Digital Resources of Japanese Texts: from a Viewpoint of Digital Humanities

Kiyonori Nagasaki

International Institute for Digital Humanities, Tokyo

and University of Tokyo

Self-introduction:

- Studied Buddhist philosophy/philology in graduate school in University of Tsukuba.

- Experienced network administrator of several Internet providers and universities during graduate school and post-doc

- Taught computer science and ethics in a local university as lecturer/associate professor during 7 years

- Resigned the university and established an independent research institute in 2009 in order to manage a Buddhist database and to form DH community in Japan.

Self-introduction:

- Studied Buddhist philosophy/philology in graduate school in University of Tsukuba.

- Experienced network administrator of several Internet providers and universities during graduate school and post-doc

- Taught computer science and ethics in a local university as lecturer/associate professor during 7 years

- Resigned the university and established an independent research institute in 2009 in order to manage a Buddhist database and to form DH community in Japan.

Range of my recent activities

- Universities and Institutes
  - Senior fellow in International Institute for Digital Humanities
  - Project associate professor of DH in University of Tokyo
  - Leader of a collaborative research group of DH in Kyoto University
  - Researcher in Japanese National Diet Library
  - A member of digitization committee in National Institute of Japanese Literature
  - Co-researcher in Okinawa Prefectural University of Arts
- Academic communities
  - Secretary of Japanese Association for Digital Humanities
  - A member of steering committee of Japanese Association for Indian and Buddhist Studies
  - A member of steering committee of SIG-Computers and Humanities in Information processing Society of Japan
  - Editor of Digital Humanities Monthly (e-newsletter)
  - A member of nominating committee of DH awards
- Research projects
  - A member of SAT Daizōkyō Text Database Committee
  - A member of Indo-Tibetan Lexical Resources project
  - A member of Digital Dickens Lexicon project
- And so on…

Large Scale Digital Resources of Japanese Texts

- Freely browsable on Web

  - National Diet Library Digital Collections
    - National Diet Library
    - Over 500,000 items
    - http://dl.ndl.go.jp/
  - CiNii
    - National Institute for Informatics
    - Over 4000000 articles.
    - http://ci.nii.ac.jp/
  - NIHU
    - National Institute for Humanities
    - Integrating over 100 research databases
    - http://int.nihu.jp/

  - Not freely "available"!

Middle or Small Scale Digital Resources of Japanese Texts (1)

- National Institute for Japanese Language and Linguistics
  - **Balanced Corpus of Contemporary Written Japanese (BCCWJ) (over 100,000,000 words)**
  - **Taiyō Corpus**
  - Bibliographical database of this field
  - …various databases
  - http://www.ninjal.ac.jp/english/products/
- National Institute of Japanese Literature
  - Bibliographical database of this field
  - …various databases
  - https://www.nijl.ac.jp/pages/database/

Middle or Small Scale Digital Resources of Japanese Texts (2)
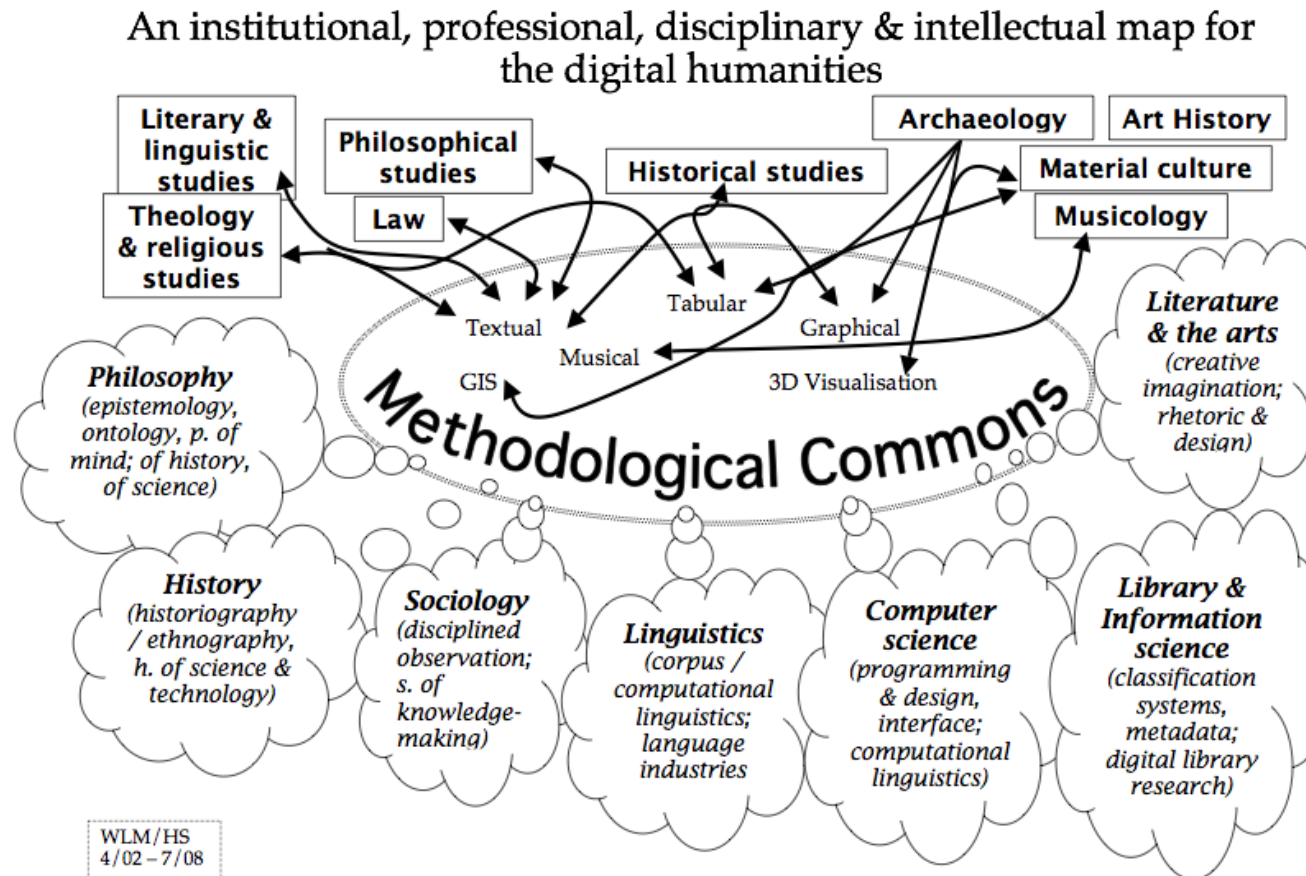
- Various databases in universities and institutes
  - University of Tokyo
    - Historiographical Institute (Japanese History)
      - http://wwwap.hi.u-tokyo.ac.jp/ships/db-e.html
    - Center for Evolving Humanities (Buddhist study and DH)
      - http://21dzk.l.u-tokyo.ac.jp/CEH/index.php
    - … various institutes and faculties
  - Art Research Center, Ritsumeikan University
    - http://www.arc.ritsumei.ac.jp/search_arc.html
  - …

Tendencies:

- Large scale resources:
  - Managed by general organizations
  - Reliable sustainability
  - Simple search function
  - Providing image files, not text data

- Middle or Small scale resources:
  - Managed by research groups or specialized organizations
  - Unreliable sustainability
  - Specific search function
  - Providing not only image file but also tag, specific metadata, and text data in some cases.
  - Providing other specific functions
  - Strong relationship with DH (Digital Humanities)

A typical DH model

- By Willard McCarty and Harold Short

An institutional, professional, disciplinary & intellectual map for the digital humanities

Literary & linguistic studies
Philosophical studies
Historical studies
Archaeology
Art History
Theology & religious studies
Law
Material culture
Musicology

Textual
Tabular
Musical
Graphical
GIS
3D Visualisation

Methodological Commons

Philosophy (epistemology, ontology, p. of mind; of history, of science)

Literature & the arts (creative imagination; rhetoric & design)

History (historiography / ethnography, h. of science & technology)

Sociology (disciplined observation; s. of knowledge-making)

Linguistics (corpus / computational linguistics; language industries)

Computer science (programming & design, interface; computational linguistics)

Library & Information science (classification systems, metadata; digital library research)

WLM/HS
4/02 – 7/08

A viewpoint of digital humanities

- "DH" which has been formed by researchers who had been addressing digitization of the humanities since several decades ago.

