Design and development of an international clinical data exchange system: the international layer function of the Dolphin Project

Jing-song Li, Tian-shu Zhou, Jian Chu, Kenji Araki, Hiroyuki Yoshihara

ABSTRACT
Objective At present, most clinical data are exchanged between organizations within a regional system. However, people traveling abroad may need to visit a hospital, which would make international exchange of clinical data very useful.

Background Since 2007, a collaborative effort to achieve clinical data sharing has been carried out at Zhejiang University in China and Kyoto University and Miyazaki University in Japan; each is running a regional clinical information center.

Methods An international layer system named Global Dolphin was constructed with several key services, sharing patients’ health information between countries using a medical markup language (MML). The system was piloted with 39 test patients.

Results The three regions above have records for 966,000 unique patients, which are available through Global Dolphin. Data exchanged successfully from Japan to China for the 39 study patients include 1001 MML files and 152 images. The MML files contained 197 free text-type paragraphs that needed human translation.

Discussion The pilot test in Global Dolphin demonstrates that patient information can be shared across countries through international health data exchange. To achieve cross-border sharing of clinical data, some key issues had to be addressed: establishment of a super directory service across countries; data transformation; and unique language translation. Privacy protection was also taken into account. The system is now ready for live use.

Conclusion The project demonstrates a means of achieving worldwide accessibility of medical data, by which the integrity and continuity of patients’ health information can be maintained.

INTRODUCTION
Clinical data exchange provides the ability to move clinical information electronically across organizations, while maintaining the meaning of the information being exchanged.1 2 Through the mutual provision of clinical data from disparate medical information systems, not only can a health enterprise offer more timely medical treatment, reduce costs, and make maximum use of medical resources, but it can also maintain the consistency and accessibility of patients’ health information, thereby ensuring continuity of treatment, reducing medical errors, and improving the quality, safety, and efficiency of healthcare services.3 5

However, several barriers, such as communication, standardization, and interoperability, remain to implementing clinical data exchange across organizations. Technically, the lack of healthcare information technology standards and inter-system communication are the major problems that all countries face.5 Although there are industry standards, such as HL7 CDA and CEN 13606, it is still necessary to localize the standards to suit the needs of a specific country or area.9–13 Both the adoption and further development of international standards are essential for clinical data exchange. Beyond that, there are additional non-technical issues, such as financial support, competition, current property preservation, and work flow regulation.14–15 For the above reasons, most clinical data exchange at present occurs between organizations within a regional system, with some occurrences at a national level.16–32

Because people travel around the world far more than ever before, the demand for clinical data exchange across national borders is becoming much greater. According to statistics, 348 million people crossed the borders of China in 2009, including 43.75 million foreigners.33 Considering the great population of foreign travelers, the need to visit a hospital abroad is quite common and makes international clinical data exchange not only necessary, but also useful in helping doctors access patients’ health information in order to provide precise assessments and appropriate treatment plans. The healthcare information generated abroad can either be preserved at a local clinical data center or sent to another electronic health record (EHR) system where the patient is referred. In this way, the integrity and continuity of patients’ health information can be maintained.34–37

This report presents the collaborative work of Zhejiang University in China and Kyoto University and Miyazaki University in Japan concerning the design and development of an international clinical data exchange system using medical markup language (MML). This system is known as Global Dolphin, an international layer function of the Dolphin Project.38 We also report on a pilot study of this application.

BACKGROUND
Medical markup language
In 1995, the Japan Association for Medical Informatics, the ‘Electronic Health Record Research Group’, published an electronically exchangeable medical data standard named MML, which uses the idea of exchanging data with attributes.39–43 Since 1998, MML has been successfully adopted by the Dolphin Project as a clinical data exchange standard.38 In addition, a localized Chinese version
of MML has been created, which makes the exchange of medical data between Chinese health institutions possible.44 45

Because of the similarity of healthcare work flows and medical records between China and Japan, we adopted MML as the clinical data exchange standard in this project.

**Dolphin Project**

The Dolphin Project was proposed in 1998 as a cooperative regional clinical system. The intent was to establish data centers for storing medical information in regional units, creating EHR accounts for each patient, and sharing medical information based on MML, HL7, or other languages. In 2001, the Dolphin Project was adopted by, and developed cooperatively in, the prefectures of Miyazaki and Kumamoto.46 47 Later, both Tokyo and Kyoto deployed the system, with the aim of providing practical EHR services.48 49

The Dolphin Project has three stages of development50: (1) the regional-level system named iDolphin which corresponds to a regional EHR; (2) the national-level system named Super Dolphin which corresponds to a National Health Information Network51; and (3) the international-level system named Global Dolphin as an international clinical data exchange.

