On Digitizing Oral Tradition
— Strategy, Practice, and Collaboration —

1 acquisition

 I should like to start with the problem of acquisition.

 Recording of the oral tradition in written document began with Kojiki edited in 712 in Japan. Kojiki is a record of myths and folktales told by HIEDANO Are and written down by OHNO Yasumaro. After Kojiki we have a lot of written record of oral tradition in our country. I suppose that we can find many similar records all over the world.

 But we could say that the first academic record of the oral tradition was made by YANAGITA Kunio, founder of folklorire studies in Japan. In 1910 he published "Tono-Monogatari", the first academic collection of folktales in Japan.

 After "Tono-Monogatari" many folklorists have collected folktales.

 In 1936 YANAGITA published a manual of research, "Handbook on folktale collection(昔話採集の手帖)" and the first tale-type index of Japanese folktales, "Tale Type Index Japanese folktales ("日本昔話名彙")

I think that YANAGITA had so strong an interest in originality of Japanese folklore and folktales that he couldn't appreciate the international point of view enough.

SEKI Keigo, succeeding YANAGITA, accomplished his original type index "Japanese Folktale Collection (日本昔話集成)"（1950-1958）and its enlarged version, "Complete Collection of Japanese Folktales(日本昔話大成)"（1978～1980）which linked Japanese local tale type to AT as much as possible.

 Before 1960s when the tape recorder became popular, Japanese oral tradition was recorded only by written media. But after 60s, the audio recording became main stream.

 Almost all of the folktale researches have been carried out with personal efforts and their results, belonging to each researcher, have been conserved personally.

So we cannot know the exact quantity of recordings, but I suppose the number might be more than 300000. But almost all of them are analog recordings and are sure to be degraded in the near future.

 In Japan we can find excellent traditional folktale tellers even today and we are making efforts to collect the oral tradition, but it is becoming difficult.

Due to the urbanization and internet revolution, our life style has changed completely. Especially, losing the traditional background of the tales, often the young listeners

cannot understand vocabularies in the tales told by the older generation. We are losing not only tellers but also listeners of folktales.

While the traditional tale tellers are becoming old, we cannot find the young listeners who transmit the tales to their next generation. In Japan we are losing our oral tradition now.

We have a great problem as to how to transmit our authentic oral tradition to the younger generation In Japan.

2 necessity of digitization

Now I want to speak about the necessity of digitization. Since 2001 East Asian Folktales Database Committee has made database of 63,763 folktales(61,763 Japanese tales,1,055 Korean, 465 Chinese) using FileMaker. Our committee began with converting analog audio recordings into digital data for conservation.

In Japan not only the tellers but also the researchers have become old and if the researchers are hospitalized or die, we often lose the means to maintain their recorded tapes. We cannot know where their tapes are kept and if we find them fortunately, it often happens that we cannot get the information such as the name of the tellers, the place of recording, etc.

During the 20 century the folktales were conserved in writing like books or magazines and the audio data was considered as secondary. After they were published, the recorded tapes were often forgotten and rejected.

Of course there are many prudent researchers who have conserved their tapes with their research information cards. But even in this case difficulties remain. Especially in Japan, as almost all the researches have been done personally and the researchers have their original ways of keeping records, it often happens that no one else can identify the contents.

So without their personal instructions, it is very difficult to identify the audio recordings exactly.

For this reason we hurried to digitize the analog audio data with the collaboration of folktales researchers to provide their precious analog tapes.

3 Our strategy and methodologies

By support of the Japan Society for the Promotion of Science(JSPS) Japanese Folktales Database Committee (JFDB, predecessor of EAFDB) has started a project to digitalize analog recordings of folktales, asking researchers to lend their precious audio datato JSPS. JFDB has changed analog recordings into digital data and

returned the both type types of data to them after conserving the digital ones.

While proceeding with this service, we found the necessity to make a database which enables us to search required data easily.

And we started to input metadata concerning the tellers, researchers, tale types, etc. using Excel.

In 2001 the capacity of our HDD was about 20G and USB flash drives were not large enough. So we have used CD-R to conserve audio data.

It was very lucky for us that the HDD capacity has increased rapidly. Nowadays a common HDD has capacity of 1 terabyte or more.

As audio compression techniques have progressed and MP3 became a stable format, we began to make our database using FileMaker.

Now our database with 63,763 entries can be conserved in one small 64G USB flash drive. It was unimaginable when we started our project in 2001.