- Humanities researchers should be expected to develop the potential for cultural resources.

  ➡ Useful resources for humanities researchers and students.

- DH researchers should be expected to develop the potential for digital cultural resources.

  ➡ Useful digital resources for DH researchers and students.

History: the dawn in Japan

1950's

The Mathematical Linguistic Society of Japan was established.

1960's

National Institute for Japanese Language and Linguistics started to treat Kana and Han characters in computer.

1970's

Large-scale computing in several universities

National Institute of Japanese Literature

National Museum of Ethnology

History: emerging communities

1980's

- Accelerated by IBM PC
- SIG-CH (Special Interest Group-Computer and the Humanities)
  - Under the Information Processing Society of Japan (Mainly aims IT)
- Japan Society of Information and Knowledge
- Japan Art Documentation Society

History: emerging communities of certain field

- Accelerated by the Internet

1990's

Japan Association for English Corpus Studies

Japan Information-Culturology Society

Japan Society for Archaeological Information

2000's

Japan Association for East Asian Text Processing

Japan Society for Information and Media Studies

The Japan Society for Archival Science

Japan Association for Contents History Studies

History: Globalization

2010's

- Japanese Association for Digital Humanities (JADH)
  - Established almost by humanities researchers who have been engaged in DH.
  - Became a constituent organization of ADHO
  - Holding international conference every year
    - Next: 9/19-21 in University of Tsukuba

- International Symposium series of Culture and Computing
  - Managed by IT researchers who are interested in culture
  - Publishing proceedings from Springer or IEEE.
  - Holding intl. conference in Kyoto or East Asia.
- Most of related communities remain

History: large grants by government

- Approximately 1 million dollars per year

1995-1999

- "Computer and the Humanities"
  - Constituted by leading researchers in several universities

1998-2002

- "Restructuring classical studies"
  - A group of digitization played an important role.

2001-2006

- "Grammatological Informatics based on Corpora of Asian Scripts"
  - Carried by Tokyo University of Foreign Studies

History: "Center Of Excellence" grant

- Approximately 1 million dollars per year

2002-

- "21th century COE"
  - Kyoto University
  - Tokyo Institute of Technology
  - Tokyo University of Foreign Studies
  - Ritsumeikan University …

2007-

- "Global COE"
  - Ritsumeikan University
  - Tokyo University of Foreign Studies

Recent DH-like JSPS-funded projects

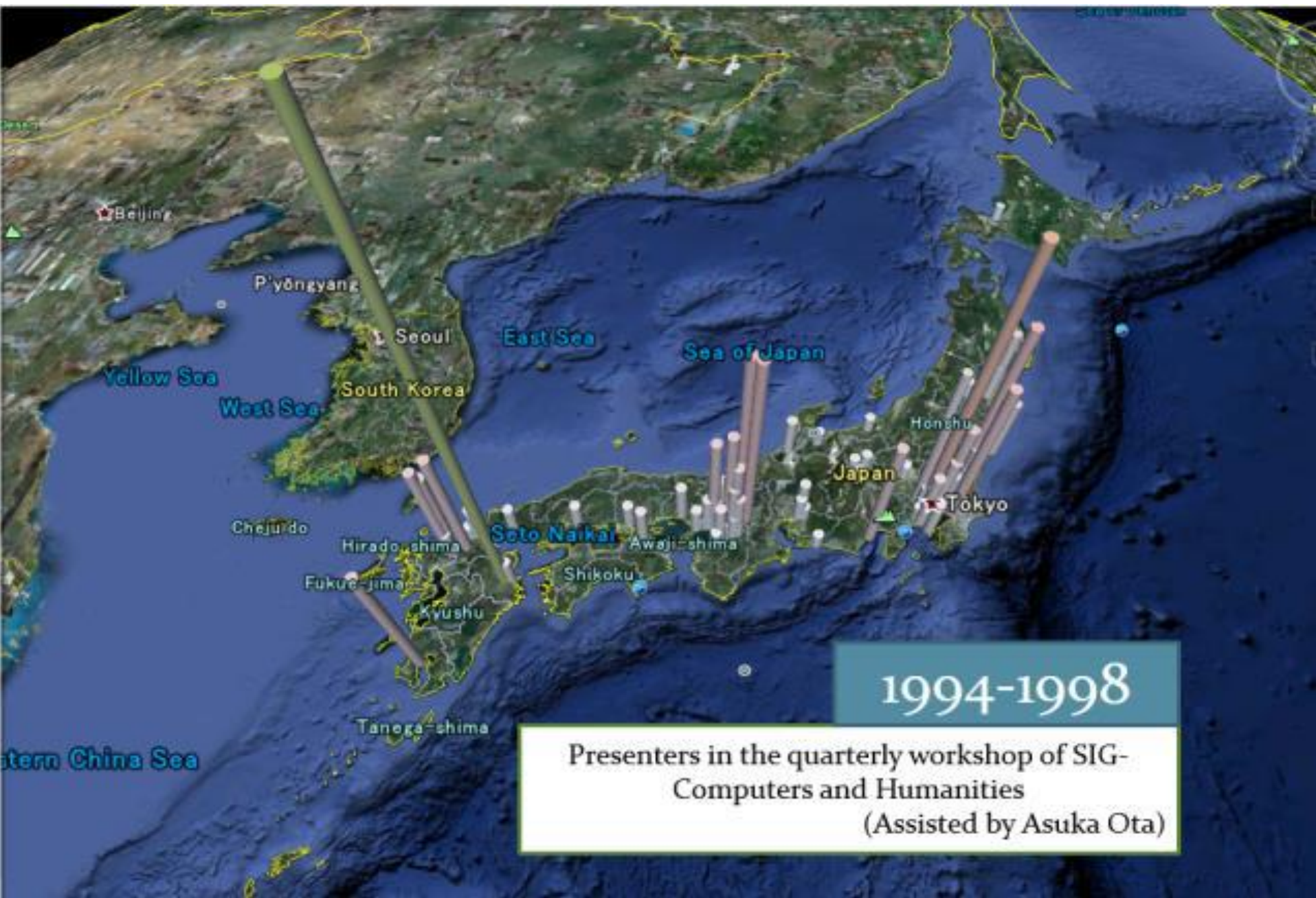- JSPS-funded projects of DH-like studies including making digital cultural resources for the humanities:

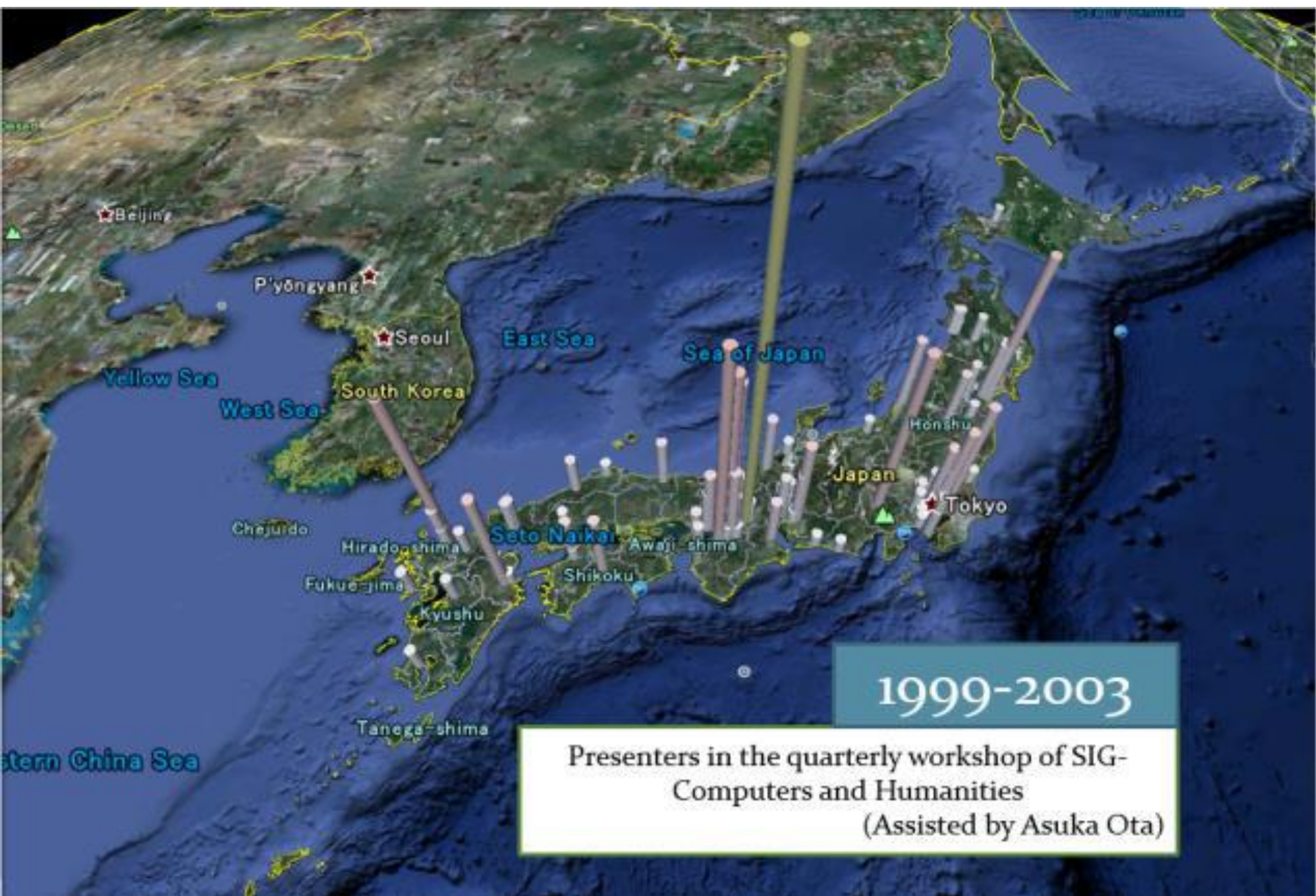| Year	Number of projects	Total (in JPY)
2009	99	251,862,000
2010	126	385,300,000
2011	172	505,390,000
2012	203	654,160,000

SIG-CH presentations

- 1989-present, quarterly workshop
- Over 800 presentations in total
- Registered by around 300 researchers
- Non-refereed papers
- Allowed anyone to make a presentation
- Strongly IT-oriented
  - Under the auspices of a kind of IT society
- Started by IT researchers, but recently humanities scholars increasing
- Note: They don't reflect entire situation of this field.
  - There are many other communities.
- It is suitable for an observation from a stable viewpoint.

1989-1993

Presenters in the quarterly workshop of SIG-
Computers and Humanities
(Assisted by Asuka Ota)

