

ORGANIZATION AND USAGE OF METADATA

Evgeny D. VYAZILOV

*Russian Research Institute of Hydrometeorological Information - World Data Center.
6, Korolyov St., Obninsk, Kaluga region, 249020, Russia.*

E-mail: vjaz@meteo.ru

KEYWORDS: metadata, environment, database, Web technology

ABSTRACT

Metadata contain strictly organized, formatting information on resources of environmental data: names of parameters, codes, properties, formats, address of a storage etc. - everything, that is necessary for information technologies to find information resources and correctly them to process. It is presenting stages of appearance (observations, collecting, processing, exchange, dissemination data) and usage (agency, organization, users) of metadata. It is developed the scheme and structure of base metadata, including the more 10 objects of metadata.

It is need to develop improved data inventories; to work out integrated data base, including different data types (factographic, spatial, textual, graphical) and metadata; to standard of metadata and to co-ordinate of studding between programs and nations in this field. It is giving the proposals for organising of metadata bases on several levels (local integrated database, statical and dynamical pages of Web sites). The main tool is proposed to use languages XML, RDF.

INTRODUCTION

For the last decades are created thousands initial, inverted and processed data sets. The information on tens of thousands expeditions, for example, only in Russia more than 35000 expeditions is collected. Hundreds various software, on basis which are calculating thousand various statistic characteristics and parameters. In all this flow of the information it is necessary very quickly to be guided. A paper catalogues, information received from technical carriers etc. do not help. In the first, the information very quickly changes. This information is corrected, reedit, and recalculated. Secondly, volume of such information is also concerning large. So observations coverage databases are estimated in tens Mbytes.

Thirdly, the reference information is a basis for transition to paperless technology of processing data. For thin sample of the data attributes of metadata are required, which sometimes are not present in initial files and about it beforehand it is not known. Metadata is an information to help for data search and processing. For successful search of the high-quality data on basis of a logic contents data and physical characteristics of databases, manage a data in expeditions, programs, centers, countries, international data exchange are necessary various metadata.

TECHNOLOGICAL STAGES APPEARANCE AND USAGE OF METADATA

The metadata are appearing on all stages of passing environmental data:

- **Observation** - information on networks and methods of observation, methods of definition of polluting parameters, ways and places of validation of measuring systems.
- **Observant platforms**- research vessels, hydrometeorological stations, satellites, planes, buys.
- **Data collection** - information on technologies of data gathering, formats of transfer data, description of transmitted complete data sets, the standards of representation and transfer of the data and metadata in the Internet.
- **Accumulation of the data** - description of data sets, organizations - suppliers, owners, users, formats of the collection, storage and exchange of the data, observant projects, Research vessels cruises, data coverage, uniform dictionary of parameters, codifies, the items of information on technologies, methods of the quality check.
- **Interdepartmental and international exchange of the data** - information on technologies and formats of the data, description of complete data sets, observant projects and programs.
- **Data storage** - information on technologies.
- **The applied processing** - information on methods of processing, check and analysis data, software, algorithms of calculation of parameters, data coverage.
- **Modeling** - information on models, methods, data formats.
- **Data distribution** - information on production (analyses, bulletins, yearbooks, climatic atlases, forecasts), forms of distribution (table, diagrams, maps, text, sound, publishing), spatial temporary scales, software, editions (atlases, including electronic), Web addresses.
- **Protection of a nature** - information on techniques, impacts of natural extreme situations, dangerous objects, ways of prevention of this influences.

The basic information necessary for creation metadata contains in various engineering documentations, therefore it require additional ordering and formalization. Metadata have specific features:

- Rather small activity of updating of the help information, both on frequency, and on volume of updating;
- Branch in time, on sources, executors of process of updating from process of use, as a consequence, necessity of maintenance of compatibility of the information subject to editing;
- Necessity of centralization of the metadata.

The structure and accuracy of the description of the characteristics of object metadata depends from scale of system, stage of data processing. The metadata circulates in system of processing data of each organization, hydrometeorological serving of a national economy or which is taking place in system of the collecting, storage, exchange and data processing at the international, national, departmental and corporate levels. On each of these levels the metadata is necessary.

