THE STUDY ON GEOSPATIAL INTEROPERABILITY BASED ON ONTOLOGY INTEGRATION

Chengfang Hu[*1], Liping Di[2], Wenli Yang[2], Xiao Xiao[1]
[1] Changjiang River Scientific Research Institute,
23 Huangpu Ave.,Wuhan,Hubei 430010, China. E-mail: hucf_2@163.com

[2]CSISS, George Mason Univ.,
4400 University Drive, MSN 6E1, Fairfax, VA,22030, U.S.
E-mail: ldi@gmu.edu, wyang1@gmu.edu

KEY WORDS: Semantic interoperability, Ontology, Integration, GIS, Metadata

ABSTRACT: In the past several years, the interoperable sharing of GIS information across organization and communities has become important due to the growing demand of GIS information research and application. Although it is need to use data from different geospatial communities, but the interoperability gaps between distinct community metadata information models have made data access a challenge. Ontology integration is an effective way to establish the mapping relation of the heterogeneous metadata information models. It is a key technique to resolve such gap; also, it is a hot topic in interoperability field. This paper will focus on the effort of concept and method analysis on ontology integration. To solve the difficulty in the deficiency of the semantic express in XML schema the XSLT technique to translate the semantic information model were applied. By adopt the ontology mapping the information sharing and interoperation between different metadata information models could be converted to mapping between one application ontology and another. This presentation discusses several key challenges in the ontology integration process. The first is structure interoperability, and the second one is semantics interoperability. Structure interoperability faces the problem of structure heterogeneity of metadata information models. Semantics interoperability describes the detail strategy of mapping the metadata syntax by different expression ways. After the method analysis, this paper gives a prototype system. The example is the transformation between geospatial metadata specification ISO 19115 and THREDDS Dataset Inventory Catalog Specification. The result shows that with the helping of ontology integration, these two sets of metadata models can be access each other almost lossless.

1. INTRODUCTION

In recent years, remote-sensing technology has made a great development in Earth observing field to improve the efficiency and effectiveness of retrieval of Earth monitoring resource. However, in the face of such huge resource, different functional department use different manner to organize it. GIS researcher always need choose the appropriate application tool to process the data which from the given functional department. The difficult bring the geospatial system interoperability gap. This unprecedented capability poses many considerable challenges to data and information systems for supporting the Earth system related research and applications that typically require integrating multi-disciplinary, multi-mission data for analyzing and decision-making. One of biggest challenges is how to integrate more two of the traditional mission-based information systems, and make them have the ability to provide seamless, cross-system, and cross-functional department data accessing. In brief, it is to implement the interoperability of two geospatial system.

Haslhofer give the definition of interoperability that it is the ability to exchange metadata between two or more systems without or with minimal loss of information and without any special effort on either system[1], the paper will focus on the study effort of two geospatial searching catalogue system interoperability, that the proposed solution can be considered as solution in the case of tow information model mapping.

Ontologies provide a fundamental classification for geographic domains that captures, for example, categories of entities recognized for a domain, as well as the relations that link these categories. In general, ontologies play an important role for knowledge representation, database design, information retrieval, and the semantic web, where they are used as an information engineering tool, for taxonomic reasoning and for first order logical inference. With respect to GIScience, ontologies have been promoted particularly for their role in reuse, sharing, and interoperability.

Eective ontology integration techniques are often needed both during ontology development, and when ontologies are used in conjunction with data. Especially for GIS searching catalogue system, the ontology of the system give the guide information of the metadata. It clearly shows the syntax and relation of the element. The process of ontology integration in fact can be turn into the process of two metadata system integration.

As from the Wikipedia, The term metadata is an ambiguous term which is used for two fundamentally different concepts. Although the expression "data about data" is often used, it does not apply to both in the same way. Structural metadata, the design and specification of data structures, cannot be about data, because at design time the application contains no data. In this case the correct description would be "data about the containers of data". Descriptive metadata, on the other hand, is about individual instances of application data, the data content.

Metadata is traditionally found in the card catalogs of libraries. As information has become increasingly digital, metadata is also used to describe digital data using metadata standards specific to a particular discipline. By describing the contents and context of data files, the quality of the original data/files is greatly increased. For example, a webpage may include metadata specifying what language it's written in, what tools were used to create it, and where to go for more on the subject, allowing browsers to automatically improve the experience of users.