1994-1998

Presenters in the quarterly workshop of SIG-
Computers and Humanities
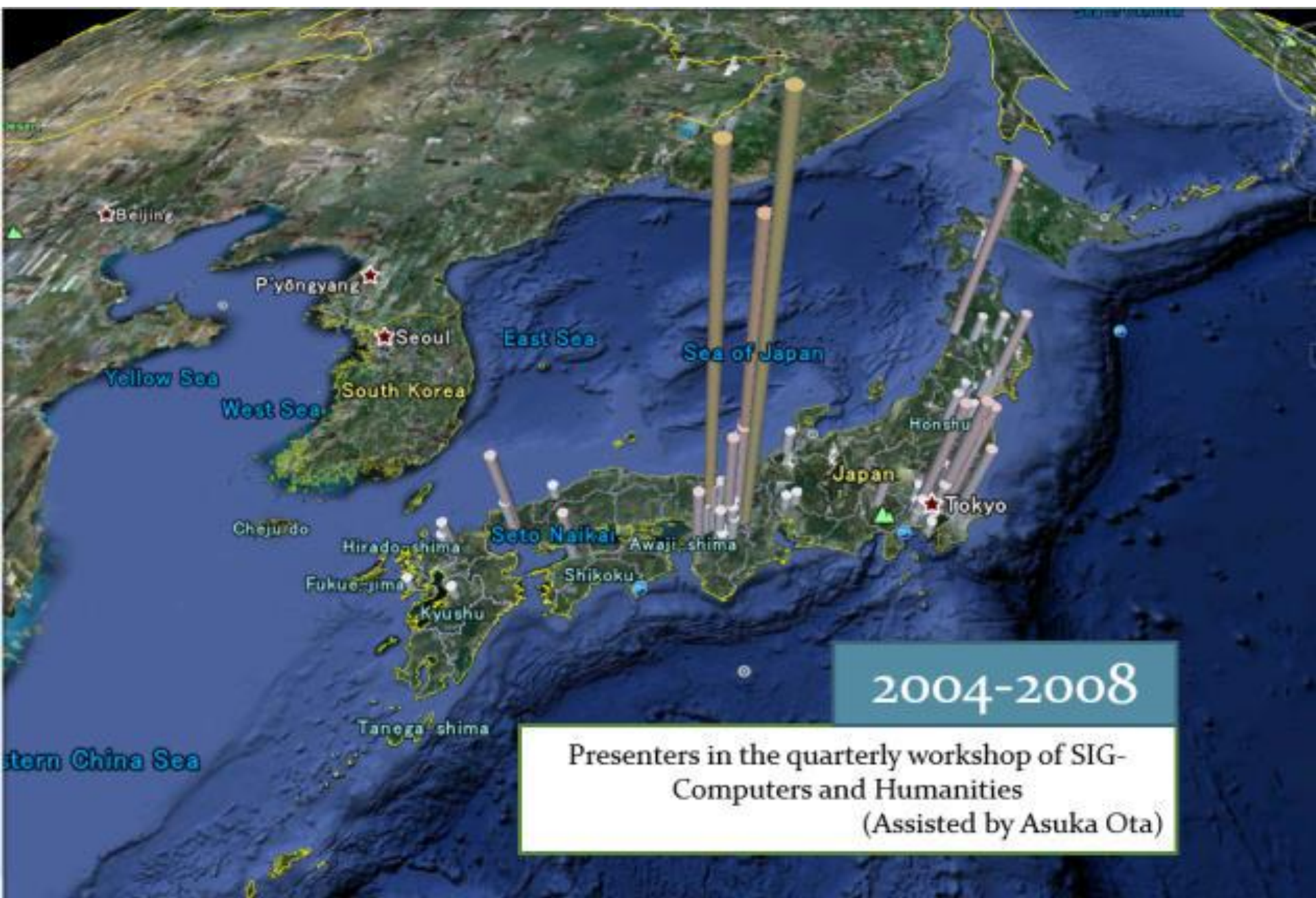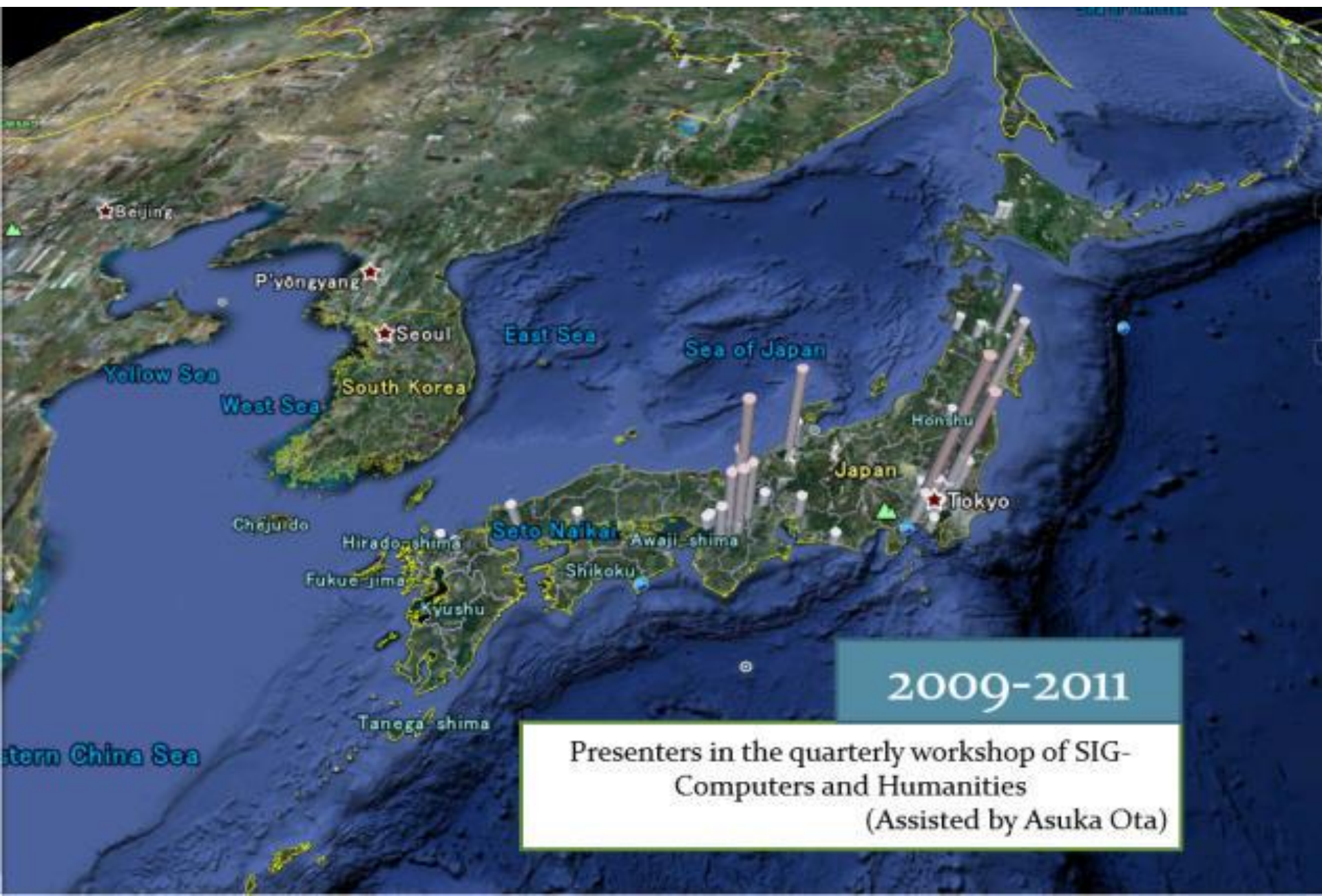(Assisted by Asuka Ota)
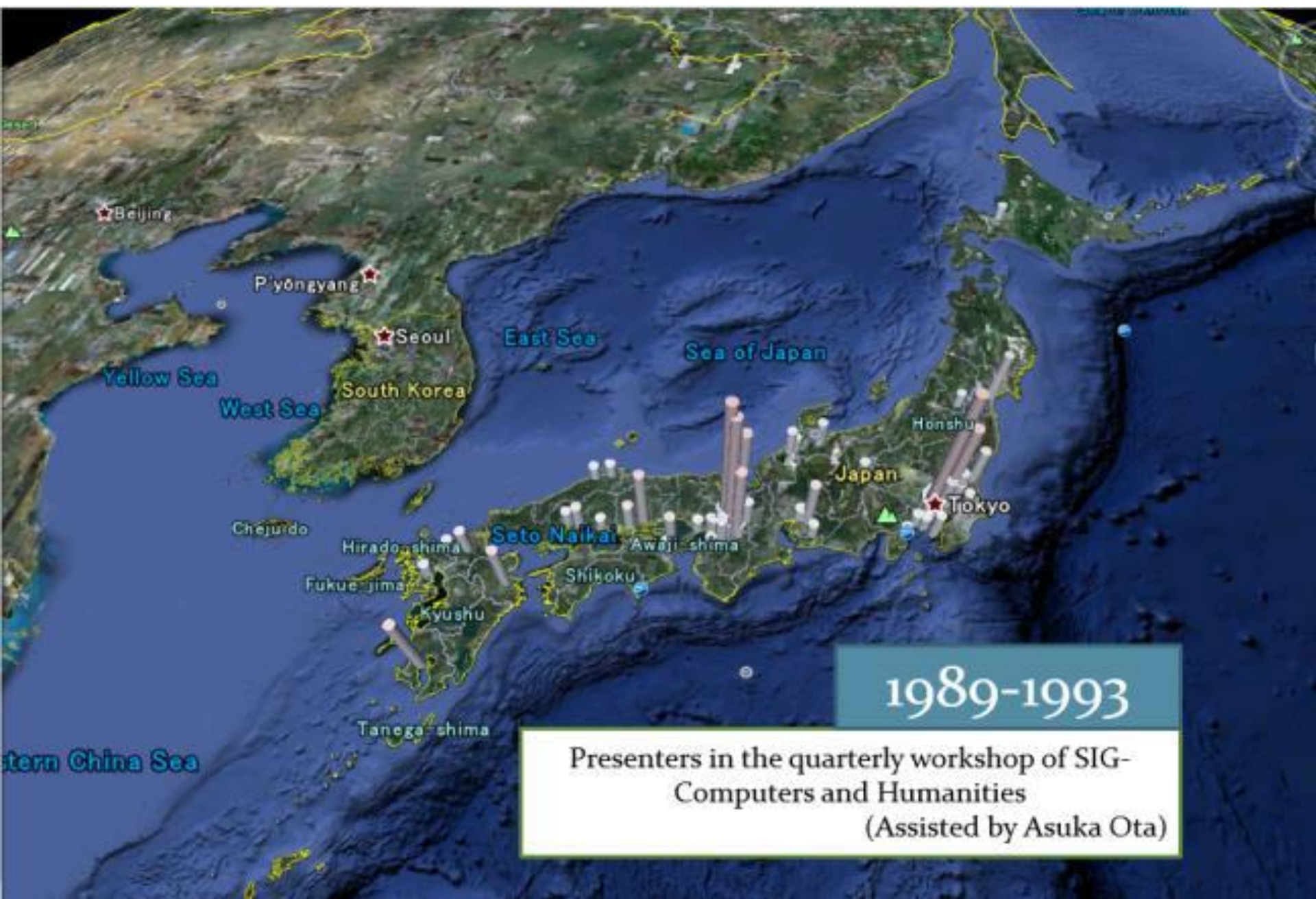
1999-2003

Presenters in the quarterly workshop of SIG-
Computers and Humanities
(Assisted by Asuka Ota)

