers for early warning systems. There are 25 seismic stations in Ulaanbaatar.

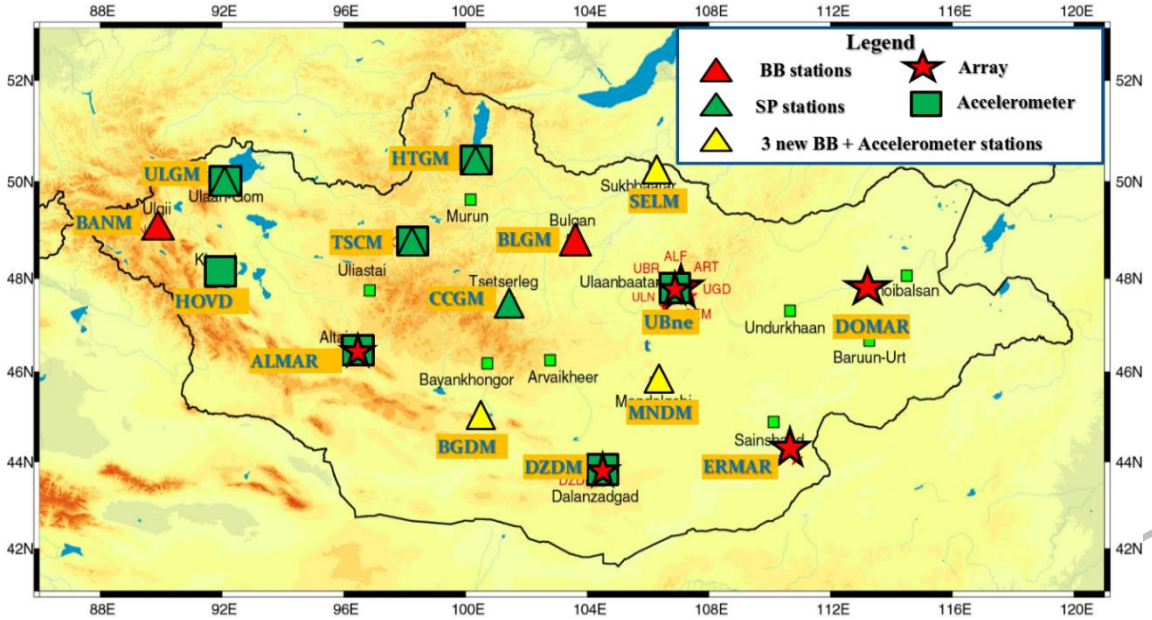

Figure 4. Seismic station network of Mongolian territory

2.4 Botany & Biology

The Institute of Botanic Garden and Institute of Biology has a database of the following information collected from various expeditions, and this year they have started to work with our institute to create a digital herbarium. This will make the above information public.

There are about 10,000 scientific documents, a catalog including descriptions of 7,800 species of insects, a catalog of habitat description of 9,000 insects, 4,000 staff models of 4,000 birds. Mongolian state fund and private herbarium collections possess 75,000 perennial, 3,500 vegetarian species of swamp areas, 10,000 weed plants, 10,000 moss plants and 2,500 fungus species. 600 traditional songs and other folklore, ~300 artworks, 20 sculptures and more than 40 kinds of instruments.

Database of Mongolian flora: (currently to belong to 134 families, 683 genuses and about 3100 species of vascular plants)

Database of Herbarium (UBA): In Mongolia there are approximately 123500 specimens collected in Herbarium.

- Database of collections of Algae (4500 sheets)
- Database of collections of Fungi (4000 sheets)
- Database of collections of Lichens (14500 sheets)
- Database of collections of masses (19500 sheets
- Database of collections of Vascular plants (81000 sheets)
- Database of Seed collections (1220 examples)

3. Data of Institute of Mathematics and Digital Technology

Our Institute of Mathematics and Digital Technology was established in August 2019 in accordance with Governmental Resolution. Our institute's main areas of research can be split into two categories of research and development of digital technologies and: the research in theoretical and applied mathematics. In the research and development of digital technologies, we are

interested in the topics of artificial intelligence and machine learning, data science, and software development.

3.1 Center for Digital Cultural Heritage

Digital Cultural Heritage Center was established in 2010 by MAS and the Max Planck Society of Germany6 CANON EOS5D Mark II digital camera and Epson Expression 10000XL using an A3 scanner. High quality printed books and ancient scriptures can be quickly digitized. In 2021, an Einscan SE 3D scanner was purchased and the equipment was added to the Cultural Heritage Digitization Center [1].