**The founding of the inter-organizational project**

Zhejiang University, Kyoto University, and Miyazaki University have a long-term collaborative relationship in many research areas, and the field of medical informatics is an important one of these joint efforts. The cities of Hangzhou, Kyoto, and Miyazaki where the universities are located are famous historic and tourist cities. In 2009, 295 700 Japanese tourists visited Hangzhou, with an average stay of 3.08 days; Japanese tourist occupancy in Hangzhou is 9% of all Japanese tourists in China.

In October 2006, the three universities signed a contract regarding the cooperative research of international clinical data exchange; by the end of 2007, as a basis for international clinical information exchange, the iDolphin-based Xizi Regional Clinical Information Center (XRCIC), which was equivalent to those in Kyoto and Miyazaki, was deployed and in operation at Zhejiang University, Hangzhou, China. The early phase of the Dolphin Project was supported by Japanese government funding in 2001, and the international layer function of the Dolphin Project was mainly supported by Chinese government funding.

**General objectives**

Unlike domestic health information exchange, international clinical data exchange has to adapt to different healthcare work flows, convert the format of medical records, establish a super directory service across countries, and translate documents from the source language to the destination language. Beyond these major issues, there are additional concerns, such as data security, privacy protection, image interoperability, and the physician interface, to be resolved. Figure 1 shows a general map of the whole Dolphin Project, including Global Dolphin, which cooperates with iDolphin and Super Dolphin. Using these systems, patients and physicians can exchange and share clinical information between different countries.

**Figure 1** A general map of the entire Dolphin Project, including Global Dolphin. MML, Medical markup language.

---


---
METHODS
System design overview
The clinical data exchange workflow is shown in figure 2. Before a person travels abroad, he/she submits an application for his/her information to be linked to Global Dolphin and is given a unique global ID so his/her medical records can be queried from the international directory service at locations around the world. In addition, the translation of the record text could be performed beforehand. When the person goes to a hospital in another country, physicians with authorization can query his/her information from Global Dolphin using the global ID through a local site—for example, the XRCIC. The person’s medical records will then be acquired from all of the national-level systems linked to Global Dolphin. If the person has not submitted an application for a global ID, Global Dolphin can still use his/her national ID or regional ID to query directly from the national-level system where the person lives and find the information required. In such a case, however, the free text-type paragraphs in the medical records cannot be immediately displayed in the destination language because the translation work is not performed in real time. The detailed clarification is in section ‘E. Translation’.

Localized Xizi regional clinical information center
In October 2007, we installed and implemented an iDolphin system at Zhejiang University in China to test clinical data exchange between China and Japan; by the fall of 2008, all of the portal sites and applications had been translated into a Chinese version. In September of 2010, this Chinese version of iDolphin was connected to a localized electronic medical record (EMR) system and then transformed into a localized regional clinical information center; the web portal is entitled Xizi Net (see online supplementary figure 1).

International directory service
Among the three levels of the Dolphin Project, the patients’ medical records actually existed at the regional-level clinical information center. In this way, when a patient visits doctors at multiple institutions in different regions, his/her information will be collected and preserved at different data centers. A directory service is needed to ensure the consistency and continuity of the patient’s medical records and make his/her clinical data an entire collection to offer to a patient-centered service. A lightweight directory access protocol is used to build a single sign-on system; this links together the same person’s respective accounts with his/her medical records under each account in different regional data centers. In this way, a user who logs into any of the data centers can access all of the clinical information in an integrated form.

In Global Dolphin, an international directory service is set up using the OpenLDAP software at the XRCIC. In the directory service, each entity, such as a patient or an institution, is considered an object; information about a particular resource, such as a medical record, image, or facility, is stored as an attribute of that object. Information within objects is secured by the access control list, so that only users with the appropriate permission are able to access it. In Global Dolphin, when the patient travels to another country, his/her original information organized in different objects administered by each regional center will be mapped by the international directory service, enabling queries from the local data center where he/she is presently located.

The international directory service is a very important infrastructure for achieving integration of health records around the world, and making clinical data exchange possible among different countries.