2004-2008

Presenters in the quarterly workshop of SIG-
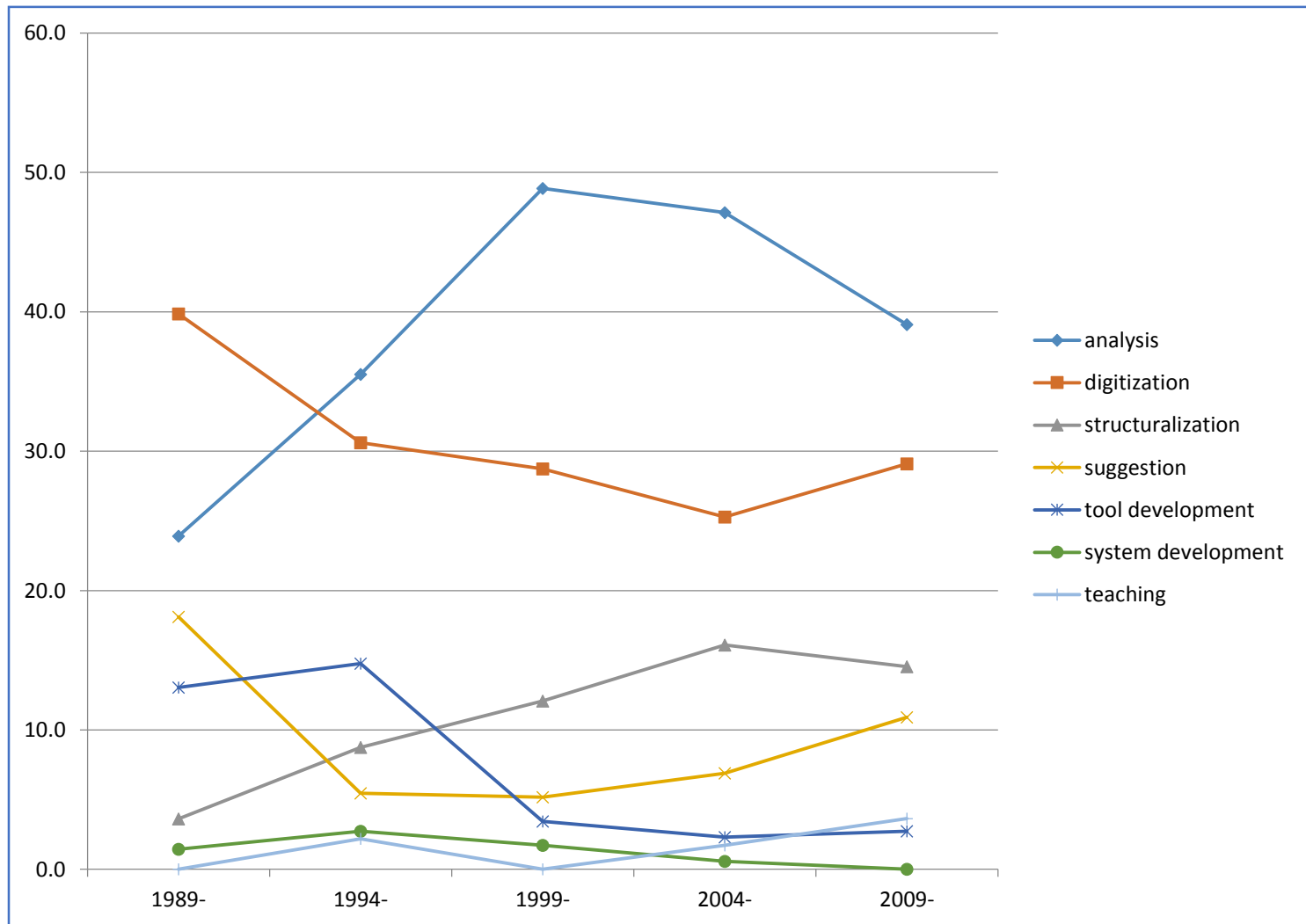Computers and Humanities
(Assisted by Asuka Ota)

2009-2011

Presenters in the quarterly workshop of SIG-
Computers and Humanities
(Assisted by Asuka Ota)

Each target field of the presentations in SIG-CH: Top 10

Each target field of the Presentations
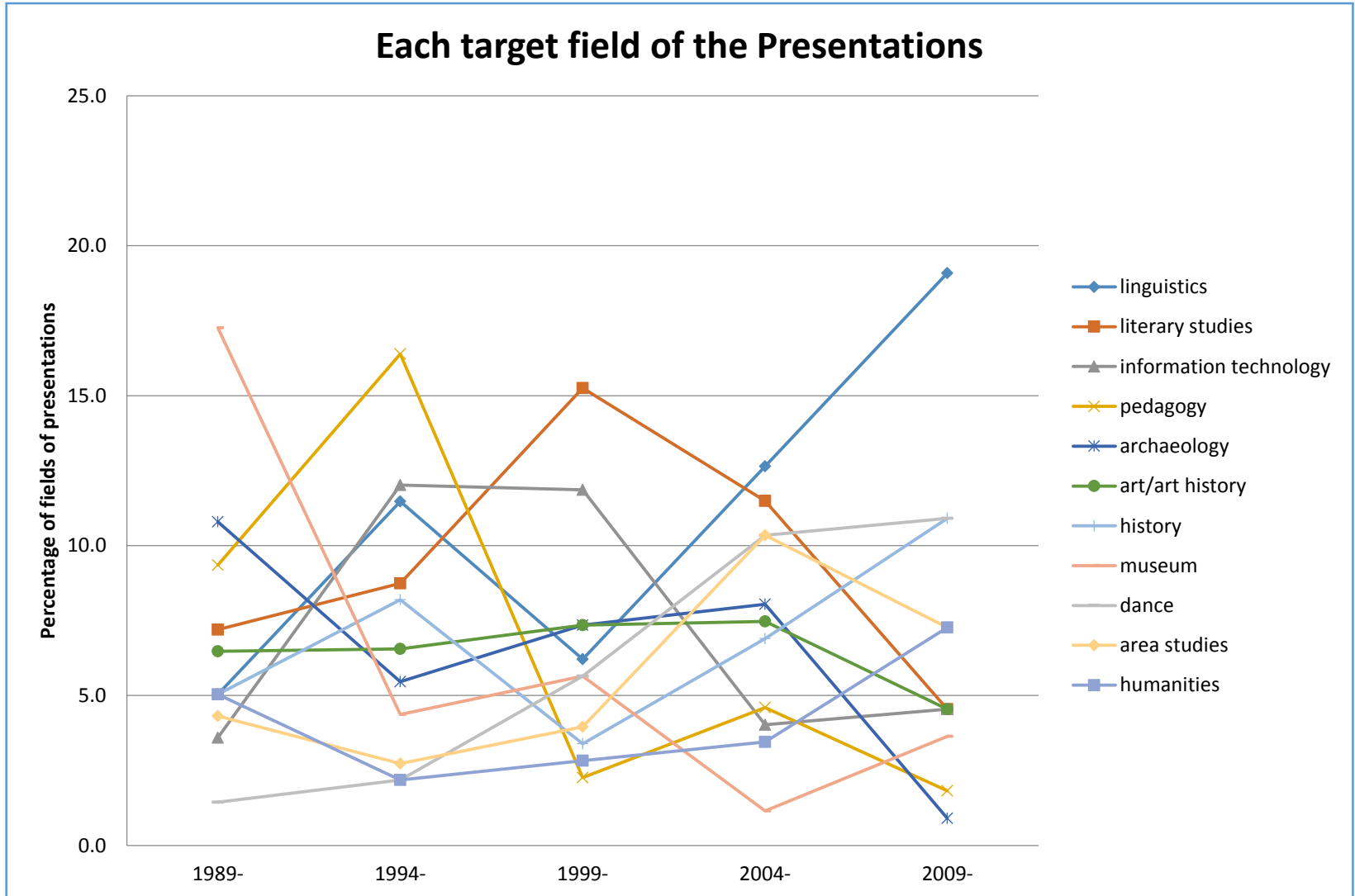
Method adopted in the presentations in SIG-CH

Types of DH-like research

- Archiving
  - Making digital cultural resources
  - Providing digital cultural resources

- Analyzing
  - Analyzing digital cultural resources
  - Analyzing human (or research) activities in digital age

- Representing
  - Making exhibition using digital cultural resources

Our responsibility for you:

- Providing:

  - useful digital cultural resources which deserve to be adopted in academic research.

  - Information of reliable digital methods to analyze Japanese digital cultural resources.

  - …while dealing with them as our research results.

  - But…

Our difficulties in general:

- Getting budget, such as research grants
  - Budget stopped, gradually everything disappears
  - Accountability is seriously required by founders and administrators.
- Surveying global / North American trends and standards for digital cultural resources
  - General style of usage in North American libraries such as JSTOR
  - Related de-facto standards such as Unicode, TEI, Discovery services
  - Often a brand-new domestic format had been invented…
- Providing useful information in English

Our difficulties in general:

- Getting budget, such as research grants
  - Budget stopped, gradually everything disappears
  - Accountability is seriously required by founders and administrators.
- Surveying global / N... ...dards of digital cultural resour...
  - General style ...OR
  - Related de-fa... services
  - Often a brand...
- Providing useful information in English

I would like to ask you: (1)

To build a system of evaluating of digital resources which built by researchers, such as award or so.

