

FINAL ACTIVITY REPORT

Guidelines on how to complete the activity report are included in italics.

Remember that this report will be made available on your project page on the GBIF website and therefore should not include email addresses, unless you have permission from all mentioned in the report that their email information can be published.

Contents

Project information	1
Executive summary	2
Project objectives	2
Activities	3
Deliverables	6
Calendar of activities	9
Project communications and visibility	10
Final evaluation findings and conclusions.....	11
Sustainability plans.....	12
Recommendations and lessons learned.....	13
Annex – Sources of verification	13

Project information

Main contact person and role:	Chin Sung Chang, Coordinator
Institution/network/agency affiliation:	Department of Forest Science, Seoul National University
BIFA Project ID:	BIFA3_14
Project title:	Data mining of historical herbarium specimens from the Korean peninsula using the Brahms database: A look at the biodiversity informatics project of North Korea
Start date and end date of the reporting period:	April 1, 2018 to March 31, 2019
Country in which the activities take place:	Republic of Korea

Executive summary

Provide a brief explanation of the project and its implementation, the context and the approach taken for the final evaluation, and a summary of the objectives achieved, lessons learned and conclusions.

Despite their abundance in collections, the needed data in the Korean peninsula including North Korea are inaccessible or insufficiently integrated to foster query-based inquiries, and are not applicable to studies at a global scale. Many foreign herbaria (A, E, TI, KYO, and others) constitute a large fraction of South and North Korean botanical collections for more than 100 years. Nevertheless, local data is often inaccessible from outside countries and the available data in both countries are not well managed and formatted thus far. The great majority of vascular plant species data (ca. 607,514 primary occurrence data) which is currently available as occurrence records in GBIF, has not been georeferenced (less than 1%) even in South Korea. This dataset is presumed to be orphaned because its data host no longer has the resources or desire to host it online. Moreover, completeness and taxonomic and geographic precision of primary occurrence data is not quite satisfactory. In some cases, the data host even loses the dataset (e.g. due to server failure), and no other backup exists. With nearly 50,000 specimens including the data about specimens stored at foreign herbaria, we have a comprehensive chronological, historical, taxonomic, and geographic coverage of Korean plants including those from inaccessible areas, such as North Korea.

We published a dataset (13,981 specimens of E.H. Wilson's, E.J. Taquet's, U. Faurie's) about the Korean peninsula through the GBIF network worldwide. On the other hand, T.B. Lee Herbarium (SNUA) has been devoted to the study of the Korean peninsula flora since 2011. Our specimens at SNUA have been collected as vouchers in floristic studies of South Korea since 1950. This project included another dataset (18,106 SNUA specimens) about woody plants collected from South Korea and published another dataset through the GBIF network. The first phase and scope of this project integrated the BRAHMS software to allow queries of foreign herbaria historical records, generate specimen georeferenced data, and photo images about the north and South Korean vascular plants. This project handled occurrence records and made them available through the GBIF web services and download files.

Project objectives

This section should include the list of objectives included in your original project proposal, stating for each how far you advanced towards their achievement. Also, include any additional objectives that were defined during the implementation of the project. In the event of unexpected challenges prevented you to reach a planned project objective, please provide detailed explanations and indicate how you plan to reach these objectives post project.

This proposed project had three primary objectives:

- 1) To establish a database, a network of specimens collected by U. Faurie/E. Taquet and E.H. Wilson deposited at foreign herbaria about North Korea that will develop and share tools,
- 2) To provide photos of specimens deposited at E, KYO, and A and to make the data accessible through the GBIF website,
- and 3) To provide data of relatively recent specimens housed at SNUA about woody plants in Korea.

Activities

Please indicate the status of the activities as outlined in the project proposal, at the time of final reporting. The table below should be completed in the same way as in the full proposal but should include information and updates on the status of each activity.

In the event of unexpected delay please provide detailed explanatory notes and indicate planned completion date after the end of the project. Add as many rows as needed.

In the event of any additional activities having being completed during the implementation of the project, please add rows as required.