For interaction on international level it is need a information on the international agreements, about foreign and international organizations, data transferred to the international exchange, formats of an exchange of the data, software of their processing, quantity of cruises, data submitted in international data exchange by each country, on areas, for departments, list of observant stations of the international exchange, plans of expeditions for a declared national programs, information on the international projects (programs), measuring systems, list of the experts, information on technologies reformatting from national formats in international and back.

References files in system of the collecting, processing, the storage's and exchange information in the country include an information about data, software, formats of a storage data, information on system of maintenance. Irrationally in each organization to have complete files of information on cruises, flights of planes, hydrometeorological stations, observant platforms, measuring systems methods of polluting substances definition, projects. Such information should be stored in national data centres.

Reference files of common usage are the dictionaries and codifies, which allow facilitating operations on a grouping, sorting, selecting data. The dictionaries provide also uniform understanding of the terms. For example the dictionary of oceanographic parameters consists of tables - information on parameter, information on physical addresses of storage of graphic or textual or HTML files and parameters in the databases placed on distributed computers. A information on separate parameters include a unique code, attributes of parameter (sphere, process, statistical characteristic, type of the data - phactographic, spatial, graphic, textual, name of parameter, name in the output form, accuracy of measuring, units of measurements, range of variability - minimum and maximum meaning, temporary and spatial scale of representation, name of the module for calculated meaning, vertical scale - minimum and maximum meaning, reference to the code tables etc. A information on physical addresses of a storage of parameters include the unique identifier of parameter, time of a beginning and ending of observing, latitude and minimum and maximum meaning), URL or PC name (directory and file name) or IP-address (database name, user name, password database, level of access, name of the table, name of a field in the table), minimum and maximum meaning of parameter, vertical scale - maximum meaning, logic name of registration data (cruise, square), size. The uniform dictionary is using by all developers of subsystems for Web technologies, databases, and programs on applied processing of data.

System of hydrometeorological serving. For a decision making persons, on objects of economy with use of the information about a condition of natural environment are necessary a information about organizations, data sets, which they store, software of processing, formats of the storage and exchange at a level of the country, observant platforms, observant nets, a information on hydrometeorological productions (current information, forecasts, reviews, bulletins, monthly journals, year-books, manuals, atlases,

in which there are of a various data type and climatic generalizations) possible inquiries and tasks.

A service should be focused on the certain classes of the users, on separate technological processes, for example, managing municipal services, planning of a craft flight, search a fish. An information on objects of economy, technological processes subject to influence of environment, results of acceptance of the decisions with use of the data are necessary for hydrometeorological service. For example, a information on the manuals, atlases, navigating maps, ports, places of shelter from storms, climatic characteristics of area of navigation and whether marine lines, forecasts long-term and short-term, current information, list of radio stations transmitting a information, list of establishments serving region by recommended rates, zone of the responsibility for hydrometeorological service are necessary for normal operation of a trade vessel.

At a corporate level the detailed information on each data set, each unit of data collection, exchange as a information on ship cruises, monthly flow of data from hydrometeorological stations and their condition (in processing, on what data carrier), about a condition of data coverage of this or that geographical area on various parameters (a information about observations, buys stations) is necessary; methods of observation, measuring systems, methods of definition of polluting parameters; a information on software, a information on formats of a storage of the data and their structure.

The individualization of oceanographic stations is carried out in time (year, month, day) and on space (latitude, longitude). The absence of any of the listed attributes puts under doubt expediency of storage of such station. Except for the listed characteristics, for complete identification of the data such a information as depth of a place of supervision, method of definition of each parameter, designation of the organizational forms of realization researches (daily, road etc.) are necessary. Besides an information common for all stations, for observation on standardizing sections at sea, etc. it is necessary to mark the additional information characteristics inherent in each form of organization of observation. An additional information about long time series of observation is the attributes a date started and ending of observation, quantity of series, parameters of observations. For standardizing sections such attributes are the name and number of a section, constant coordinates of points of a section.