Data transformation
There are differences in the MML standard between the Japanese and Chinese versions; and, even in Japan, the clinical data formats in individual iDolphin are dissimilar from each other. For example, Haniwa Net in Miyazaki uses MML 2.3 and Maiko Net in Kyoto uses MML 3.0. Moreover, there are other heterogeneous systems, including other EHRs and EMRs. To achieve data exchange between disparate systems, Global Dolphin has to implement a data transformation step in the clinical data exchange process to map the medical record structurally.

In this project, the InterSystems Ensemble platform was used to design and develop the process of data transformation. Ensemble has a graphical mapping tool, which allows the user to ‘draw’ lines from fields in one set of data to fields in another. Once the MML file is uploaded to the system, it is captured and delivered in a universal, abstract format as a message object to the data transformation tool. The message consists of a header and a body object. According to the message routing rules and transformation rules, the adapter automatically chooses a mapping style to transform the message body object into a new one in order to fit the target format; through the format-free zone developed on Ensemble, heterogeneous systems can communicate with each other in a consistently compact, efficient, and streamlined manner.

We developed two sets of transformation rules in this project: one is for mapping different MML versions and the other is for mapping MML and HL7. Figure 3 shows an instance of data mapping from MML to our Chinese-localized EMR system (which has an HL7 interface) using the Ensemble tool.

Translation
Clinical information exchange between different countries must overcome a language barrier; this is the biggest difference from domestic health information exchange, and an issue that cannot be solved by medical industry standards or data transformation only. In the Global Dolphin system, the MML content is divided into three types, which are discussed below: the structured record, the free text-type phrase, and the free text-type paragraph.

1. For the structured record, we analyzed the terms and writing patterns of health information documents. Most of the
words recorded are medical terminology, such as disease names, and medical symbols that conform to standards, such as SNOMED, ICD10, and LOINC. For this type of record, a terminology-mapping table was established so that a data transformation method could be used to change the words from the source language to the destination language. Many medical terms in the documents are followed by an attribute or subordination enclosed in parentheses. For example, TP in ‘syphilis (TP)’ represents the TP test for syphilis. These attributes or subordinations are organized into a subclass column of each term for rapid mapping. In this way, the structured record can be automatically translated, and a new language can easily be added. Figure 4 shows a segment of the terminology-mapping table.

2. For the free text-type phrase, as it is written in natural language, the utilization of terminology mapping and replacement methods is impossible. However, the phrases are short and do not contain complicated grammar or medical terms, which means that they can be translated fairly accurately by translation software. In our project, the Google Language API was adopted to automatically translate the short phrases. While the MML file is parsed, the elements of a free text-type phrase, such as ‘<Department> Second Department of Internal Medicine (Outpatient) </Department>’, are marked, and the content is transferred as a parameter to the Google Language API to acquire the destination language.

3. The free text-type paragraph cannot be translated by a machine and requires human translation. The system extracts free text from the MML file and publishes it anonymously through a web interface for the translators of a third party to perform the manual translation work (see Figure 3).

Figure 3 An example of data transformation using the Ensemble tool.

Figure 4 A segment of the terminology mapping table.
online supplementary figure 2). Once the translation is submitted, it is automatically appended to the MML file, and users can see the complete information in their own language. Because the translation work is not performed in real time, the patient must submit an application to prepare his/her clinical information before he/she goes to another country. The third-party translators we use in our work have medical domain knowledge, and are not ad hoc translators, such as family or bilingual staff.55 56

Patient privacy protection is very important when the free text is sent out for translation. However, China and Japan do not have rules such as the Health Insurance Portability and Accountability Act yet. In this project, we have referenced some of the Health Insurance Portability and Accountability Act rules about protected health information (PHI) and a pattern matching deidentification means of protecting personal privacy.57

Language translation may produce misunderstandings and errors, so the translated clinical information will be followed by an icon, which indicates the translation method—that is, ‘!’ for human translation and ‘?’ for machine translation. To reduce the effect of translation errors of diagnoses, the original text is appended to every translated portion, so the users can check the health information themselves. In addition, to display Chinese and Japanese text on-screen at the same time, we use Unicode to encode and decode the MML file; other codes would be transformed to Unicode before exchange.

Security
Information security is one of the most important aspects in international clinical data exchange.58 59 We ensure data security during access, transmission, and storage by the means described below.