Description of activity	Partners involved	Contribution of activity to goals listed in table 4.3	Status of activity as of final reporting Completed? Yes/No	Explanatory notes, inc. planned completion date if necessary	Source(s) of verification
Digitizing and publishing georeferenced species occurrence data based on specimens held in Asian collections					
Provide data about 13,981 specimens collected by E. H. Wilson & Faurie/Taquet through the Korean peninsula (including South and North Korea)	Dr. H. Kim	Mobilize existing knowledge within the Korean Peninsula	Yes	13,981 occurrences Completed to input the records in the database and loaded the data in the GBIF	Some of the photo images are currently presented in our website (http://hosting03.snu.ac.kr/~quercus1/virtualherbarium.htm) with the GBIF network
Provide data about 18,106 specimens (of woody plants) collected from 1953 to 2017 in South Korea	Dr. H. Kim	Mobilize existing knowledge within the Korean Peninsula	Yes	18,106 occurrences, Records input completed and data cleaning process will be made	
Compiling inventories of biodiversity data holdings (for example, by implementing metadata catalogues)					

Description of activity	Partners involved	Contribution of activity to goals listed in table 4.3	Status of activity as of final reporting Completed? Yes/No	Explanatory notes, inc. planned completion date if necessary	Source(s) of verification
Develop the online flora using the scratchpad* ¹	Drs. H. Kim & Haining Qin	Deal with biodiversity data at the appropriate scale in Asia	YES (part)	Data about woody plants were completed. Distribution maps of selected herbaceous taxa were presented and the rest will be completed no later than Dec. 31, 2019.	http://florakorea.myspecies.info/en http://kpf.myspecies.info/
Preparing data papers					
U.J. Faurie/E. Taquet and E. H. Wilson from 1901 to 1917 in Korea	Dr. H. Kim	Mobilize existing knowledge within the Korean Peninsula	NO	Prepare the paper about Wilson data	Will be submitted within three months
SNUA collections about woody plants from 1953 to 2016 in Korea	Dr. H. Kim	Mobilize existing knowledge within the Korean Peninsula	NO	Prepare the paper about SNUA woody plant data	Will be submitted within three months
Other activity types					
A checklist of North Korean vascular plants * ²	Dr. H. Kim	Mobilize existing knowledge within the Korean Peninsula	NO	Will be finished with six months as the top priority	

*¹ Develop the online flora using the scratchpad

We have presented the distribution maps of woody species recorded in Korea and are doing in the middle of presenting the distribution map of selected herbaceous taxa based on old historical collection in the scratchpad. We will build up distribution maps of all herbaceous plants based on T.B. Lee herbarium specimens as well as historical collections within a year (December 31, 2019).

*2 Produce a local annotated checklist

Four years ago we produced a provisional checklist of Korean Peninsula flora. The completed checklist (a provisional checklist of KPGF) included all of the plant species distributed both in South and in North Korea (see attached file). Over the past 4 years we spent to present a checklist for China, Japan, Far East Russia, and Taiwan (see attached the examples) in order to build up the database about Northern Asian flora project.

The preset approach has been structured as a comprehensive work on presenting a checklist about North Korean taxa using nine volumes of 'Flora Coreana (ed. Im, R.J., 2001)' and other recent literature (eg. Endemic plants of Chosen, La, U.C et al., 2015). The BRAHMS has been used to set to list synonyms as well as accepted names and to accept the links during data entry. We usually adopt the Plant List which provides the accepted names with links to all synonyms. This approach reveals the priority setting on completion within six months. The checklist will be first published through the GBIF in 2019.

Deliverables

This section should summarize the project deliverables completed by the final reporting date, with a description of the associated outputs. Please highlight any changes from the original plans provided in the full project proposal.

In the event of an unexpected delay, please provide detailed explanatory notes and indicate planned completion date. Add as many rows as needed.

In the event of any additional deliverables having being completed during the implementation of the project, please add rows as required.

a. Data

Details of datasets mobilized and/or pending mobilization as an outcome of the project: Please use list from mid-term report and update this as at final reporting.