An information on the data should identify them with a various level of generalization from a base of the receipt, given up to unit, of the data (cruise, data set for month of observation at hydrometeorological stations). But as against an information on sources data (observant platforms) here it are specified reference number, specified period of observation. The allocation in separate databases of the characteristics of units of receipt of the data and descriptions of observant platforms allows considerably speeding up search of an information about the data. Begins possible to have the identical characteristics for the descriptions of information about files. In the descriptions of files information about data of the ships of weather, in standardizing sections should be reflected, etc. Such information are receiving at a level stations practically it is very difficult, and sometimes and it is impossible.

In establishments it is very important to adjust the strict account of receipt of the data, and also results of the **international and interdepartmental exchange of the data**. The results of this account should be reflected in the appropriate metadata bases. So in National

Oceanographic Data Center (NODC) of Russia there is a table, in which on each cruises the availability of these data, relation to Declare National Programs, and also fact of transfer in international data exchange is marked. Thus, the following help information is necessary for functioning NODC:

- A information about data (files, data sets, databases, their contents and structure);
- A information about marine organizations, experts;
- A information on observant networks, projects, programs, methods of supervision etc.;
- A information on platforms and sources of data (research vessels cruises, coastal stations, buys, satellites);
- A information on a complex of software in various organizations of the country;
- The standard qualifiers, dictionaries, codifies (dictionaries of the terms; codes of ships, countries, projects; codes of observable parameters - wave, cloudy, weather etc.);
- A information on algorithms, models, software etc.;
- An information about Web sites. An example of a information about various sites with the information on the publications, public organizations, metadata, international projects and programs, international and national marine organizations is the page Russian NODC " Travel on information resources in field of oceanography", <http://www.meteo.ru/nodc/travel/index.htm>

STRUCTURE OF METADATA

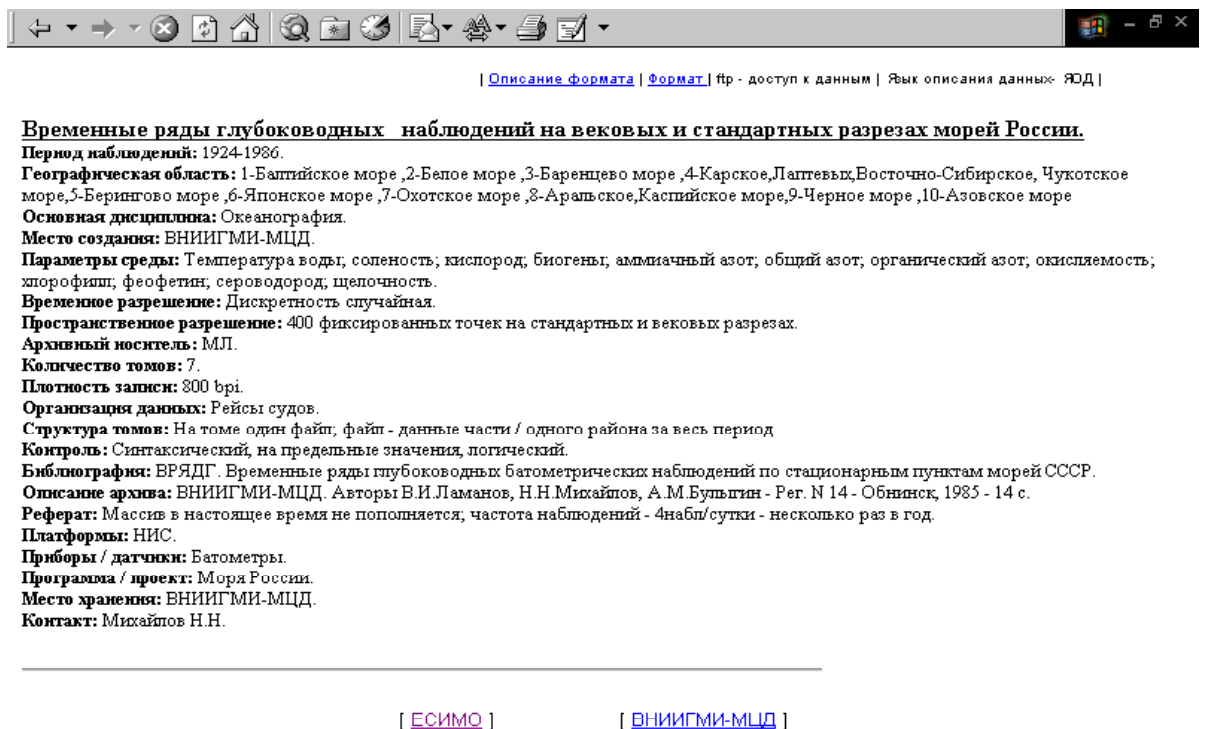
The structure of metadata in the field of oceanography is presented in a **fig. 1**. The metadata should allow realizing search of input data at several levels of their physical organization depending on qualification of the users, their rights, used tools of information searching and etc.