This paper will focus on the effort of geospatial catalog service interoperability crossing OGC catalogue[2] and geoscience community THREDDS catalogue.

2. MATERIALS AND METHODS

2.1. General System Architecture Interoperability Methods

The purpose of GIS catalogue system is to provide the according researchers or the public with access to the digital geographic data. Users may use this type of catalogue site to search, find and download geographic data and the associated metadata that he wanted. In this way, the catalogue system gives a great advance for sharing geospatial data. And furthermore, if we gather the various GIS searching catalogue as one site, and use only one main portal to access, that will help the researchers to find his need data more effectively and rapidly. There are two challenges during resolving this problem, one is searching mechanism interoperability and the other is metadata system interoperability.

This paper choose two popular geospatial searching catalogue to research, one is OGC CSW developed by GMU CSISS, the other is THREDDS developed by UNIDATA.

There are several ways for two GIS system to interoperate. And these two goal catalogues are all B/S structure, they are in accordance with the basic three layers architecture which can be divided into three layers: data layer, server layer and display layer.

The design of B/S architecture system always following such principles[3]:

1. This system adopts the design concept of standardization and modularization.
2. Considering the real situation of various kinds storage method，it must design universal storage repository suitable for metadata content. It may be as the form of database or as the ordinary files system.
3. Humanized interface design principle makes general users use without training. It sticks to the design philosophy that as long as you can type， the user can conveniently apply this system.

The function of every layer is as followed:

1. Display layer: process interaction and communication with users. Despite display layer is not more important than other layers, but it almost gets all glory, because it is the only layer that the user can see. This layer is in charge of the interaction between the system and the user. Web client resides in the user's computer, and generally is applied for receiving the form of Web browser（form）.
2. Server application server: Web server is on the address of Web host, and applied for making dynamic Web pages and organizing the form of the system. The server receives the user client request and analyzes it. And after the process at server side, the information of result will be transform to the client. It processes the information that the user needs. The functions of this layer include the following three parts： a.Access（obtain and save）the data in database layer. b.Obtain data from display layer. c.Execute necessary operation and/or process data. Server application layer obtains data from database layer and process it according to the demand of the display layer. The procession logic layer also can obtain the data, provided by the display layer and process it according to the database layer.
3. Data layer: store all the data that this system processes. Provide data service for procession logic layer or display layer.

From the above characteristics, we give the methods about the B/S architecture system interoperability. It will have three ways to get the solutions. The schematic diagram can be described as the figure 1. In the process of interoperability, we suppose the direction is from system A to system B, and remain the system B original frame and no need to change anything. After the interoperability, the user in system A can operates the system B without alter site entrance.

Method 1 Develop middleware bridge the system A's server to system B's database.

This method need to develop a middleware to connect server A to database B. That means the searching engine A have the capability to inquiry the database B. For A side, it must give the effort to consult with side B, and ask them to open the right of their database, and also need to give the construction illustration of the database B.

Method 2 Develop client A apply to the server B engine API.

This method means the client in side A is a multifunction client, it is complex that can adapt for both server A and server B. Go on this way, the programmer need to know the server B's API very well, and design a complicated one, also need consider reduce the conflict of two server's response. For this method, the server B need have its own

API, which allow the third party to develop the client. Generally speaking, the application follows the service-oriented architecture (SOA) have this characteristic, and the development task of server and client can be separated.

Method 2 Develop transform middleware bridge client A and client B.

The Web client communicate with the Web server usually through the mode of "request——respond". The Web client sends the request to the Web server, and then the Web server responds according to the request. In this method, we translate the request of A and B, then sent the reform request to the server B. the server will sent back of the response which is the format of B's design. The middleware get the response then translate it to the format of response A, and display it to the client A's user interface.