If the dataset is not yet published, please indicate it as “not published” and provide a detailed explanation and expected date of publication. Add rows as required.

Title of dataset	Taxonomic/ geographic scope	Approximate number of records (specimens)	Current format (e.g. undigitized, digitized)	Status of dataset: Published/not published – inc. date/expected date of publication	Explanatory notes	DOI or URL
------------------	-----------------------------------	--	---	---	-------------------	------------

Flora of the Korean Peninsula (Faurie/Taquet data and E.H. Wilson data)	Vascular plants/South Korea including some parts of North Korea	6463+6433(12,896) and 1,085 [total 13981]	photos (5,567) – low quality of digitized	published	Types and other information with photos were included	https://www.gbif.org/dataset/65bdd8e3-a27b-4b88-998d-dfb27d528206
T.B. Lee Herbarium Vascular Plants Collections (SNUA specimens)	Vascular plants/South Korea	18,106	Undigitized	published		https://www.gbif.org/dataset/266726e3-9f9b-44ed-86f2-d590f6c9626e

b. Other deliverables

Describe other deliverables (e.g. publication of data papers, catalogues, reports etc.). produced and/or planned to be produced as a post-project deliverable. Please provide indicative dates/estimated time for completion for planned post-project deliverables.

Please provide links in the sources of verification. Attachments should be provided in the Annex.

Name and type of deliverable	Status of deliverable Published/not published – inc. date/expected date of publication or estimation of time for completion	Explanatory notes	Source(s) of verification
Preparing data papers	Not published yet (June 2019)		
A checklist of North Korea vascular plants	Not published yet (Dec. 2019)	The present approach has been structured as a comprehensive work on presenting a checklist about North Korean taxa using nine volumes of	

		<p>'Flora Coreana (ed. Im, R.J., 2001)' and other recent literature. BRAHMS has been used to set to list synonyms as well as accepted names and to accept the links during data entry. We usually adopt the Plant List which provides the accepted names with links to all synonyms. This approach reveals priority setting on completion within six months. The checklist will be first published through the GBIF in 2019.</p>	
--	--	--	--

Calendar of activities

The calendar should be completed in the same way as in the Full Project Proposal (4.6) but should include any changes. Please provide reasons for any changes in the Notes column in the table below.

Proposed dates	Activity	Lead partner	Notes
June 2018	Attendance of project team member at BIFA Capacity Enhancement Workshop	Hui Kim	Project leader took part in the meeting and gave a presentation, but Dr. Kim attended the workshop for one week.
April-Sept 2018	Data input/Bar coding of SNUA specimens	C.S. Chang	April-Sept 2018. Completed the input
July 2018	Visiting TI(University of Tokyo Herbarium) and KYO(Kyoto University), five-day work *1	C.S. Chang/S.Y. Kwon/ San Kang (undergraduate student)	July 2018. Visited Kyoto University and the University of Tokyo for five days
Sept. 2018	Georeferencing and data cleaning*2	C.S. Chang/H. Kim	Sept 2018
Jan-March 2019	Information dissemination and preparation of data papers	C.S. Chang/H. Kim/Haining Qin	Jan-March 2019 (plan)

a. General explanatory notes

Visiting TI and KYO *1

Collection number of most specimens collected by Faurie/Taquet at E, A, P, K and KYO (some missing) were confirmed easily using our acquired photos, but majority information on specimens (except type collections) deposited at TI (University of Tokyo, herbarium) and KYO (Kyoto University) were obtained without collection number (“s.n.”) and stored in our database. Many additional photos of these specimens (ca. 1,100 additional photos of Faurie/Taquet’s specimens) were collected with our current visit to two Japanese herbaria.

1st MAP symposium *2

Presentation title: Data cleaning process in historical collections -Old labels give a clue for new science. A case of the Korean peninsula and Northeastern China.