4   Metadata standards and application software

EAFDB with 63,763 entries (61,763 Japanese data,1,055 Korean, 465 Chines) enables us to browse and search materials by 42 metadata:

① data number ②CD number ③title ④title given by the teller ⑤name of teller/informant ⑥ name in Kana ⑦date of birth ⑧ethnic group ⑨address of teller ⑩place of birth ⑪date of research ⑫place of research ⑬name of researcher ⑭organization ⑮original tape number ⑯owner of original tape ⑰classification ⑱tale type by Taisei (Japanese tale type catalogue 1, SEKI:1978-80) ⑲tale type number by Taisei ⑳tale type number by Aarne-Thompson ㉑tale type by Tsukan (Japanese tale type catalogue 2) ㉒tale type by Okinawa Tsukan (Japanese tale type catalogue 3) ㉓legend type (Japanese tale type catalogue 4) ㉔tale type by IKEDA (Japanese tale type catalogue 5) ㉕Korean tale type (Korean tale type catalogue) ㉖ tale type by Ting (Chinese tale type catalogue 1) ㉗tale type by Kim (Chinese tale type catalogue 2) ㉘tale type by Eberhard (Chinese tale type catalogue 4) ㉙audio data size ㉚permission for open access ㉛permission given by ㉜beginning formula ㉝ending formula ㉞listener's response ㉟notes on transmission ㊱transcription ㊲ transcription 2 ㊳comments ㊴key words ㊵summary ㊶original ㊷references

These 42 metadata are necessary only to our EAFDB and we can reduce its number for international database.

5　Fundamental Problems of EAFDB

　EAFDB has two fundamental problems.

5-1 First problem is errors of input data.

　Data input has executed by about 20 persons who have excellent knowledge, experience, and interest in folktales. They have a well unified manual. But their abilities, experience, knowledge, and interest are diverse and their decision as to what to input may also be diverse.

　For example, the selection of key words is likely to be influenced by their individual interest. It may also be affected by their abilities, experience, and knowledge. For example, among the workers there are specialists of Japanese folktales whose interest in international comparison is not as much, and specialists of Chinese folktales whose knowledge about Japanese and Korean folktales is not enough. And it always happens that we make simple errors while inputting data.

　Indeed we have a lot of errors in EAFDB.

5-2 Second problem is a regional bias.

　In Japan as almost all of folktale researches are executed by personal efforts, regional bias of collected materials is inevitable. In an area where there was a good researcher, we can find rich and high-quality data, but where there were no researcher we cannot find any data.

　Concerning the folklore research, the same phenomenon can happen anywhere all over the world, but in Japan the case is serious.

6　How to resolve the problems

　What can we do to resolve these two problems. I am very optimistic about this.

6-1 Concerning the first problem, we can correct our errors whenever we find them.

　As EAFDB has not enough financial support, only one excellent scholar, called Miss IWAKURA Chiharu, has checked the errors for almost 15 years.

　She is also our only program designer who has accomplished and improved EAFDB program using FileMaker.

　In order to check errors, FileMaker and Excel are very convenient softwares.

Even if we have an excellent final checker as Miss IWAKURA, a lot of errors remains in EAFDB.But if we design a good and simple program, we can correct errors any time and improve our databases to have fewer errors.

6-2　Concerning the second problem, EAFDB has a solution.

　Almost all of the EAFDB data are audio data now. We can add printed data to EAFDB .

In Japan the folktale researchers have published the results of their researches in books, magazines and pamphlets. Printed materials remain long time. So it is better to digitize the analog audio data first, which is likely to disappear in the near future. But in order to solve the regional bias we should digitize printed materials.

And we must continue, of course, our researches in the data lucking areas.

As our folktales researches are endless, we can continue our researches and input their results as much as possible.

7　language, platform, and possibilities for developing a common working model

7-1 language

Since 2011 EAFDB has started to make a database which enables us to compare the folktales of China, Korea and Japan. Almost all the data of this database are in Japanese now, but in future we are aiming for a database accessible in at least four languages; Chinese, Korean, Japanese and English.

This project seems like a dream, but it is not so difficult to accomplish. Because computer technology is already globalized enough.

If we use computer to make a database in any language, we can make it easily into one accessible in multiple languages.

At least Excel and FileMaker, which we use, are multilingual and if we add new data in another language, existing data are not affected.