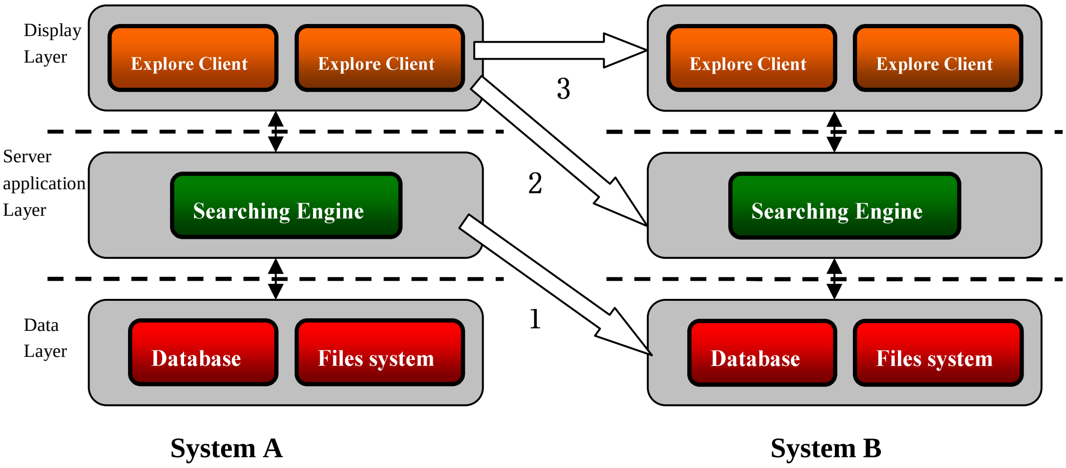

Figure 1 Schematic diagram of searching system interoperability methods

From the analysis above, we draw the following conclusion which representive as table 1:

Table 1 Comparison of the system interoperability methods

	CONTENT	STRONGPOINT	SHORTCOMING	COUPLING	FEASIBILITY
METHOD 1	Middleware connect A server to B database	Straight design plan, the sever A can realize searching function more flexible without restriction of server B's searing engine	Getting the accessing database right is difficult and impossible	High	Low
METHOD 2	Develop client A apply to the server B	Only need to research the API of server B, no need to communicate with other side database	Development task is heavy. Need spend more time to study the API specification.		
METHOD3	Middleware translate between client A and client B	Lightweight middleware, easy pulg in and expanding.	Need to study the request and response format of side B.	Low	High

As the description of the above table, the coupling of the method 3 is the lowest, and also it is the most feasible one, we will choose method 3 to complete the interoperability research.

2.2. Steps of the Interoperability

Figure 2 gives the sketch map of the interoperability step. The steps are as following:

1. The user use client A to sent a searching request. And select using catalogue B on the client interface to execute it.
2. Middleware start after the catalogue B selection state is activation, and translate the A request to B request format.
3. Middleware sent the request to server B site url, and waiting the response.
4. After getting the response from server B, the middleware translate it into the specification A, and then sent back it to client A to display.

The user in client A will have no feel that he is visiting the site B. However, in this way, the user will get more information about multi-site

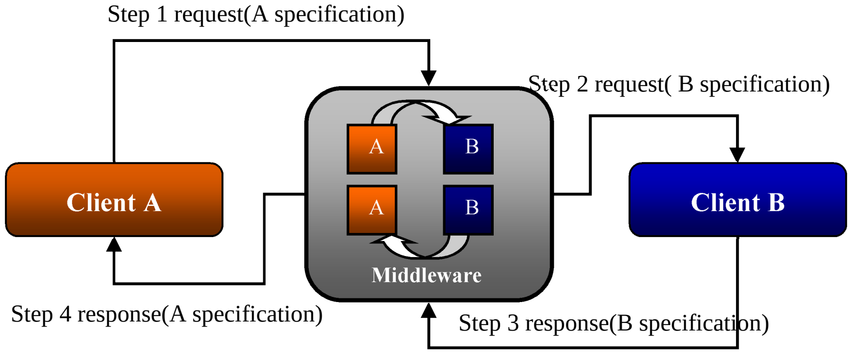

Figure 2 Steps of clients interoperability

The Open Geospatial Consortium (OGC) is international standards organization that is leading the development of standards for geospatial and location based services. In recent years, OGC has developed a set of web-based interoperability protocols for data access and services which including Web Map Services(WMS), Web Feature Services(WFS), Web Coverage Services(WCS) and Catalog Services for Web(CS/W). Catalog Services for Web (CS/W) is a specification that support the registry and discovery of geospatial information.