Chin Sung Chang (韩国首尔大学)



亚洲植物多样性编目国际研讨会

第二轮通知

一、会议信息

会议时间：2018年9月25-26日（9月24日注册）

会议地点：北京

会议语言：英语

二、大会报告

1. Mapping Asia Plants: progress and outlook

马克平 (中国科学院植物研究所)

2. Mapping biodiversity patterns across Southeast Asia

Alice C. Hughes (中国科学院西双版纳热带植物园)



3. Data cleaning process in historical collections - Old labels give clue for new science. Case of the

Korean peninsula and Northeastern China.

Chin Sung Chang (韩国首尔大学)

4. Facing the Anthropocene challenge: challenges and opportunities of species and vegetation databases

in addressing ecological problems in a non-analogue future

Alejandro Ordonez (丹麦奥胡斯大学)

地 址: 北京市海淀区香山南辛村20号水杉楼 邮编:100093 电话: 010-62836603 / 6629
Address: 20 Nanxincun, Fragrant Hill, Haidian District, Beijing, China 100093 Tel: +86-10-62836603 / 6629

Project communications

- 1) The following Internal communication was conducted in Feb. 2019.
 - present & publish the results in academic societies/journals (The Korean Society of Plant Taxonomists/Korean Society of Forest Sciences).
- 2)
 - The part of our project was given as a presentation in the international symposium (by the Institute of Botany, Beijing, China) on Sept. 25th and 26th. The title is “Data cleaning process in historical collections - Old labels give clue for new science. Case of the Korean peninsula and Northeastern China”.

Project communications and visibility

Describe the way the results of your project have been and will continue to be communicated and shared with the project stakeholders and broader GBIF community. Please also review the page describing your project available from

<http://www.gbif.org/programme/bifa> . Highlight any additional documents, events, news items or links that you would like to add to your page and provide links/attachment in the Annex.

An evaluation of the project activities

The herbarium specimens in SNUA have been collected over the past 50 years from South Korea geographical areas. We made the data globally available and paid attention to use both the Korean and the English geographical names. A thorough input was conducted in order to compile digital occurrence records corresponding to ca. 500 species. As we planned, a total of 18,106 data was stored in the Brahms database and loaded in the GBIF website. We visited two Japanese herbaria, TI and KYO and obtained additional photos about Faurie/Taquet collections. We completed labeling species name and collection number for each photo taken. It is important to manage carefully these datasets, promoting constant quality improvements. The present study completed our aims to assess the quality of the dataset and records.

Through the project, we were able to obtain more information about Faurie/Taquet collections from Japanese herbaria . This BIFA project experience made us published the remainders of SNUA dataset and another historical dataset about North Korea.

Final evaluation findings and conclusions

This section of the report should cover for example:

- *An evaluation of the project activities and their outputs/deliverables*
- *An assessment of the overall outcomes, impacts of the project and how it contributes to the overall objective of the BIFA programme*
- *Comments on the project implementation and completion, and its efficiency and effectiveness, strength and weaknesses etc.*
- *Any feedback on the project's relevance from the partners and stakeholders*
- *Indications and reasons for any changes which have been made to the project's original plans, and actions to follow-up*
- *The management arrangements for the project, including support from the GBIF Secretariat*
- *Areas of success to build on, after the project's implementation period*
- *Conclusions from your experience during the implementation of the project*

The collective work about historical collections and knowledge of the experts across the foreign herbaria provided access to the information on the Korean peninsula. It was important to learn about how to manage carefully the datasets and to promote constant quality improvements through this project.

Although a large variety of tools was available to support data transformation and data cleaning tasks, we learned about the openrefine tool through the workshop. We use both the Brahms database and this openrefine about duplicate elimination, specific cleaning phase, and data analysis.

Concerns about the quality and 'fitness for use' of the data mobilized have grown over the years. Therefore, the present study aimed to assess the quality of the dataset and records and make us keep all appropriate documents that prove the veracity of data. Through the project, we were able to obtain more information about Faurie/Taquet collections from the Japanese herbaria. This BIFA project experience made us publish the remainders of SNUA dataset and another historical dataset about North Korea.

Through this study, we would like to show a clear need for institutionalization of a data publishing

framework in Korea. National Institutions here have been slow to make data available due to a lack of awareness, a lack of technical capacity and lack of adequate mechanisms. This study has clearly demonstrated the great value of the GBIF mediated data in terms of this aspect nationally.