3.2 Artificial intelligence and machine learning laboratory

In terms of equipment and laboratory capacity, we have established an artificial intelligence and machine learning laboratory in 2020 and are working on research in this area. There are five high performance computers. /Dell Vostro 5090 Intel (R)Core (TM) 9-9900K @ 3.60GHz, 16 cores 32GB 2666MHz,500GB SSD + 2TB HDD, 1Gbp / s Nvidia RTX 2070 8GB(2304 CUDA cores)/

3.3 Computing center of MAS

Nowadays, simulation and calculation methods are widely used in scientific research, and the need for high performance computing centers is created. Therefore, the Computing Center has been established at the Institute of Mathematics and Digital Technology of the Academy of Sciences since 2019 in order to perform the calculations of the research work carried out in the academic institutions and universities of our country using cloud technology.

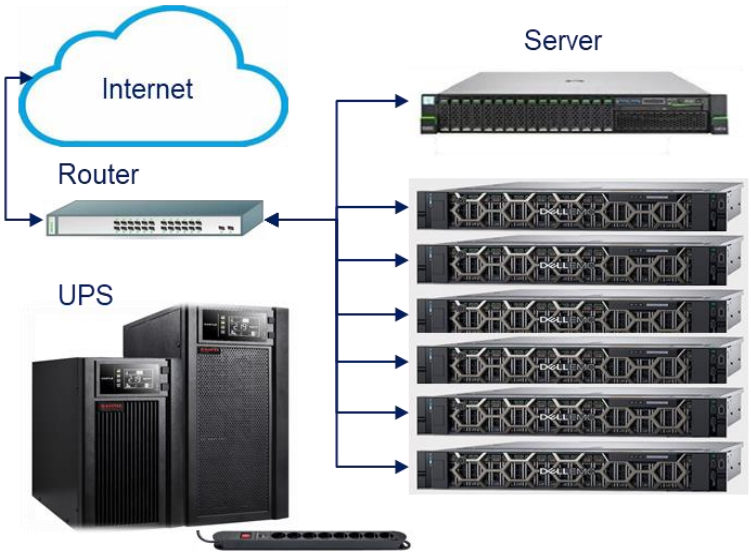

Figure 5. Computing Center structure of MAS

This center is open to all researchers from Mongolia. Currently, the computing center is installed with high energy, materials science, molecular dynamics modeling [6], and machine learning software, and additional software can be installed at the request of users for research purposes.

Table 1. Characteristics of the computing center

№	Server	Core number	NVIDIA	Memory	Capacity
1.	Fujitsu	64	V100	197 Gb	1.8 Tb
2.	Dell R740	96	-	329 Gb	1.3 Tb
3.	Dell R740	96	-	329 Gb	1.3 Tb
4.	Dell R740	96	V100S	394 Gb	2.6 Tb
5.	Dell R740	96	A 40	197 Gb	3.5 Tb
6.	Dell R740	72	-	197 Gb	2.6 Tb
7.	Dell NX3240	48	-	65 Gb	118 Tb
8.	UPS				20 kW

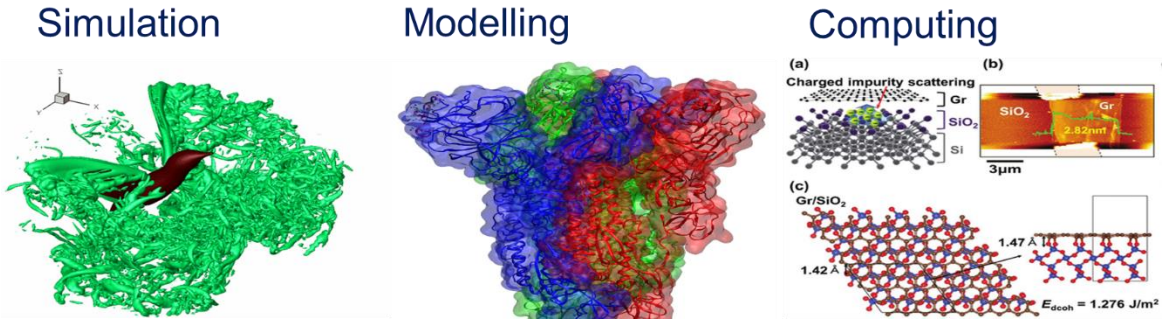

Figure 6. Performing tasks in the computing center

3.4 Research and development

In the Information technology department of our institutes, we have implemented these following projects in the last 3 years.

- Study of key technologies for Mongolian to Chinese translation based on deep learning [3]
- Research on artificial intelligence technology to recognize human faces [5]
- Color science and some applications [2],[4]

Depending on the projects and programs implemented in our department, we have created the following types of databases.