Top level - most accessible via Web technology and intended for the wide users. Here there should be common items of information on the data sets and databases available in various organizations; information about organizations; the observant projects; the experts; information resources on a sea environment available in the Internet and etc. At number of documents for instance more 10 for each type of metadata it is offered to use dynamic resources of organization of access to the information with usage of languages XML (Extensible Markup Language), RDF (Resource Description Framework) [4,5]. Thus there is a possibility of standardization of performance and search of metadata with usage Web technologies. Now in NODC of Russia [1,2] this level is created as html pages [3] for (for search information on data sets, formats, organizations and etc.). For example page with metadata is on **fig.2**.

The second level of metadata is intended for users - specialists, which want to estimate an amount, quality, entirety of this or that databases or data sets. Therefore bases of metadata representing items of information on observant platforms (ships, coastal servers, satellites etc.), methods of observations used instruments here should be used, the detailed data directories of this or that area and etc. This level of metadata forms in the environment of a relational data base management system (DBMS) with possibilities of output of the information in the Internet as dynamic pages with usage of languages XML, RDF. Application of the specially developed description language of the hydrometeorological

data, given for performance, representing the information on an amount of observations on space, time and parameters here is possible. For example, data coverage is presented on **fig.3**.

The third level of metadata is intended for search of the initial information in databases and output given in the form, convenient for the professional user or information system. This metadata in main are intended for the database administrator. The created here bases of metadata will allow under the logical characteristics given to find their physical addresses of storage or procedure of account. It is a level of metadata forms in the environment of a relational DBMS without output of metadata in the Internet. A classification of data will play the large role at this level. The base of this level is the dictionary of parameters permitting on the logical characteristics of the parameter (a name, method of obtaining) and conditions of search (geographical area, period of observations) to receive or physical addresses of data storage (name of the table), or application name for calculation of value of this parameter or its characteristic. For example, table of cruise information is given on **fig.4**. For this level it is created system «Information Resources in Field Oceanography», including information about all metadata objects (observing nets, platforms, sources of data, information on data sets, equipment, projects, etc). Here is using XML. Fragment of this system is on **fig.5**.



| [Описание формата](#) | [Формат](#) | фр - доступ к данным | Язык описания данных: ЯЭД |

Временные ряды глубоководных наблюдений на вековых и стандартных разрезах морей России.
Период наблюдений: 1924-1986.
Географическая область: 1-Балтийское море ,2-Белое море ,3-Баренцево море ,4-Карское,Лаптевых,Восточно-Сибирское, Чукотское море,5-Берингово море ,6-Японское море ,7-Охотское море ,8-Аральское,Каспийское море,9-Черное море ,10-Азовское море
Основная дисциплина: Океанография.
Место создания: ВНИИГМИ-МЦД.
Параметры среды: Температура воды; соленость; кислород; биогены; аммиачный азот; общий азот; органический азот; окисляемость; хлорофилл; феофетин; сероводород; щелочность.
Временное разрешение: Дискретность случайная.
Пространственное разрешение: 400 фиксированных точек на стандартных и вековых разрезах.
Архивный носитель: МЛ.
Количество томов: 7.
Плотность записи: 800 ври.
Организация данных: Рейсы судов.
Структура томов: На томе один файл; файл - данные части / одного района за весь период
Контроль: Синтаксический, на предельные значения, логический.
Библиография: ВРЯДГ. Временные ряды глубоководных батометрических наблюдений по стационарным пунктам морей СССР.
Описание архива: ВНИИГМИ-МЦД. Авторы В.И.Ламанов, Н.Н.Михайлов, А.М.Бульгин - Рег. N 14 - Обнинск, 1985 - 14 с.
Реферат: Массив в настоящее время не пополняется, частота наблюдений - 4набл/сутки - несколько раз в год.
Платформа: НИС.
Приборы / датчики: Батометры.
Программа / проект: Моря России.
Место хранения: ВНИИГМИ-МЦД.
Контакт: Михайлов Н.Н.