The interest in publishing data through the GBIF is increasing among Asian researchers. The increase in data publishing activity may call for the creation of a GBIF node in northeastern Asia including North Korea in the future to support biodiversity experts in international data work. We are integrating Asian species checklists (northeastern China, far eastern Russia, Japan, and Taiwan as well as the Korean peninsula) and will provide the critical taxonomic foundation for the development of Asian biodiversity databases here using outside funding (eg EABCN project from the Korea National Arboretum).

Sustainability plans

Please provide a description of how the partners involved will build on the results of this project in their future work. This could include future collaborative activities, such as plans to complete any unfinished project activities and how the future impact of the project could be monitored and/or measured.

We focus on spatial gaps in biodiversity information within Asia. Overcoming the gaps in biodiversity information about North Korea has proved a serious challenge for us. Our target data set as the second phase of this project includes ca. 40,000 historical collections from North Korea deposited in Japan and integration and mobilization will be conducted. Also, we will expand the data set to include specimens (ca. 1,000) collected from North Korea in Chinese herbaria (PE and IFP) and far eastern Russia (VLA, Herbarium of the Institute of Biology and Soil Science, Far East Branch of RAS) as a part of the project.

Our project seeks to improve the information available to other countries, in turn helping them to monitor and advance progress towards all other goals. The imbalance links to the South and North Korea are the massive gaps in knowledge about the details of the diversity and distribution of units of biodiversity of vascular plants. A major important context for the current project is the international platform on northeastern vascular plants. The platform includes capacity building and the generation of data and knowledge among China, Russia, Japan as well as Korea. Another important context for our project is to solve the conflicting scientific names currently used among the five countries. Biodiversity tends to be understudied in spite of its richness, and for which the huge volume of existing biodiversity data are still not available to each other. With regards to the conversion and validation of taxonomic information (e.g. scientific names) the challenges are more complex in eastern Asia. Therefore, the study about taxonomical comparison with the research fund from Korea National Arboretum has been undertaken for three years thus far. The GBIF project could integrate and coordinate the effort for broad and effective execution of this initiative.

After the project is over, we will share and keep hosting these data through the scratchpads biodiversity online. Working in this way may allow large and long term goals to be achieved in these small stages

without losing sight of our vision. Data quality is a key issue for the ongoing success of digital biodiversity platforms. Therefore, we are more interested in a recent issue ‘fitness for use’.

It is necessary to find strategic solutions through GBIF regarding this matter. We continue our data mining project about the northeastern Chinese specimens including historical collections with Chinese collaborators. We will search for another funding on northeastern China including Kormaov’s collections (1895 to 1897 in China, Korea, and Russia)

Recommendations and lessons learned

This section should describe your experiences that could help in designing and implementing biodiversity mobilization projects more effectively, including the best practices to adopt and the pitfalls to avoid.

The following comments were presented in our midterm report, but we hope that the GBIF secretariat considers the following problem existed in the KBIF. Please check the dataset presented by the KBIF this year as well as the previously published dataset and present an appropriate recommendation on this matter.

The National Science Museum by regulations of the Ministry of Science, ICT and Future Planning of the Korean government takes part in the GBIF global network as a national member. Although a considerable amount of research funding are operated by these Korean National Institutions including Korea National Arboretum of Korea Forest Services, Korea Institute of Science and Technology information, KBIF National Data Repository, and Korean Natural History Research Information System, data gaps and data immobilization in digital accessible information in the Korean peninsula have hampered prospects of safeguarding biodiversity. Unfortunately, the Korean government sector is not actively engaged in database disclosure and data creation.