⇒ It is very helpful for our accountability

Our difficulties in general:

I would like to ask you: (2)

- Getting budget, such as research grants
  - Budget stopped, gradually everything disappears
  - Accountability is seriously required by founders and administrators.

- Surveying global / North American trends and standards for digital cultural resources
  - General style of usage in North American libraries such as JSTOR
  - Related de-facto s   services
  - Often a brand-ne

- Providing useful in

To help to get information about such things because they need in early stage of project, as it is difficult to change anything after formulating strategy and policy.

Our difficulties in general:

- Getting budget, such as research grants
  - Budget stopped, gradually everything disappears
  - Accountability is seriously required by founders and administrators.
- Surveying global / cultural resources
  - General style of u
  - Related de-facto s services
  - Often a brand-new do format had been invented…
- Providing useful information in English

I would like to ask you: (3)

To help to write useful information in English and to gather it on any convenient place such as Wiki-like site if you can.

Change the situation…?
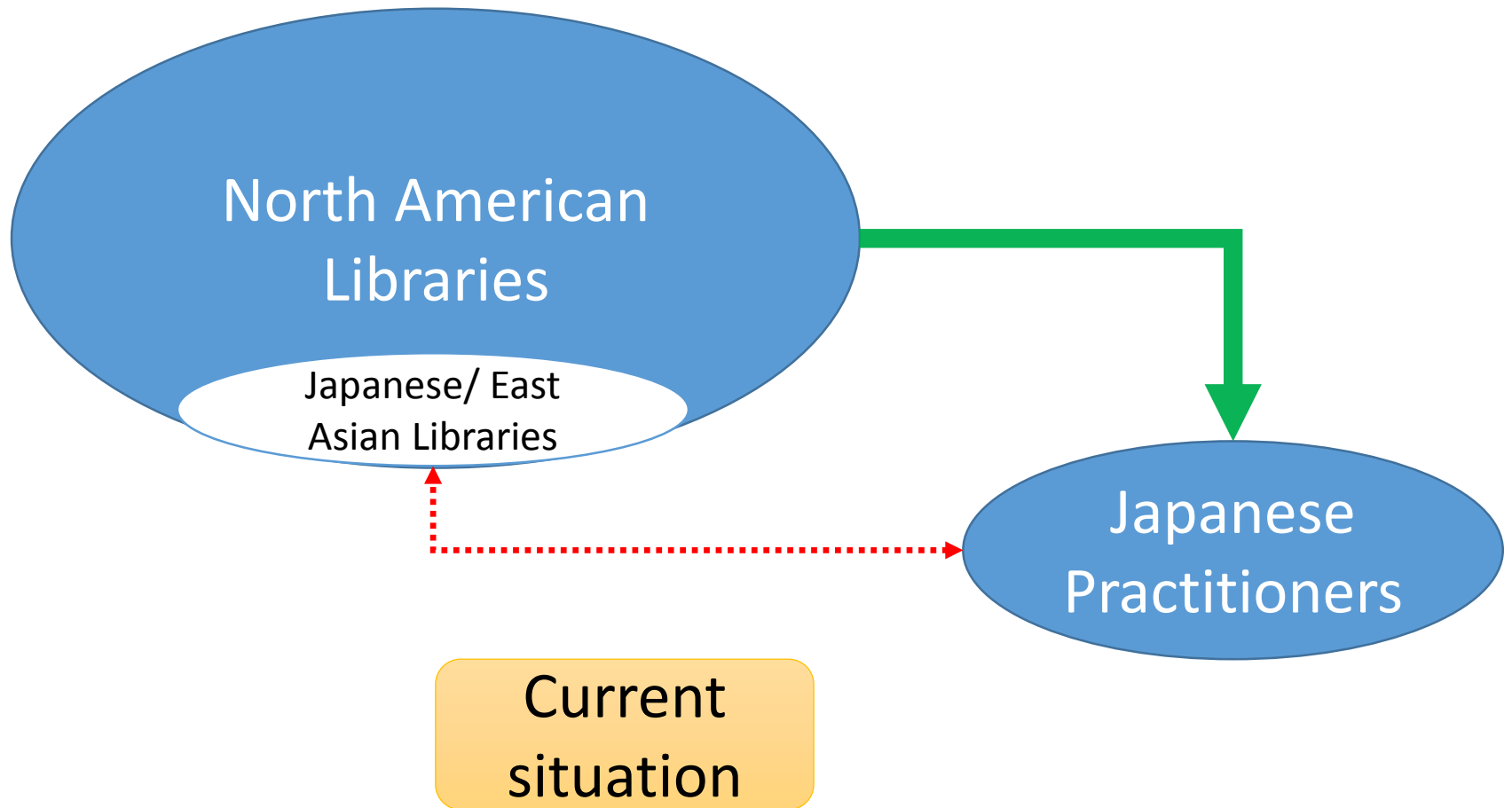
Change the situation…?
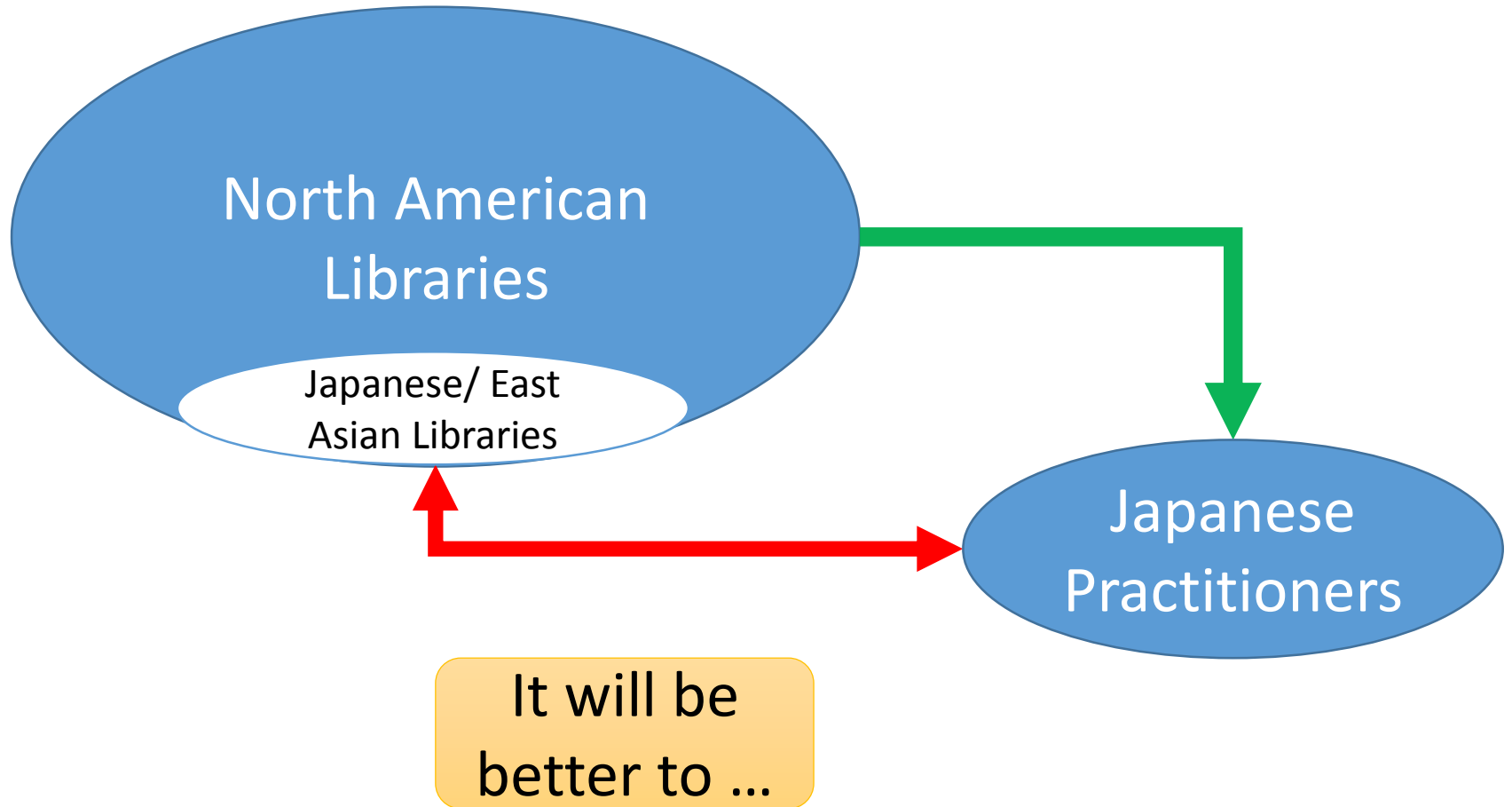
An example of efforts in a project (1)