THREDDS (Thematic Realtime Environmental Distributed Data Services) is a system to simplify the discovery and use of scientific data and to allow students, educators and researchers to publish, contribute, find, and interact with data relating to the Earth system in a convenient, effective, and integrated fashion. A THREDDS data server always has a THREDDS Catalog that describes what datasets are available, called the THREDDS data server catalog. These catalogs provide a simple hierarchical structure for organizing a collection of datasets, a means of accessing each dataset, a human understandable name for each dataset, and a structure on which further descriptive information can be placed.

Table 2 gives the samples of the response from these two searching engine site. The left one is a segment of CSW ISO 19115 response, and the right one is the THREDDS response.

Table 2 segment response of two searching system

A segment of CSW ISO 19115 response	THREDDS response
<?xml version="1.0" encoding="UTF-8"?> <csw:GetRecordsResponse xmlns:csw="http://www.opengis.net/cat/csw/2.0.2" xmlns:gmd=" http://www.isotc211.org/2005/gmd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:gml=" http://www.opengis.net/gml" xmlns:gco="http://www.isotc211.org/2005/gco" xmlns:ows=" http:www.opengis.net/ows" xmlns:ogc="http://www.opengis.net/ogc" xmlns="http://www.isotc211.org/2005/gmd" version="2.0.2" elementFormDefault="qualified" xsi:schemaLocation="http://www.opengis.net/cat/csw/2.0.2 http://schemas.opengis.net/csw/2.0.2/CSW-discovery.xsd http://www.isotc211.org/2005/gmd http://schemas.opengis.net/csw/2.0.2/profiles/apiso/1.0.0/apiso.xsd"> <csw:SearchStatus timestamp="2008-07-09T12:26:03-05:00" status="complete"/> <csw:SearchResults nextRecord="0" numberOfRecordsReturned="1" numberOfRecordsMatched="1"> <MD_Metadata> <fileIdentifier> <gco:CharacterString>2b109350-e5b7-102a-ac4e-67ff417bcb93</gco:CharacterString> </fileIdentifier> <language> <gco:CharacterString>eng</gco:CharacterString> </language> <parentIdentifier> <gco:CharacterString>cbca2a00-e5b6-102a-92ed-7280417bcb93</gco:CharacterString> </parentIdentifier> <hierarchyLevel> <MD_ScopeCode codeListValue="series" codeList=" http://www.isotc211.org/2005/resources/Codelist/ML_gmxCodelists.xml#MD_ScopeCode"/> </hierarchyLevel> <contact> <CI_ResponsibleParty> <organisationName> <gco:CharacterString>DOC/NOAA/NWS/NCEP</gco:CharacterString> </organisationName>	<?xml version="1.0" encoding="UTF-8"?> <catalog xmlns="http://www.unidata.ucar.edu/namespaces/thredds/InvCatalog/v1.0" xmlns:xlink="http://www.w3.org/1999/xlink" name="Unidata THREDDS Data Server" version="1.0.1"> <service name="thisDODS" serviceType="OPENDAP" base="/thredds/dodsC/"/> <dataset name="Realtime data from IDD"> <catalogRef xlink:href="idd/models.xml" xlink:title="NCEP Model Data" name=""/> </dataset> <dataset name="Test Restricted Dataset" ID="testRestrictedDataset" urlPath="restrict/testData.nc" restrictAccess="tiggeData"> <serviceName>thisDODS</serviceName> <dataType>Grid</dataType> </dataset> </catalog>

From the table 2, we can see that all of these responses are the type of XML format. Because of the same representive form, we have the precondition to make the interoperation from each other. We use XSLT technique to resolve it.

XSLT is used to transform an XML document into another XML document, or another type of document that is recognized by a browser, like HTML and XHTML. Normally XSLT does this by transforming each XML element into an (X)HTML element. With XSLT we can add/remove elements and attributes to or from the output file. we can also rearrange and sort elements, perform tests and make decisions about which elements to hide and display, and a lot more.

Base on the XSLT, we need define the transformation model first, then use XSLT to read the input xml file and after procession, we can get the output of result xml file.

The key challenge is how to make the transformation model file. This is the problem of metadata information mapping.

2.3. Information Model Interoperability

In the last few years, the problem of generating mappings between ontologies has been extensively investigated. As a first step in the integration of such independently developed ontologies it is usually necessary to establish appropriate correspondences (or mappings) between the terms used in the various ontologies. Their vocabularies will most likely diverge, either because they use different namespaces, or because they use different names or naming conventions to refer to their entities. As a consequence, these ontologies will most likely be unrelated from a logical point of view, even if they intuitively overlap.