A process of validation and data cleaning, including a retrospective georeferencing process, has not quite performed yet by the KBIF. A dataset composed of more than 1,000,000 records was compiled in Korea, but was considered of poor quality because none of the records had properly georeferenced

The GBIF website has already indicated major problems of taxonomic precision, geographic precision, temporal precision about data published by KBIF. The KBIF is necessary to present how to manage orphaned data (over 90% of the published data), and how to process data cleaning, such as detecting and correcting corrupt or inaccurate records in order to solve these problems within a certain time period. The GBIF should recommend that the KBIF needs to establish linkages with the databases of the GBIF and high qualified data made available will not be subject to a limitation on dissemination to other countries. At present, the KBIF is internationally isolated and lacks data creation and sharing.

Annex – Sources of verification

Sources of verification are for example links to relevant digital documents, news/newsletters, brochures, copies of agreements with data holding institutions, workshop related documents, pictures, etc.



Fig.1. Working at KYO on June 28, 2018.



Fig. 2. Data input at SNUA on Sep 18, 2018.



Fig 3. Join the workshop on June 5th at PE

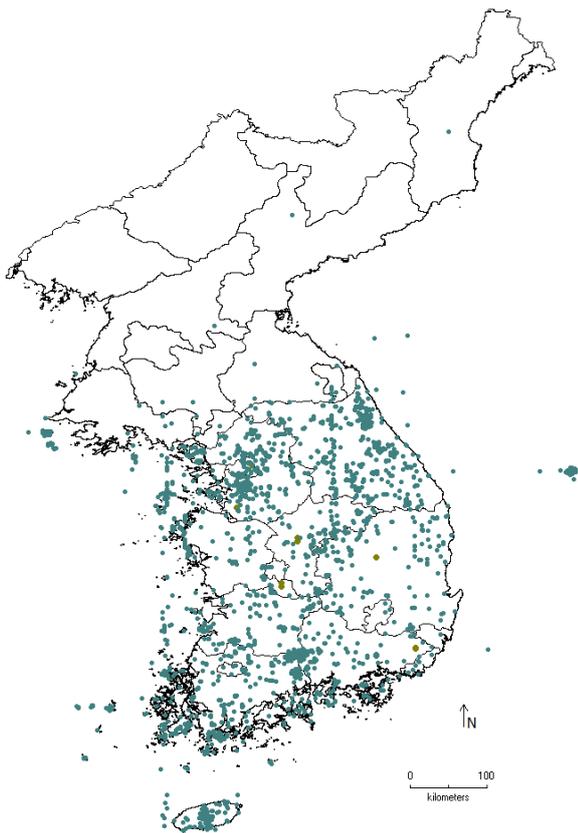


Fig 4. Collection sites represented by 18,000 specimens deposited at SNUA from 1951 to 2017 with a map.