- The SAT Daizōkyō Text Database Committee
- Since 1994, led by Professor Masahiro Shimoda
- Published a Buddhist text database including over 100,000,000 Han characters on Web in 2008
- Linking other digital resources so that users can use it easily in English
  - Search Chinese/Japanese texts in English term
  - Browse English terms and aligned English sentences with Chinese texts
- Has used Web collaboration systems since 2005
- Has been incubated in both domestic and international DH communities since its early stage.

संगणिकीकृतं तैशोत्रिपिटकं
Saṃgaṇikīkṛtaṃ Taiśotripiṭakam
The SAT Daizōkyō Text Database

大藏經 テキストデータベース

新大藏 大正

ホーム　　　　検索　　　　ご挨拶　　　　組織　　　　利用条件　　　　使い方　　　　English　　　　History　　　▲

大正藏検索

dream | AND ▼

☐ Punctuation　Hangul　Eng　Search　[settings]

長阿含經 (No. 0001 佛陀耶舍譯 竺佛念譯) in Vol. 01

|◄ ◄|| ◄● 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 ●► ||►► ►| [Ln]　[返点:無/有]　CITE

T0001_.01.0001b09: 佛説 11 長阿含經卷第一
T0001_.01.0001b10: 　12 後秦弘始年佛陀耶舍共竺佛念譯
T0001_.01.0001b11: (一)第一分初 13 大本經第一
_.01.0001b12: 如是我聞。一時佛在 14 舍衞國 15 祇樹 16 花林
_.01.0001b13: 窟。與大比丘衆千二百五十人倶。時諸比丘
_.01.0001b14: 於乞食後集花林堂。各共議言。諸賢比丘
_.01.0001b15: 唯無上尊爲最奇特。神通遠達。威力弘大。
_.01.0001b16: 乃知過去無數諸佛。入於涅槃。斷諸結使
_.01.0001b17: 　17 消滅戲論。又知彼佛劫數多少。名號姓字。
_.01.0001b18: 所生種族。其所飮食。壽命脩短。所更苦樂。
_.01.0001b19: 又知彼佛有如是戒。有如是法。有如是慧。
_.01.0001b20: 有如是解。有如是住。云何諸賢。如來爲善
_.01.0001b21: 別法性知如是事。爲諸天來語。乃知此事。
_.01.0001b22: 爾時世尊在閑靜處。天耳清淨。聞諸比丘
_.01.0001b23: 作如是議。即從座起。詣花林堂。就座而
_.01.0001b24: 坐。爾時世尊知而故問。謂。諸比丘。汝等集
_.01.0001b25: 此。何所語議。時諸比丘具以事答。爾時世
_.01.0001b26: 尊告諸比丘。善哉善哉。汝等以平等信。出
T0001_.01.0001b27: 家修道。諸所應行凡有二業。一曰賢聖講法。
T0001_.01.0001b28: 二曰賢聖默然。汝等所論正應如是。如來
T0001_.01.0001b29: 神通。威力弘大。盡知過去無數劫事。以能
T0001_.01.0001c01: 善解法性故知。亦以諸天來語故知。佛時
T0001_.01.0001c02: 頌曰
T0001_.01.0001c03: 　　　比丘集法堂　講説賢聖論
T0001_.01.0001c04: 　　　如來處靜室　天耳盡聞知
T0001_.01.0001c05: 　　　佛日光普照　分別法界義
T0001_.01.0001c06: 　　　亦知過去事　三佛般泥洹
T0001_.01.0001c07: 　　　名號姓種族　受生分亦知
T0001_.01.0001c08: 　　　隨彼之處所　淨眼皆記之
T0001_.01.0001c09: 　　　諸天大威力　容貌甚端嚴

Suggested keywords ✖

Click a keyword.
七夢 五夢 化夢 哑子得夢 夢 夢中 夢中心行 夢境 夢幻 夢幻
泡影 夢心 夢想 夢所見 夢王 夢者 夢見 夢觀 大夢 如夢 如夢
所見 如幻如夢 幻夢 無有顧戀希望染著 猶如夢 異夢 皆如幻
夢 眠夢 睡夢 臥寐 諸法如夢 長夢

☐ 佛説尼拘陀梵志經:0011 (施護譯)
☐ 佛説大集法門經:0012 (施護譯)
☐ 長阿含十報法經:0013 (安世高譯)
☐ 佛説人本欲生經:0014 (安世高譯)
☐ 佛説帝釋所問經:0015 (法賢譯)
☐ 佛説尸迦羅越六方禮經:0016 (安世
☐ 佛説善生子經:0017 (沙門支法度說
☐ 佛説信佛功德經:0018 (法賢譯)
☐ 佛説三摩惹經:0019 (法天譯)
☐ 佛開解梵志阿&T075586;經:0020
☐ 佛説梵網六十二見經:0021 (支謙部
☐ 佛説寂志果經:0022 (竺曇無蘭譯)
☐ 大樓炭經:0023 (法立譯法炬譯)
☐ 起世經:0024 (闍那崛多譯)

Digital Dictionary of Buddhism ✖

電子佛教辭典
パスワードがない場合は「guest」でログインしてください。
Users who do not have a password can log in with the userID "guest".

検索語：入於涅槃

入於涅槃 to enter nirvāṇa (rù yú nièpán)

入 to enter (rù)
於 at (yú)
涅 nir (niè)
涅槃 (Skt. nirvāṇa) (nièpán)
槃 a pan (pán)

BDK-SAT parallel corpus ✖

Search results in BDK-SAT corpus:
(If any Chinese text below is clicked, the entire text is displayed on the center window.)
All texts in this corpus are downloadble at BDK Web site in PDF format.

1. 我今於中夜當入於涅槃
   -> Today I will enter nirvana In ...
2. 佛道轉于法輪入於涅槃而不捨於菩薩之道
   -> Although attaining the enlight...
3. 名聞普周遍輸入於涅槃
   -> Their fame will spread univers...
4. 我所作事畢當入於涅槃
   -> When the thing I have to do is...

An example of efforts in a project (2)

- Proposing Han characters which are not encoded in Unicode to ISO/IEC 10646 (IRG and WG2).

- Proposing variant characters of Siddham script which lacks in current Unicode to ISO/IEC 10646 (WG2).

- Attempting to form a group for eastern text encoding adopting Text Encoding Initiative Guidelines) with related DH researchers.