[[ЕСИМО](#)] [[ВНИИГМИ-МЦД](#)]

Fig.2 For example of data set description on site

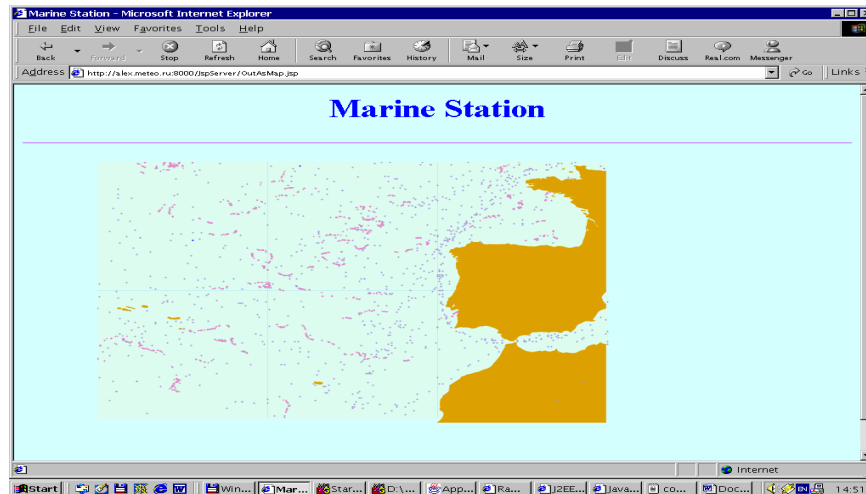


Fig 3 Data coverage (massages BATHY, TESAK, BUY, and SHIP from Global System Telecommunication) for region North – East Atlantic on November 2000

	TAPE	DATASET	CNU	UI	GART_DAT	NISH_DAT	SHIP	COUNTRY	ORG	ATIO	DRIZON	JT_Y	UT
602	HG1A1E	БАЗРЕЙС	602		04-JAN-57	04-MAR-57	UMITAKA-MAI	ЯПОНИЯ	МЦД-Б	38	1022	1969	5 л
603	HG1G1E	БАЗРЕЙС	603		01-FEB-58	08-FEB-58	BROUGH	НОВАЯ ЗЕ.	МЦД-Б	37	335	1969	5 л
604	HG1D1E	БАЗРЕЙС	604	11	19-DEC-56	31-MAR-57	СЛАВА-15	СССР (РОС ВНИРО		54	504	1969	5 л
605	HG1D2E	БАЗРЕЙС	605	12	16-NOV-57	25-MAR-58	СЛАВА-15	СССР (РОС ВНИРО		88	864	1969	5 л
606	HG1G1E	БАЗРЕЙС	606		14-JAN-55	19-FEB-55	GLACIER	США	МЦД-Б	13	207	1969	5 л
607	HG1G1E	БАЗРЕЙС	607		30-OCT-55	30-APR-56	EDISTO	США	МЦД-Б	29	479	1969	5 л
608	HG130E	БАЗРЕЙС	608		02-JAN-56	26-MAR-56	GLACIER	США	МЦД-Б	12	163	1969	5 л
609	HG1G1E	БАЗРЕЙС	609		03-JAN-57	10-FEB-57	ATKA	США	МЦД-Б	2	22	1969	5 л
610	HG1D2E	БАЗРЕЙС	610		09-DEC-56	01-MAR-57	STATEN ISLA	США	МЦД-Б	24	208	1969	5 л
611	HG1G1E	БАЗРЕЙС	611		20-FEB-57	15-MAR-57	NORTHLAND	США	МЦД-Б	6	85	1969	5 л
612	HG1G1E	БАЗРЕЙС	612		04-NOV-56	15-FEB-57	GLACIER	США	МЦД-Б	11	146	1969	5 л
613	HG148E	БАЗРЕЙС	613		03-DEC-57	08-FEB-58	GLACIER	США	МЦД-Б	36	517	1969	5 л
614	HG1G1E	БАЗРЕЙС	614		28-OCT-57	13-FEB-58	GLACIER	США	МЦД-Б	33	603	1969	5 л
615	HG1A1E	БАЗРЕЙС	615		07-JAN-58	25-FEB-58	BURTON ISLA	США	МЦД-Б	11	167	1969	5 л
616	HG1G1E	БАЗРЕЙС	616		04-JAN-58	12-FEB-58	WESTWIND	США	МЦД-Б	46	835	1969	5 л
617	HG1G1E	БАЗРЕЙС	617		04-JAN-60	06-JAN-60	ATKA	США	МЦД-Б	20	394	1969	5 л
618	HG1D3E	БАЗРЕЙС	618		14-JAN-60	13-FEB-60	EASTWIND	США	МЦД-Б	20	346	1969	5 л
619	HG1G1E	БАЗРЕЙС	619		09-DEC-59	13-MAR-60	GLACIER	США	МЦД-Б	57	988	1969	5 л
620	HG1D3E	БАЗРЕЙС	620		08-FEB-60	06-MAR-60	BURTON ISLA	США	МЦД-Б	37	571	1969	5 л
621	HG108E	БАЗРЕЙС	621	3	26-OCT-68	15-MAR-69	АХТАРЕС	СССР (РОС АТЛАНТИ		116	975	1969	Бес
622	HG1A1E	БАЗРЕЙС	622	1	14-AUG-68	06-JAN-69	СУЧАН	СССР (РОС ТИПРО		159	1248	1969	Бес