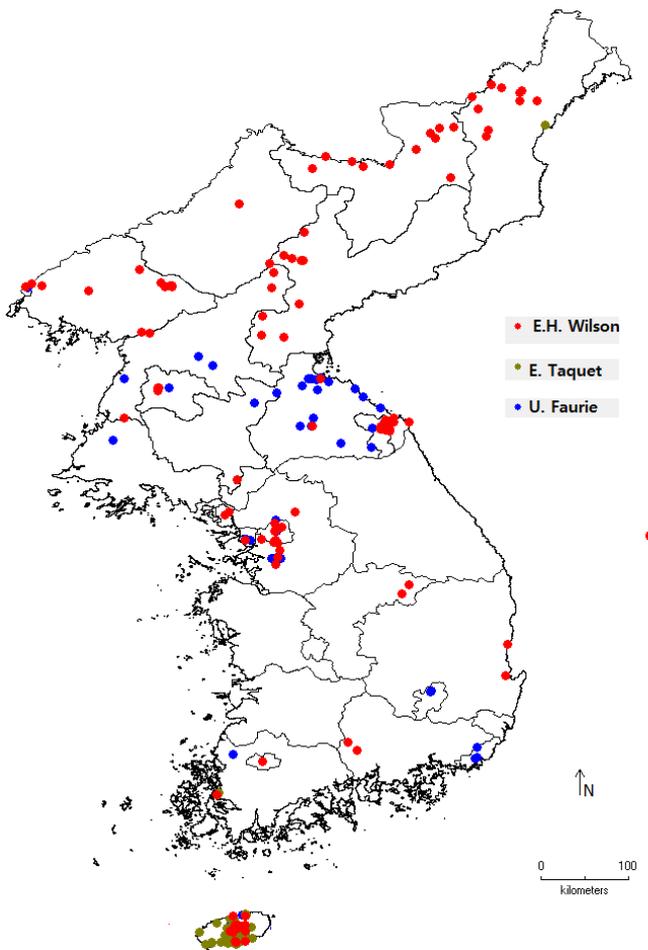
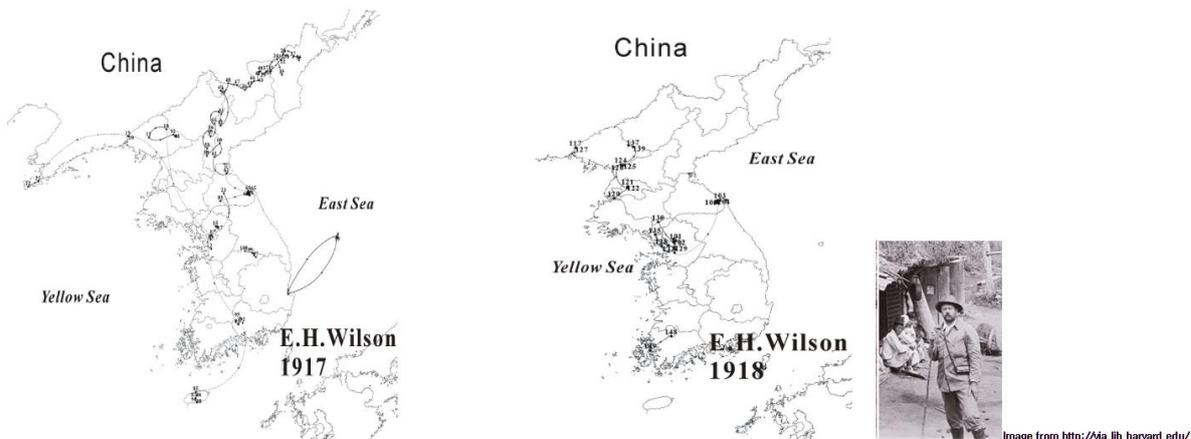


Fig. 5. Historical records of E.H. Wilson, U.J Faurie, and E. Taquet deposited at several foreign herbaria



Fig. 6. Some examples of ca. 6,900 photos taken from several herbaria

E. H. Wilson's Collections



Specimens

collector	number	dups	species	day	month	year
Wilson, E.H.	9302	A	Abies holophylla Maxim.	11	10	1917
Wilson, E.H.	9090	A	Abies holophylla Maxim.	8	9	1917
Wilson, E.H.	8912	A	Abies holophylla Maxim.	15	8	1917
Wilson, F.H.	8781	A	Abies holophylla Maxim.	7	7	1917

Fig. 7. Virtual herbarium for photos of Wilson's, Taquet's, and Faurie's collections deposited at foreign herbaria (<http://hosting03.snu.ac.kr/~quercus1/Wilson.htm>; <http://hosting03.snu.ac.kr/~quercus1/Faurie.htm>; <http://hosting03.snu.ac.kr/~quercus1/Taquet.htm>)

학회소개 학회지 학술대회 회원명단 계산판 갤러리

한국식물분류학회

학술대회

학술대회발표

초록접수

학회 > 학술대회 > 초록접수

초록접수

[초록정보]

초록제목	GBIF에 출판된 국내기관 자료의 현황과 자료 유용성 분석 - 사적 자료의 발굴 및 GBIF 등재
초록내용	GBIF에 출판된 국내기관 자료의 현황과 자료 유용성 분석 - 사적 자료의 발굴 및 GBIF 등재 김 휘 ¹ , 권신영 ² , 장진성 ^{P2} ¹ 국립목포대학교 한약자원학과, ² 서울대학교 산림과학부