For these systems, the metadata information model is an instance expression of its ontology. Metadata is a machine processable data that is used for the interpretation and the processing of multi-media content for their adaptation, filtering, or semantic knowledge extraction. Metadata describes different types of information: multimedia contents, semantics of these contents.

There are have several interoperability issues during metadata information mapping, include: Syntactic interoperability, Semantic interoperability and structural interoperability.

A. **Syntactic Interoperability:**

If two or more systems are capable of communicating and exchanging data, they are exhibiting syntactic interoperability. Specified data formats, communication protocols and the like are fundamental. XML or SQL standards are among the tools of syntactic interoperability.

Some of the metadata elements in THREDDS and CS/W profile refer to the same meaning, and use the different expression syntax. In this type, element mapping can be done directly from one to another. The schema of ISO 19115 profile is based on ISO/TS19139 XML schemas[4].

For example /catalog/dataset/@name can map to /MD_Metadata/identificationInfo/MD_DataIdentification/citation/CI_Citation/title. The syntax "name" in THREDDS and "title" in ISO profile are pointing to the same object that by which the cited resource is known. In the schema, their type are "xsd: string" and "scXML:CharacterString" which can be transfer smoothly without any problem.

B. **Semantic Interoperability:**

Beyond the ability of two or more computer systems to exchange information, semantic interoperability is the ability to automatically interpret the information exchanged meaningfully and accurately in order to produce useful results as defined by the end users of both systems. To achieve semantic interoperability, both sides must refer to a common information exchange reference model. The content of the information exchange requests are unambiguously defined: what is sent is the same as what is understood.

During our mapping job, not only syntax mapping should be completed, but also structure mapping is need to be done. The following will describe a examples of structure mapping.

In THREDDS catalog, the attribute of an element may play an important role to deliver the information. The different attribute value make the element present the different means. For example:

<date type = "modified"> 2011-08-11 11:20:28z </date>

means the modified date of this dataset is 2011-08-11 11:20:28. The value of /catalog/dataset/date@type is restricted in the enumeration values: "created", "modified", "valid", "issued" and "available". The five value means the "date" element can present the date in different five conditions.

However, in ISO19115, we can not translate this "date" element and its attribute "type" to corresponding structure, ISO19115 doesn't use attribute to restricted the date type, instead it use two coequal elements that one is the date information, the other is the property of the date to represent this information. The result is as follow:

```
<date>
    <gco:Date> 2007-08-18 10:24:28z </gco:Date>
</date>
<dateType>
    < CI_DateTypeCode codeList="../ISO19139/resources/codeList.xml?CI_DateTypeCode"    codeListValue="revision"/>
</dateType>
```

C. **Structural Interoperability:**

In THREDDS catalogue xml file, several dataset elements may be contained at the same time, each one with its corresponding metadata description information. The metadata relation of these datasets may totally be independence or inherits from one to another. If two datasets are independent, the content of the metadata will have no intersections. If the two datasets are parent son relation, the metadata of the son dataset will includes two parts: the metadata inherited from its parent, and the metadata of its own. The first part of metadata will not appear within this dataset element scope, it is only appear in its parent element with the flag of attribute "inherited" value equal to "true". This structure decide that the datasets at the same level in the catalogue tree may contain some similar description metadata if they have the same parent dataset, but they also have some difference because they still remain the special characteristic metadata respectively.

In CS/W, the major elements to describe the metadata are "MD_Metadata" and "DataGranule" respectively corresponding to ISO 19115. As to the CSW profile specification, "MD_Metadata" can not contain another

“MD_Metadata” element, the value of MD_Metadata.hierarchyLevel.MD_ScopeCode@codeListValue is used to describe this “MD_Metadata” is Dataset or Datasetcollection. And “DataGranule” can not include another “DataGranule” either. So the CS/W client can not get the hierarchy structure information directly from the XPath of the xml file. The CS/W profiles have the opposite mechanism to store the parent son relation. It uses a reference element storing the parent ID information to keep this hierarchy structure. In ISO19115 profile, it is MD_Metadata.parentIdentifier.