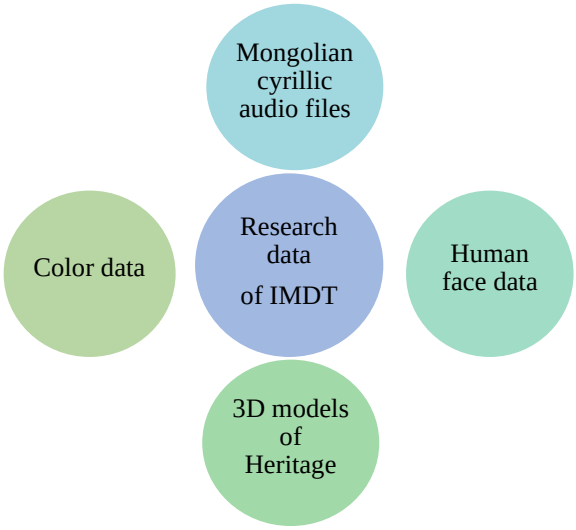

Figure 7. Data of Institute of mathematics and digital technology

3.4.1 Face recognition system based on AI

Using Artificial intelligence and machine learning laboratory and the Computing center of MAS, we have studied the face recognition training methods in deep learning and tested some of them. And to use it to identify criminals in cooperation with the Mongolian police center.

Figure 8. Block diagram of face recognition system

We trained the object detection model Yolo v5, which classifies people with and without masks, in order to train people with masks. Using this model, we developed a method to distinguish the area around the eyes and forehead which is not covered by a human face mask. In this way, Casia-Webface (a total of 494414 people with 10576 differences) extracted the area around the eyes and forehead from an open database and trained a masked person to model, which was 86% accurate on 1300 images [5].

3.4.2 Color science project

Using Artificial intelligence and machine learning laboratory and the Computing Center of MAS, is the color database compiled by the Color Science Project team. This research will develop a number of methodologies for the introduction of color science technology, including the use of multispectral technologies in the preservation of tangible cultural heritage. We also combined multispectral technology with data processing and analysis to study how color dye and cultural heritage change color depending on material, properties, and environmental influences, and conducted a number of case studies on color measurement and color data analysis methods [2],[4].

Figure 9. Color science project implementation

In order to make the open data collected from these studies available to researchers, we created an online platform with a color database.

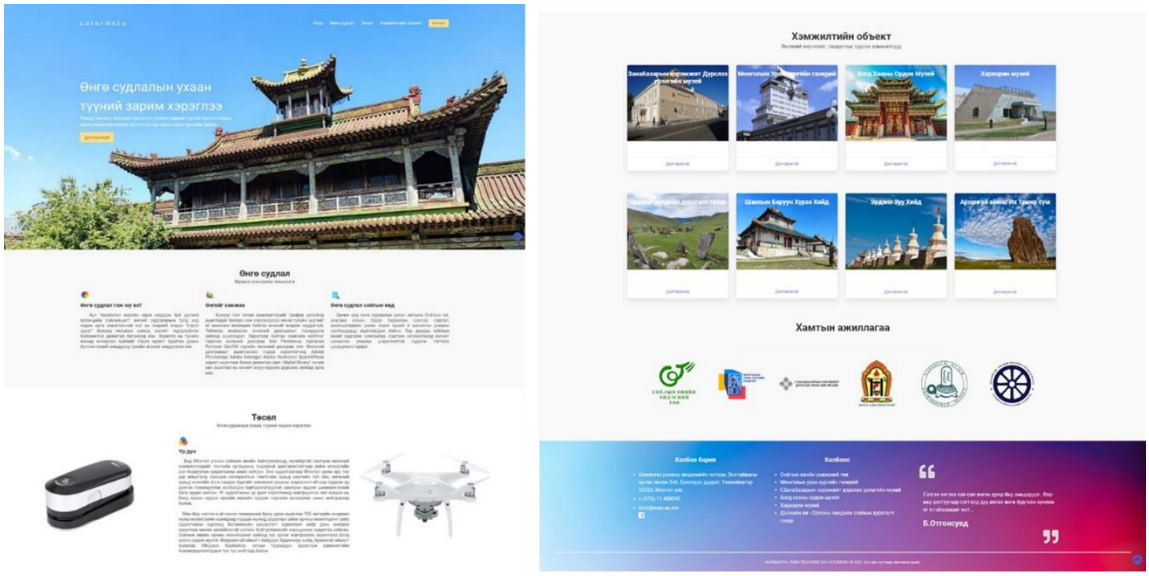

Figure 10. Color data of heritages

3.4.3 Translation of speech from Mongolia to China based in deep learning

One of the research examples performed in the Computing center of MAS is the translation system of speech from Mongolia to China based on deep learning. We created three types of data for this project [3].