Fig.4 Table of cruise information

All metadata are stored in the database of a DBMS, where their urgency is supported, and they are used periodically (on a top level) for upgrade of the dynamic information on a site, permanently for execution of searches at the second and third levels. The bases of metadata now form as the tables; the fragment of links of the tables in base of metadata is shown in a fig. 6. In further, after development of resources of storage of metadata in structure, most approximate to languages XML and object model DOM (Document Object Model), which are standard W3C (World Wide Web Consortium). The reorganization of base of metadata is probably as triples (object name, field name and attribute value).

Сортировать по организациям | Сортировать по названиям судов | Список НИС России на 1.01.2000 г.

Основные сведения об экспедиционных судах России (СССР)		
Организация	Название судна	Позывной
ГМСл Азербайджана	11 ПЯТИЛЕТКА	
ЗАПРЫБПРОМРАЗВЕДКА	1500 ЛЕТ КИЕВУ	LYFB
КАСПНИРХ	18 ПАРТСЪЕЗД	
ЗАПРЫБПРОМРАЗВЕДКА	60 ЛЕТ ВЛКСМ	EWYB
ГМСл Латвии	CENTAVRUS	
ГМСл Латвии	HIDROGRAFS	
ДВНИГМИ	K-406	
ГМСл Эстонии	LOOD	
ГМСл Эстонии	NEKMANNGRUND	
ГМСл Эстонии	RIJRITA	
ГМСл Эстонии	REVALSTEIN	
ВНИИМОРГЕО	А.АРХАНГЕЛЬСКИЙ	
СЕВРЫБПРОМРАЗВЕДКА	А.БОРИСОВ	
МОУкрНИГМИ	А.ВАВИЛОВ	
ГосНИНГИ	А.ВИЛЬКИЦКИЙ	
КАМЧАТСК УГМС	А.ЕРМАЛЮК	
ДВНИГМИ	А.И.ВОЕЙКОВ	UAAX
ПИНРО	А.ОТКУПЩИКОВ	UWHH
ТИНРО	АБАКАН	UQHM
ПРИМОРСК УГМС	АБРЕК	

Fig 5 Description of information resources for fields Oceanography

Language XML allows creating own languages on its basis for application in concrete spheres of activity and, as a matter of fact, is the carrier of the data between units of uniform information space. W3C has developed the standard for representation metadata in networks Internet - language of the description of resources RDF, based on language XML. Use of the tables of styles XML allows not only to transform formats, but also gives the mechanism of a manipulation by the data after reception metadata on the screen of the user. For example, may to sort, to make search or to add the information directly from browser. Use of objecting model of the document - DOM allows transforming metadata from a format of a database in XML. Use of language XML allows standardizing structure of various objects metadata, attributes, their names, type of the data. That is on the basis of language XML with assistance IOC the specialized marine language for the description data in field of oceanography is created.