As the analysis above, we should do some structure transfer from THREDDS to CS/W. The hierarchy pointer direction stored in THREDDS catalogue xml is from parent to child, we will parse this and change to reserve it from child to parent in CS/W. And for the purpose of meeting the CS/W searching method to the best, we will separate the dataset in one THREDDS catalog xml file into server dataset units, every unit will with its full metadata description. Although it will bring some overlap storage space in CS/W database, but when the CS/W retrieve a dataset unit for example “MD_Metadata”, CS/W has no need to do some more search for getting its parent metadata. Figure 3 give the relation of this structure mapping.

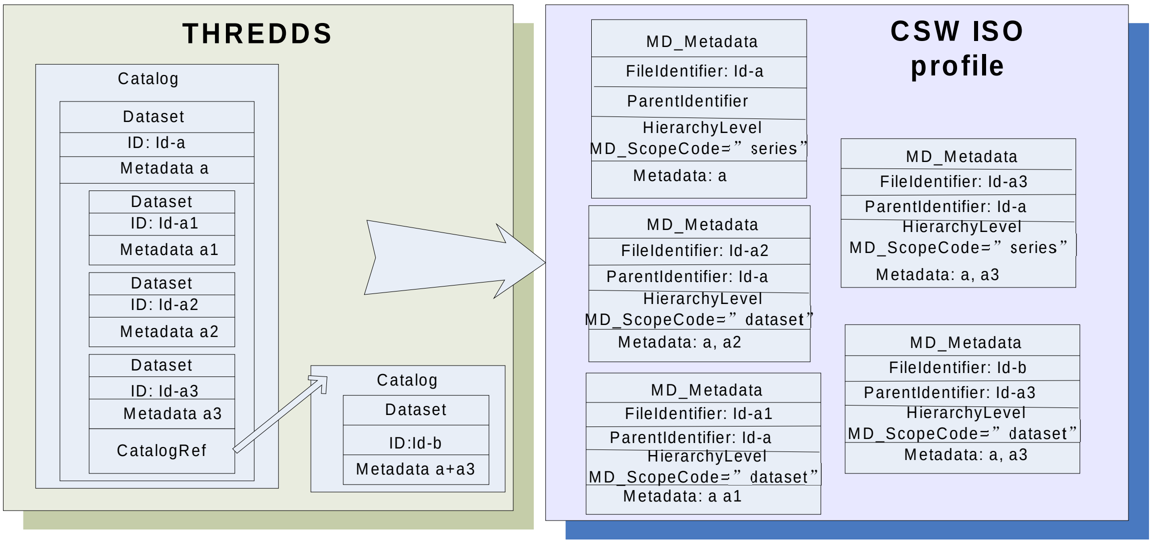

Figure 3 The relation of structure mapping

3. CONCLUSIONS

The interoperable sharing of GIS information across organization and communities has become important due to the growing demand of GIS information research and application. In this paper, we first address the methods of GIS system interoperability, compare their characteristics. Then discuss the approaches and the challenges of metadata information mapping, including Syntactic interoperability, Semantic interoperability and structural interoperability. On the other hand, this paper also describes the detail strategy during the process of element mapping. With the analysis job above, it is possible for us to map the information in THREDDS catalogues into CS/W that implement the interoperability between THREDDS and CS/W.

Acknowledgment. This work is supported by the scientific research foundation for returned scholars (GY2010106) and the Doctor Start-up Program of Changjiang River Scientific Research Institute.

References

1. B. Haslhofer and W. Klas, “A survey of techniques for achieving metadata interoperability,” ACM Comput. Surv.,vol. 42, no. 2, 2010.
2. Voges et al., 2005 Voges, U., Senkler, K. (Eds.), 2005. OpenGIS? Catalogue Services Specification 2.0—ISO19115/ISO19119 Application Profile for CSW 2.0 (Version 0.9.3). OGC Document Number: 04-038r2, 89pp.
3. Aimin Yang, Junping Wu, Lixia Wang, 2009. Based on B/ S of the three-tier architecture of the questions the management system Proceeding CEA'09 Proceedings of the 3rd WSEAS international conference on Computer engineering and applications
4. ISO, 2006 ISO, Text of ISO 19115:2003 Cor. 1, Geographic information –Metadata – Technical Corrigendum 1, ISO/TC 211 N 1979. April 11,2006 41pp