A security gateway is set up between the internet and local area networks to separate the applications and database from the external network, preventing outside visitors from directly accessing the internal server. The gateway server is equipped with reverse proxy and a secure socket layer (SSL) virtual private network (VPN), and users need a digital certificate issued by the center to acquire access rights to the applications. Before a user executes a function, the access control list will check whether the user is permitted to perform that operation. In this way, the system can make sure that the right person performs the correct operation at the correct time.

When health information is transmitted via the internet, there is the possibility that it could be stolen and decrypted. Therefore documents transmitted via the internet are encrypted with a digital signature on the SSL VPN gateway, so that only the target user can decode the file stream and obtain a meaningful document.

The safety and authenticity of the data stored in the center is very important; a data backup and authenticity protection application was developed and installed to accomplish this goal based on our previously published work.60—62

Image data interoperability
Only the minimum necessary data will be stored in the regional clinical information center—for instance, a CT image will be uploaded to the data center only if it is attached to a CT report file, and additional images will remain in the PACS. We developed a service-oriented architecture-based interoperable image data application to enable access to the PACS in hospitals through a gateway server.63 When users need more images beyond the storage of the regional center, they can access images from a PACS through a simple object-access protocol-based web service, and the regional center plays a role in certification authentication and access control, in addition to forwarding the request to the location of the image access service.

User interface for physicians
The main users who need to access the clinical information of a patient from a foreign country are medical professionals. A user interface for physicians is provided in the center for them to query, select, view, and write the medical records (see online supplementary figure 3A–D). Accordingly, physicians can make new clinical documents, using free text or structured types, and can view patients’ medical images from remote hospitals.

RESULTS
Currently, there is only one local hospital (660 beds, approximately 1000 outpatients daily) connected to the XRCIC, and, up until the end of October 2010, the XRCIC had collected a total of 245 000 Master Patient Index accounts from the hospital dating back to December 26, 2000, which made information on these patients available to the Global Dolphin system. Other detailed information for the last decade is listed in table 1.

In Japan, the scale of each regional project in Miyazaki and Kyoto is shown in table 2. Registered medical institutions in each region include one large general hospital and other small practices, most of which are private-physician practices. Almost all of the uploaded medical records were collected from the large general hospital in each region, Miyazaki University Hospital (612 beds, approximately 700 outpatients daily) and Kyoto University Hospital (1182 beds, approximately 3600 outpatients daily). The center has only the minimum necessary clinical data in its own database, including MML documents (text) and binary files attached to the MML documents (images).

We have exchanged the clinical data of 39 test patients from Japan to China for study, nine of whom had applied for a global ID. A total of 1155 documents were exchanged, including 1001 MML files and 152 images. Of the 1001 MML files, 197 were free text-type paragraphs that needed to be translated, and 103 of these paragraphs contained a total of 153 instances of PHI; the deidentification tool masked 90.2% of the PHIs in all of the texts; nine dates, three locations, and three names were missed.

At present, the state of our work is between a pilot study and a live application. Although there are no actual patient overlaps,

<table>
<thead>
<tr>
<th>Table 1 Document numbers of the Xizi Regional Clinical Information Center up until the end of October 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document type</td>
</tr>
<tr>
<td>No of patient information records</td>
</tr>
<tr>
<td>(unique number of patients)</td>
</tr>
<tr>
<td>No of diagnosis records</td>
</tr>
<tr>
<td>No of examination reports</td>
</tr>
<tr>
<td>No of test reports</td>
</tr>
<tr>
<td>No of surgery records</td>
</tr>
<tr>
<td>No of clinical summaries</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2 Scale of each local project in Japan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miyazaki</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Registered medical institutions</td>
</tr>
<tr>
<td>Uploaded medical records (unique number of patients)</td>
</tr>
<tr>
<td>No of documents sent (text)</td>
</tr>
<tr>
<td>No of documents sent (images)</td>
</tr>
<tr>
<td>Year started</td>
</tr>
</tbody>
</table>
the Global Dolphin is currently linked with the three regional information centers, and is on standby whenever someone needs it.

**DISCUSSION AND CONCLUSION**

Health information exchange has been identified as an essential strategy for addressing the crisis of cost, quality, and safety in healthcare all over the world. However, there are four common barriers that hinder the implementation of clinical data exchange, which are widely agreed upon around the globe: communication, standardization, funding, and interoperability. These barriers include both technical and non-technical aspects. As a result, most clinical data exchange occurs between organizations within a regional level, with some exchanges occurring at a national level. In Global Dolphin, we have designed and developed an international clinical data exchange system with several key applications and infrastructure that are technically ready for the sharing of health information among different countries.