In order to test it, I would like to show you a small Korean-Japanese database of Korean data made by Prof. In-Hak Choi. This database is an incomplete one with metadata labels are shown only in Japanese. But as this test shows us, it is not difficult to search data by entering Korean or Japanese.

So we are optimist about Language and possibilities for developing a common working model.

If we have enough multilingual support from the researchers and enough time and finance, we will be able to translate our database into a multilingual one.

As for language and platform and possibilities for developing a common working model I think as follows.

First of all each researcher can use their mother language to make their own database of folktales. The computer technologies are multilingual so it does not cause any problem.

7-2　Platform

But the common platform is important. We need international agreement about this. The most important thing is to decide the minimum of common metadata on the

international basis.

And if we respect this minimum standard, we can add any regional metadata as many as we need for regional database. It will not cause any problems as long as it is respected. The capability of computer today, and in future, can enable us to build an international database by compiling and combining all the regional data around the world.

We can realise an international database if we respect our minimum standard.

7-3 Application Software

Concerning application software we must choose highly versatile and relatively inexpensive one. As I do not have any other software but FileMaker, I have no other way than to say FileMaker is useful. But I believe it's very important to input fundamental data of each country in one simple and common software used around the world, like Excel.

Of course each researcher in any country can design their original program most suitable for their country using original software, but if the data in their database are incompatible, It may be difficult to incorporate them for international comparison.

In order to assure the possibilities for developing a common working model, the compatibility of data is indispensable.

Fortunately the FileMaker is highly compatible, EAFDB is ready to provide for free our database with 63,763 data to everyone.

As this database with audio data can be stored in a 64G USB flash drive, I can send it to anywhere all over the world.

8 Internet Plus' strategy

Now I would like to talk about our Internet Plus' strategy.

EAFDB aims to present all of the data to everybody on our internet site, using FileMaker Server. But 63,763 data are too many for general visitors. This database is intended for researchers now. We also have some technical problems to solve. To select suitable data and information for general visitors, we must consider their needs and purpose.

It would be more effective to limit the number of data according to their interest.
We have repeated trials and errors to revise our homepage.
http://minwadata.fm.senshu-u.ac.jp/EastAsiaMinwaDB/IndexEAMinwaDB.html

8-1 EAFDB Internet Plus' Strategies 1

We are very sorry to say that this site is only for Japanese-speaking people for now.
We also have Chinese, Korean, English, French, German and Spanish pages, but they

present only our activities and some sample data.

On the other hand on Japanese page we present several trials. So I would like to introduce one or two of them.

Here, the most important theme for us is how to transmit our oral tradition to the coming generation.

Japanese primary education is conscious of the importance of our oral tradition and children have several occasions to listen to folktale-tellers in their schools, kindergartens, nursery schools, libraries, etc.

Local governments and tourism organizations are also eager to revitalize their regions by using folktales.

And there are many people who want to tell folktales to children and adults. But almost all of them have no experience of listening to the traditional folktales in their childhood. They know folktales only through illustrated books or collections of stories. We would like to call them "new-type" here. Using Internet,EAFDB could encourage "new-type" and the regional activities on business, education, tourism, etc.

We are convinced that our database is a useful and encouraging guide for them.

8-2

For example we are introducing several versions of Momotaro (well known Japanese version of Jack the Giant Killer).

We can find almost only one type in illustrated books, but we have rich regional variants of this tale in Japan.

It would be better for the tellers in Okayama Prefecture to tell an Okayama version of Momotaro to the Okayama children in Okayama dialect.

It often happens that the tellers are obliged to choose the stereotype stories and to tell them in common language. It is important that tellers can tell the local folktales in their own local language.

8-3

Fortunately we can find excellent local folktales tellers well motivated to transmit traditional folktales of their regions. So we are trying to introduce their activities on video.

Japanese folktale tellers often introduce new techniques and dramatic performances to attract their listeners just like American storytellers, but sometimes their stories become stereotyped because of their techniques or performances.

We would like to present to them occasions to realize that the regional variety and richness of language of folktales are much more important than techniques.

8-4

With EAFDB we could do other things, I suppose. As we all know, internet technology is ubiquitous.

Everybody can carry, transmit and receive information anytime and anywhere in the world. It's very easy. Everybody can share information on internet. And, following certain rules, we can use them for research, education, regional activities, tourism, business, etc.

