JRC Scientific and Technical Reports

Temporal Metadata for Discovery

A review of current practice

Makx Dekkers
AMI Consult SARL

Massimo Craglia (Editor)
European Commission DG JRC
Institute for Environment and Sustainability

EUR 23209 EN - 2008

The mission of the Institute for Environment and Sustainability is to provide scientific-technical support to the European Union's Policies for the protection and sustainable development of the European and global environment.

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information

Makx Dekkers

Address: AMI Consult Société à responsabilité limitée
B.P. 647
2016 Luxembourg
E-mail: mdekkers@ami-consult.com

Massimo Craglia (Editor)

Address: European Commission Joint Research Centre
Institute for Environment and Sustainability
Spatial Data Infrastructures Unit
TP262, Via Fermi 2749
I-21027 Ispra (VA)
ITALY
E-mail: massimo.craglia@jrc.it
Tel.: +39-0332-786269
Fax: +39-0332-786325

http://ies.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 42620

EUR 23209 EN
ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

Table of contents

1 BACKGROUND AND RATIONALE

This report contains the findings of a study on temporal metadata for discovery commissioned by the Spatial Data Infrastructures Unit of the European Commission Joint Research Centre to AMI SARL. of Luxembourg.

The study was commissioned in 2007 to investigate issues of expression and exchange of temporal information related to resources of importance to the INSPIRE Directive (2007/2/EC) of the European Parliament and of the Council establishing an Infrastructure for Spatial Information in the European Community. The purpose of such an infrastructure is to assist policy-making in relation to policies and activities that may have a direct or indirect impact on the environment. The Directive was approved on the 14th March 2007, and came into force on the 15th May 2007.

INSPIRE should be based on the infrastructures for spatial information that are created by the Member States. Such infrastructures should be designed to ensure that spatial data are stored, made available and maintained at the most appropriate level; that it is possible to combine spatial data from different sources across the Community in a consistent way and share them between several users and applications; that it is possible for spatial data collected at one level of public authority to be shared between other public authorities; that spatial data are made available under conditions which do not unduly restrict their extensive use; that it is easy to discover available spatial data, to evaluate their suitability for the purpose and to know the conditions applicable to their use.

To achieve these aims, the Directive focuses in particular on five key areas: metadata, the interoperability and harmonisation of spatial data and services for selected themes (as described in Annexes I, II, III of the Directive); network services and technologies; measures on sharing spatial data and services; coordination and monitoring measures. The text of the INSPIRE Directive is available from the INSPIRE web site (www.ec-gis.org/inspire). Member States are required to bring into force national legislation, regulations, and administrative procedures necessary to comply with the Directive by the 15th May 2009.

To ensure that the spatial data infrastructures of the Member States are compatible and usable in a Community and trans-boundary context, the Directive requires that common Implementing Rules (IRs) are adopted in a number of specific areas. These IRs will be adopted as Commission Regulations or Decisions, and will be binding in their entirety. The Commission is assisted in the process of adopting such rules by a Regulatory Committee composed by representatives of the Member States and chaired by a representative of the Commission (this is known as the Comitology procedure [1]). The Committee was established in June 2007. According to the INSPIRE Directive, the IRs on metadata must be adopted within one year of the entry in force of the Directive, i.e. by 15th May 2008.

The INSPIRE Directive requires (Art. 5) Member States to ensure that metadata are created for the spatial data sets and services corresponding to the themes listed in Annexes I, II, and

[1] An explanation of the process for the development and adoption of the implementing rules is contained in Section 3 of the Work Programme 2007-09 see http://inspire.jrc.it/reports/transposition/INSPIRE_IR_WP2007_2009_en.pdf

III, and that those metadata are kept up to date. Article 5 also specifies the information that must be included in the metadata, while Article 11.2 indicates the minimum combination of search criteria that discovery services implemented by the Member States must provide in querying the content of metadata.

Neither Article 5 nor Article 11 makes specific reference to temporal information as being required in metadata or as a criterion for searching metadata. However, Article 8 requires that information on the temporal dimension of data is included in the Implementing Rules for the interoperability of spatial data sets and services.

Against the background of these considerations, the international Drafting Team advising the European Commission on the preparation of the draft Implementing Rules for Metadata recognised that almost all themes have some requirements for temporal information which could be a factor in searching for data, and that there are other themes, particularly in Annex III (Atmosphere, Meteorology and Oceanography) where the data are fundamentally temporal, and where the classification of repositories, let alone data, is commonly organised by date and time. It would therefore be ineffective to address these themes without including temporal information in the INSPIRE discovery metadata.