**Flexibility and expandability**

The three-level configuration of the Dolphin Project is loosely coupled with other systems on the same level and between each of the levels, which makes it easy to incorporate a new system or level to update the existing system, as well as to communicate with other heterogeneous systems.

**Modularity and reusability**

All of the applications in Global Dolphin are modularized and encapsulated as services that can be easily reused and transformed into service-oriented architecture, so that system interoperability can also be achieved.

**Consistency and continuity**

The Global Dolphin system makes the exchange of health information available at the international level, keeps the health information of patients consistent, and improves medical services and the continuity of healthcare.

However, to assure that the Global Dolphin system works in practice, we have to consider more than merely technical aspects. For instance, the different standards for protecting privacy between countries could cause problems, e.g., a patient’s information in China may not be as well protected as in Japan; Global Dolphin system will always try to conform to the stricter privacy protection standards. In addition, some Chinese medical expressions and clinical sections either do not exist in Japan, or differ from those in Japan, including race, traditional Chinese clinics, traditional Chinese diagnosis category, etc. Furthermore, the Chinese health insurance system is also markedly different from that of Japan. Thus we have to adjust to these differences in work flow and modify the health information information module in order to make the exchanged data useful and understandable for each area.

Because of the international approach, we also have to consider health insurance coverage when we develop contracts. Normally, local health insurance is unavailable in most hospitals across borders. However, in our project, the local hospital we have chosen is approved by many multinational health insurance companies.

Although Global Dolphin has many advantages, some limitations are inevitable. The translation of free text-type paragraphs cannot be completed automatically; in our study cases, it usually took 1 day to finish the translation, which would become a severe issue in cases where the document is required immediately. In addition, an inaccurate translation could cause risk, especially when it concerns hypersensitivity. Furthermore, when the free text is translated by a third party, there is the possibility that the patient’s identity could be inferred from the PHI missed or information other than the PHI. A further study on natural-language processing and a semantic-based deidentification method will be carried out and will hopefully resolve the above issues.

Global Dolphin provides an efficient way of dealing with syntax and structural interoperability in the exchange of clinical information, but system interoperability consists of more than data structure and sequencing information; the upper level of syntax is semantic interoperability. In the future, more semantic interoperability technologies, such as ontology language, will be studied and incorporated into the system. For instance, we anticipate using semantic interoperability technologies to match a patient who does not have any IDs in our system.

According to the 59 study cases tested, most of the clinical data exchange in our work was largely within the same technical system (Dolphin Project) and used an MML standard. However, we have developed an HL7 interface based on Ensemble and can communicate with HL7-compatible systems (ie, the hospital connected to the XRCiC). In the future, we will link disparate systems to evaluated data exchange across institutions.

Global Dolphin is a trial of international clinical data exchange, and, as our work continues, we hope it will finally achieve the goal of facilitating access to, and retrieval of, clinical data to provide safer, more timely, efficient, effective, and equitable patient-centered care. Moreover, with the global accessibility of clinical data, many value-added services can be offered by the system in the future, such as international epidemic control, public health assessment, translational medicine research, and medical tourism.

**Funding**

This work is partly supported by the Fundamental Research Funds for the Central Universities and by National Science and Technology Infrastructure Program (Grant No 2011BAH15B08) of China. The early phase of the Dolphin Project was supported by the Research and Development Fund of the Ministry of Economy, Trade and Industry of Japan in 2001.

**Competing interests** None.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**REFERENCES**


Design and development of an international clinical data exchange system: the international layer function of the Dolphin Project

Jing-song Li, Tian-shu Zhou, Jian Chu, et al.

JAMIA 2011 18: 683-689 originally published online May 12, 2011
doi: 10.1136/amiajnl-2011-000111

Updated information and services can be found at:
http://jamia.bmj.com/content/18/5/683.full.html

These include:

Data Supplement
"Web Only Data"
http://jamia.bmj.com/content/suppl/2011/08/09/amiajnl-2011-000111.DC1.html

References
This article cites 46 articles, 7 of which can be accessed free at:
http://jamia.bmj.com/content/18/5/683.full.html#ref-list-1

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in
the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
Editor's choice (1219 articles)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/