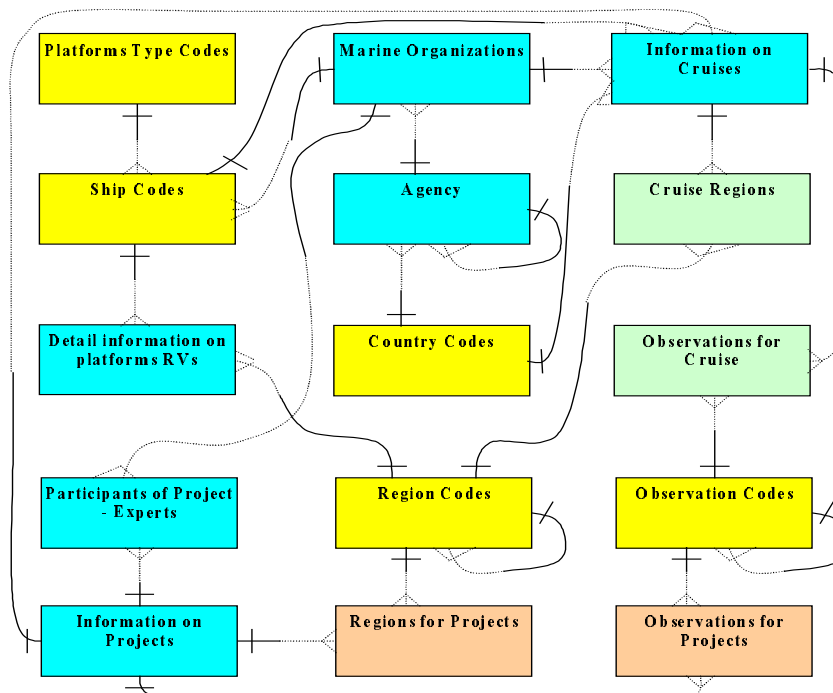


Fig.6 Fragment of scheme metadata base for oceanography

CONCLUSIONS

Metadata contain strictly organized, formatting information on resources: names of parameters, codes, properties, formats, address of a storage etc. - everything, that is necessary for information technologies to find information resources and correctly them to process. The scheme of base metadata, including the description of contents of the tables with metadata bases is developed. The structure of base metadata is worked out.

It is carried out allocation of object metadata that allows more precisely to divide the responsibility for creation of such information and systematically to create them depending on their importance and necessity.

Our research needs are to develop improved data inventories; to work out integrated data base, including different data types (factographic, spatial, textual, graphical) and metadata; to standardising of metadata and co-ordinating of studding between programs and nations.

In spite of that this metadata are creating for field oceanography, more approach may use in field emergency.

References

1. <http://www.meteo.ru> - Russian Research Institute of Hydrometeorological Information - World Data Center
2. <http://www.meteo.ru/nodc> - Oceanographic Data Center of Russia
3. <http://www.oceaninfo.ru> - Web site of Federal subprogram of Russia "System of information for World Ocean"
4. <http://www.w3.org> – World Wide Web Consortium
5. <http://www.w3.org/RDF/> - The Resource Description Framework

Short author biographies: Evgeny D. VYAZILOV, Dr. Sci. for oceanography, Head of Lab. Oceanographic Data Center of Russian Research Institute of Hydrometeorological Information - World Data Center (<http://www.meteo.ru/nodc>). Specialisation: Environmental Data Management; Metadata, Information Systems, Databases, Decision Support System. Main Achievement: Cruises Inventory Data Base, including 32 000 cruises; Computerised Handbook for Informational Oceanographic Resources (metadata base); Web site "Decision Support in Natural Disasters" <http://www.meteo.ru/nodc/project2/action.htm>; thesis "Methodology of Environmental Data Using on Marine Objects for Decision Support"; Web site of Federal subprogram of Russia "System of information for World Ocean", <http://www.oceaninfo.ru>