생물다양성정보학(Biodiversity Informatics)은 자료의 관리, 발표, 발견, 탐구 및 분석과 같은 정보 해석을 위해 디지털 과학 기술을 적용한 학문으로 주요 결과물은 본포도, 지리적 분포, 분류학(종의 정이명 목록, 체크리스트), 계통수, 보전 및 관리 전략, 생태모델의 미래 예측 등이지만 근간은 정이명 학명을 통한 자료 통제, 자료의 디지털화와 자료의 양, 그리고 자료의 이용 및 플랫폼이다. GBIF(Global Biodiversity Informatics Facility)는 전 세계적 규모의 생물다양성 정보기구로서 현재 10억 건 이상의 생물다양성 자료를 제공하고 있으며, 우리나라는 2001년 가

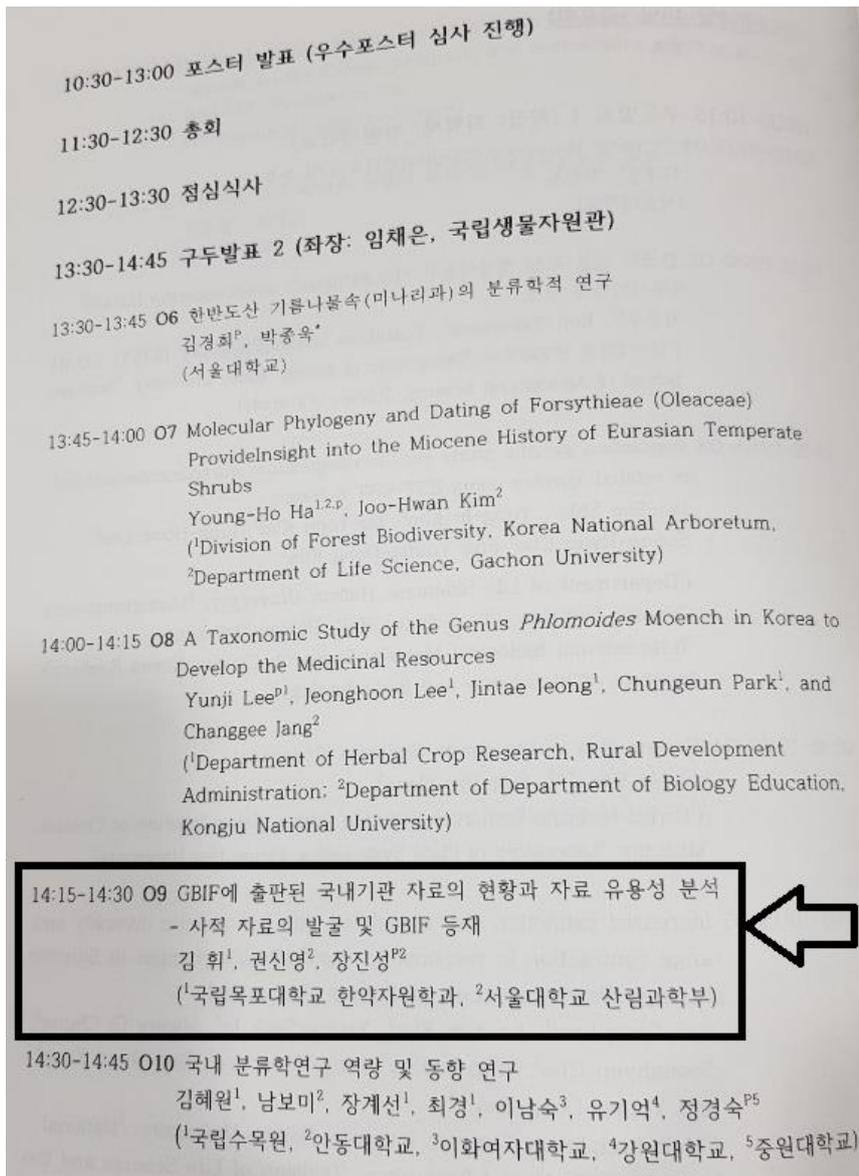


Fig. 8. Presentation in Korea Plant Taxonomy Society (Feb., 14, 2019). Presented in Korea