ISO/IEC JTC1/SC2/WG2/IRG
Ideographic Rapporteur Group

Last Updated: 2013.11.25; 10:30 HK time

Breaking News:

IRG#41 was held in Tokyo, Japan (2013.11.18 – 2013.11.22)

(Documents for IRG#41, actively updated from 08/08/2013)

[Google Search] ● IRG Site ○ The Web

JTC1/SC2/WG2 – ISO/IEC 10646-UCS Website

IRG Principle and Procedures (SummarySubmissionForm,Blank Submission Form, IWDS updat Flow Chart)
(Latest Version is IRGN 1975, confirmed at IRG41)

IRG Meetings:

Meeting	Document Range	Location	Date
42	IRGN1981–	Qingdao, China	2014.05.19–2013.05.2
41	IRGN1950–IRGN1980	Tokyo, Japan	2013.11.18–2013.11.2
40	IRGN1923–IRGN1949	Hong Kong, China	2013.05.20–2013.05.2
39	IRGN1880–IRGN1922	Hanoi, Vietnam	2012.11.12–2012.11.
38	IRGN1810–IRGN1879	Gyeongju, Korea	2012.06.18–2012.06.2
37	IRGN1757–IRGN1844	Mountain View, California, USA	2011.11.07–2011.11.
36	IRGN1710–IRGN1798	Chongqing, China	2011.04.11–2011.04.
35	IRGN1673–IRGN1748	Macao, China	2010.11.08–2010.11.
34	IRGN1617–IRGN1710	Nagaoka, Japan	2010.06.21–2010.06.2
33	IRGN1580–IRGN1664	Da Nang, Vietnam	2009.11.23–2009.11.2
32	IRGN1483–IRGN1606	Hong Kong, China	2009.06.15–2009.06.
31	IRGN1450–IRGN1567	Kunming, China	2008.11.10–2008.11.

CJK_F v1.0 Seq. No.	Rad Stroke Fs		G Hz G Source	J Hz J Source	K Hz K Source	Sat Hz Sat Source
鞡 01401	24	1				鞡 Z_SAT02644
顰 01402	24	2			顰 KA-KC06116	
黼 01403	25	2				黼 Z_SAT02643
嚘 01404	25	2			嚘 KA-KC04905	
嚼 01405	25	4				嚼 Z_SAT04043
嚼 01406	28	5				嚼 Z_SAT04566
矙						矙

Welcome to the official home of

JTC1/SC2/WG2 - ISO/IEC 10646 - UCS

Latest update: 2014-03-18

Latest Contributions can be accessed via the WG document register or anonymous ftp
ftp://std.dkuug.dk/ftp.anonymous/JTC1/SC2/WG2/docs

|		
N4571	4th edition delta fdis charts	2014-02-27
N4570	Subdivision of work for PDAM2	2014-03-18
N4569	PDAM2	2014-02-26
N4568	DAM1 Content	2014-02-27
N4567	Background for power symbols document N45	2014-02-26
N4566	Unicode Liaison Report to WG2	2014-02-26
N4565	Approved Disposition of Comments Amendme	2014-02-26
N4564	Approved Disposition of Comments DIS 4the e	2014-02-26
N4563	Annex S Source Separation example	2014-02-25
N4562	Feedback on Nushu N4566	2014-02-25
N4561	Nushu ad hoc report	2014-02-25
N4560	Siddham ad hoc report	2014-02-25
N4559	Not Used	
N4558	Request of Changing One CJK_C Source and G	2014-02-25
N4557	Brief summary Siddham	2014-02-24
N4556	Reply to N4513 and N4533 Comments on Nus	2014-02-22
N4555		
N4554	Resolutions - meeting 62	2014-02-28
N4553	Draft Minutes - meeting 62	Pending
N4552		
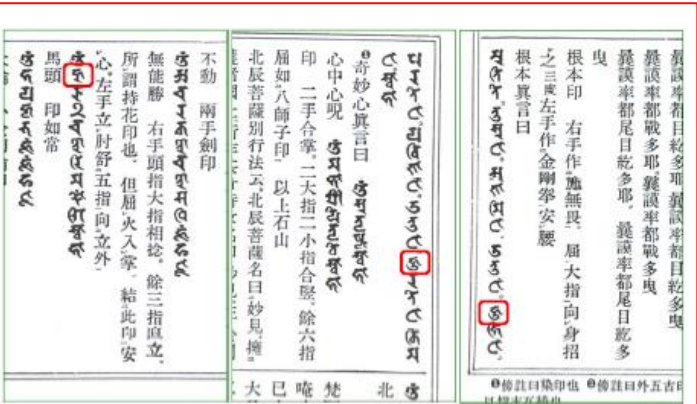
Figure 1: Differentiation of Vowel Sign U in the Mantra for Bhaisajyaguru (薬師如来),
Usuzoushi, Taisho Vol. 78, Text No. 2495, p. 655 (left) and p.680-681 (mid-right).
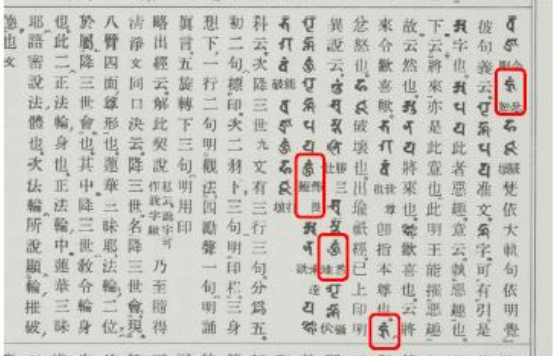
Figure 2: Another Semantic Differentiation of Vowel Sign U in Single Document
Byakuhoushou (by Chouen, 13C), Taisho Vol. 96 (Zuzoubu Vol. 10), p. 313

2.2 Co-Usage of Vowel Sign UU Variants

Figure 3 and Figure 4 are submitted in L2/14-055. For the detailed analysis of their
semantics are described in L2/14-055.

Additional information (1):

- SIG-Transcribe JP in JADH started "翻デジ2014" with support from Japanese National Diet Library to transcribe its digital images by use of Omeka so that users can search Japanese texts.

Brief summary of "翻デジ2014"

- Transcribing the image files in Japanese or Chinese texts in digital collection of the library.
- "翻デジ2014" will play a role of incubating and supporting transcribing projects:
  - Providing a function of collaborative transcription to the public.
  - Transcribed text data should be public domain in general.
  - Collaborators should transcribe it manually so far.
  - Planning to combine OCR with manual correction and semi-automated correction which will be provided by "Crowd4U" project (http://crowd4u.org/).

TranscribeJP

プラグイン 表示 ユーザー 設定　ようこそ, Super User ログアウ

ダッシュボード

アイテム

コレクション

アイテムタイプ

タグ

Scripto

If you would be interested in it, please contact me!

まあ、コンな風で、大蔵経は、高麗本が一番完全なも
のです、彼の黄檗版の蔵経は、明版を翻刻したのです
から、ヤハリ高麗蔵に比べては完全とは云はれませぬ、
そこで、私は忍澂上人の事を聞いて初めて蔵経開版の
ことを発願したので、之が「縮刷蔵経」開版の第一の動機
です、

それから、第二の理由が、支那の歴史で見ると、何か
高僧の名著でもあつて、皇帝の御意に召すと、「勅して
大蔵に入る」とあつて、直に蔵経の中へ入れて了ふ、
さうして皇帝の序文とか、宰相の序文とか云ふものが
蔵経の首に付いてあつて、中々貴いものであるから、
何卒日本でも、佛教の貴い典籍は斯様工合に尊重して、
セメテは天皇陛下から序文を戴き、将軍家に跋でも書
いて貰つて、一つ日本の大蔵経を作りたいと思つたの
か、即ち私の蔵経宿初を発願した第二の原因で、これ
はまだ、私が二十歳代の時なのです、

電子翻刻テクストをWikiに保存

wiki formatting help

w discussion

子翻刻窓を隠す] [wiki] [履歴]

Watch page　Unprotect page　Import page　新字旧字混在　旧仮名遣い　Wikiのみ

Additional information (2):

- As you may know:

- National Institute of Japanese Literature has gradually been starting a large digitization project of Japanese cultural materials.
  - But the institute gave up to transcribe them.
    - A typical Japanese weakness…
  - The project aims to make digital facsimiles.
  - Please attend to the project with your warm eyes.
    - And give me any advices if you would find something.

Additional information (3)

- JADH conference will be held at University of Tsukuba, Ibaraki, Japan, 19-21, September.
  - "Bridging GLAM and Humanities through Digital Humanities"
  - Deadline of CFP is 7 May
- Please submit your research or practice!

Additional information (4)

- If you are able to read Japanese and interested in Japanese DH situation, please see "Digital Humanities Monthly" http://www.dhii.jp/DHM/
    - Essays of researchers and practitioner of DH
    - DH-related event report
    - And so on…