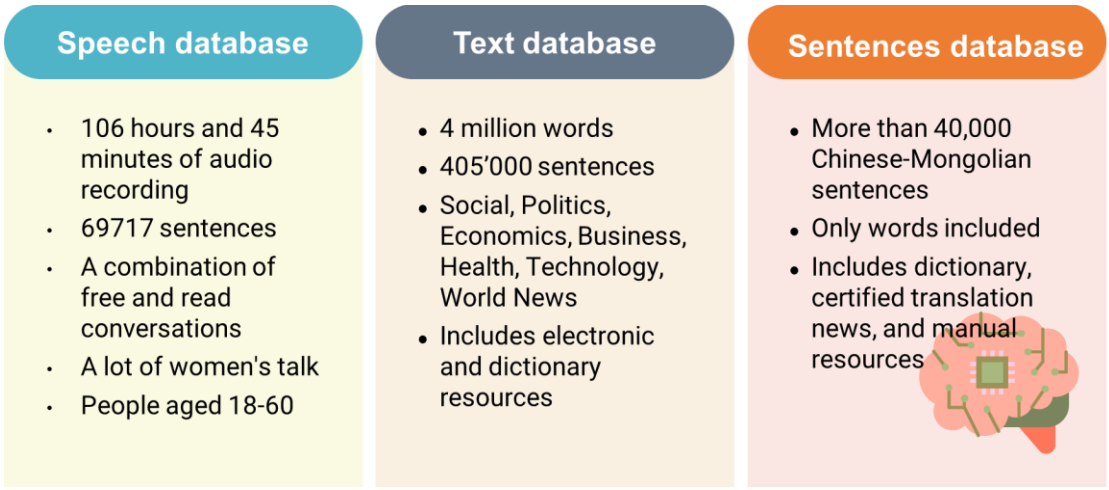

Figure 11. Database of translation of speech from Mongolia to China

4. Conclusion

As a developing country, it is very important to have a Computing Center in the institutes and universities in our country. The Computing Center with high performance computing was established at the Mongolian university of Science and Technology one month ago. Every year, we increase the capacity of the computing center of MAS, and we are trying to develop this E-science in Mongolia by keeping the use of the computing center open to the public free of charge.

For Mongolia, as research and development improves, the composition of the data will increase. Also, by increasing the capacity of the computing center and making it more open to researchers, the quality of research work will increase. Therefore, we have to adopt the examples

of developed countries and increasing a certain amount of funds for E-science is absolutely necessary.

References

[1] Nergui B, Batzaya E, Status of e-Science Activities in Mongolia (Heritage Science), Proceedings of conference ISGC 2016, PoS(ISGC 2016)036.

[2] Otgonsuvd Badrakh, Batzaya Enkhjargal, Galbaatar Tuvdendorj, Some study results of color changes depending on Mongolian environmental condition), Proceedings of conference ISGC 2021, PoS(ISGC2021)017.

[3] Uuganbaatar Dulamragchaa, Sodjamts Enkhbat, Lkhagvasuren Altankhuyag, Khajidmaa Battumur, Khajidsuren Uuganbaatar, Purevsuren Tumurbaatar, Khalkh Mongolian Speech, Text, and Parallel Dataset Collection for Machine Learning, Journal of Institute of Mathematics and Digital Technology, Vol. 3 No. 1 (2021).

[4] Otgonsuvd Badrakh, Garmaa Dangaasuren, Nergui Baasan, Galbaatar Tuvdendorj, Recognition of Paintings Using Color Data Analysis, Journal of Institute of Mathematics and Digital Technology, Vol. 4 No. 1 (2022).

[5] Davaasuren Nyamdavaa, Luubaatar Badarch, Battulga Ulziisaikhan, Battulga Ulziisaikhan, Uuganbaatar Dulamragchaa, Munkhbayar Battsogt, Real Time Face Recognition Software Development for People With and Without Face Masks, Journal of Institute of Mathematics and Digital Technology, Vol. 4 No. 1 (2022).

[6] Munkhdalai Chagdarjav, Batgerel Balt, Batsaikhan Mijiddorj, A Study on Interactions of the SARS-CoV-2 Spike Proteins with the Human Toll-like Receptor 4 using Molecular Dynamic Simulations, , Journal of Institute of Mathematics and Digital Technology, Vol. 4 No. 1 (2022).