We can get folktales information on folktales and folklore anywhere in the world and we can learn and transmit them and find new possibilities in them.

9　From folktale database to oral tradition database

As EAFDB focuses on folktales, I especially have presented characteristics of EAFDB as folktale database, but it is also suitable for oral tradition research.

Folklorists have tried to collect not only folktales but also everything about the informants' folk life through their researches.

9-1 classification of folksongs

Among the metadata of EAFDB we have 'classification' and there we have nine categories; 1)folktales, 2)legends, 3)anecdotes, 4)myths, 5)superstitions and proverbs, 6)folksongs, 7)word plays and enigmas, 8)folk customs, 9)others

Using these categories, let's try to search other data besides folktales.

For example we have 2249 data about folksongs.

If we search for data with "marriage" as a key word, we can find 22 data about folksongs concerning marriage.

And making a search for superstitions and proverbs, we get 1609 data. And if I add "rain" as a key word, we can find 41 data.

When we search for entries with "marriage" as a key word without specifying category, we find 660 data. We get 1123 data for "rain".

If we put both "marriage" and "rain", we will get 31 data.

Though audio data of EAFDB are centered on folktales, they contain a lot of folklore information about informants.

But concerning the printed data, it happens that they might be strictly specified to folktales, So it is necessary to pay much more attention when we input them into oral tradition database.

9-2　Uses of 90 sectors of Classification of EAFDB

In the classification table of EAFDB there are 90 sectors and we can add items to integrate other databases if necessary.

For example we can add song categories to our Folksong's column, such as ①61 work

songs, ②62 ritual songs, ③63 Political satire songs, ④64 Daily life songs, ⑤65 Love songs, ⑥66 Children's songs, ⑦67 historical legend songs, ⑧68 Others, according to the folksong classification by Wu Chao（吳超）

Wu Chao has classified children's songs into (A)lullaby, (B)counting rhymes, (C)questions and answers song, (D)play song, (E)chain song.

So we could add subclassification as follows; 66a lullaby, 66b counting rhymes, 66c questions and answers song, 66d play song, 66e chain song.

Just like folksongs we cold add 41Origin of Universe(宇宙起源母題),②42Origin of Human(人類起源母題) , ③43Origin of Culture(文化起源母題), ④44Origin of Animals & Plants(動植物起源母題), etc.

And EAFDB has only audio data and written data at present, but we could add video data to them as well.

Anyway I am sure that EAFDB could be integrated not only into any other folktale database but also into any oral tradition database, and vice versa.

10 Conclusion

At the end of my speech on Digitizing Oral Tradition: Strategy, Practice, and Collaboration, I would like to summarize five points as conclusions.

1.

It is not difficult to make an excellent digital archive of oral tradition with international collaboration.The global network of computer and internet might realize this dream. We can expect the performance of computer will continue to improve so that transnational and multilingual collaboration will be promoted smoothly.

2.

In order to realize our international collaboration, the first thing for us to do is to select minimum standard of metadata. We should minimize the number of the common international metadata. Researchers of each country or each region might add their own metadata if necessary.

We should fully respect each researcher's own choice and ideas and their freedom to do so should be assured.

3.

Among the metadata, I think that the most important one is the Tale-Type and the Type index established by Anti Aarne and known as AT or ATU number.

As everybody knows today, the classification by Aarne and Thompson has been made on the basis of the European Folktales and it is not enough to classify the folktales of other areas and cultures. But we have a great history and rich results of research with

this classification as de facto standard.

Researchers in each region or each country could use their original or regional tale-types as well as AT or ATU, such as the Korean Type-Index made by In-Hak Choi or Japanese one by SEKI Keigo.

4.

The next important thing is to select application software for international digital archive, but I think it is not necessary to choose one in a hurry. The point is to ensure the compatibility of the input data of each region and country.

If we use simple and compatible software like Excel or FileMaker, for example, it will not cause any problem to design and construct an international database by integrating all date of the world.

5.

Finally I would like to add a few words concerning how to conserve audio data. In this case also data should be compatible.

For the website of EAFDB we are using MP3 files now, but considering future innovation of compression technology we also conserve WAVE files. We need much more space in HDD for WAVE files, but the capacity of HDD has been increasing with unimaginable speed.

Nowadays we can conserve WAVE audio files very easily.This might be very important not only for audio data but also for video data. Compression technic improves day by day, but we should conserve carefully original data uncompressed.