The Metadata Draft Implementing Rules (IR) identified general requirements, which may not be achievable with present mechanisms. These requirements are currently either inadequately specified in metadata standards or are poorly (or not at all) implemented in existing search mechanisms using metadata. Since the remit of the Drafting Team is to identify detailed requirements and to describe best practice, or to recommend new directions, the IR deferred the detailed specification of temporal extent metadata for handling fundamentally temporal themes to the results of this study.

In the INSPIRE Metadata Implementation Rules[2] (Version 3, dated 26 October 2007) which has been the basis for this study, four elements of temporal information are defined:

- The time period covered by the content of the resource (also called the temporal extent of the resource)
- The date of publication of the resource
- The date of last revision of the resource if the resource has been revised
- The date of creation of the resource if it has not been revised.

The first element is related to the content of the resource, e.g. if a document, that itself may have been published in 2000, describes meteorological statistics of a country in 1999 – in that case the temporal information will indicate “1999”. The next three elements are related to the lifecycle of the resource; in the example above this would include the year, “2000”, that the document was published, revised or created.

[2] http://inspire.jrc.it/reports/ImplementingRules/INSPIRE_Metadata_ImplementingRule_v3_20071026.pdf

A major use of INSPIRE metadata will be as an exchange format for metadata that exists in various related domains, particularly covering geographic, geologic, meteorological and environmental information.

It is therefore important to have an overview of the way these various related domains record and express temporal information for the INSPIRE metadata elements mentioned above.

This document is the result of a study that looked at the relevant domains and derived a set of recommendation for the drafting team of the INSPIRE Metadata Implementing Rules in order to help establish the environment to enable appropriate mappings between existing domain-specific metadata and INSPIRE metadata and to gain some insight in other issues that may be relevant when building services based on the INSPIRE infrastructure.

2 OBJECTIVES OF THE STUDY

The objective of the study was to investigate research issues related to temporal information, together with actual usage of temporal information in existing collections and services in relevant domains, particularly geography, geology and meteorology.

As a result of the study, recommendations were to be proposed for approaches to mapping date and time information in areas relevant to the INSPIRE metadata implementing rules, specifically areas concerned with geographic and meteorological information and other areas related to environmental information.

The specific tasks to be performed were detailed as follows:

1. Review of the literature on temporal information – and temporal reference (i.e. extent and location) metadata (including ISO standards);
2. Review practice and significant use cases in different thematic communities including at a minimum the geographic, geological and meteorological communities on the way temporal information and temporal reference are stored in the data and handled in metadata;.
3. Identify criteria which distinguish temporal information required for data evaluation or usage, and temporal information required for discovery:
    a. To comment on the practical applicability of the criteria;
    b. To gather metadata examples of temporal reference metadata in different thematic communities, in any code or structured text;
    c. Identify mechanisms for using metadata describing different temporal elements for searching.
4. Produce recommendations to the Commission for the Metadata Drafting Team on
    a. Required temporal reference (i.e. extent and location) information for discovery, including distinction between “search” and “present”;
    b. Ways to implement the recommendations through INSPIRE discovery services
    c. Implications in respect to performance related to searching through large data volumes
    d. Proposals to amend relevant standards (e.g. ISO) or specifications (e.g. OGC) if required.

3 APPROACH

As given by the terms of reference, the work was performed in four stages, in line with the task description above:

3.1 Review of literature

The review looked at publicly available information on current research related to expression of time, specifically in the areas of interest (environmental data, scientific datasets, geography, geology and meteorology), identifying major strands of research and investigation. Furthermore, various standards related to the expression of time were considered, including ISO 19108, ISO 8601 and the Date and Time format of the World Wide Web Consortium as well as approaches used by the US Federal Geographic Data Committee and standards for geological eras developed by the International Commission on Stratigraphy.

3.2 Review of practice

Through a questionnaire distributed to contact persons in relevant organizations, current practices and opinions were identified in the main domains, including the environment information domain, the geographic information domain, geological information domain and the meteorological domain. Furthermore, experts from a more general background were contacted to cover the issues from a more general, cross-domain perspective.

3.3 Identification of criteria and analysis

Based on the outcomes of the survey responses, the various approaches were analysed, identifying commonalities and differences, and benefits and drawbacks of these approaches with respect to their relevance in cross-domain environments.

3.4 Recommendations

Based on the analysis in stages 1-3, a set of recommendations were drawn up that, taking into account the various roles that temporal information can play in the data in the domains covered, propose an approach to express the temporal aspects in the Temporal Reference element in the INSPIRE Abstract discovery metadata element set.

4 RESEARCH RESULTS

4.1 Main standards

During the research phase, the main base standards identified for expression of temporal information with relevance for the INSPIRE infrastructure are:

ISO 8601:2004

This International Standard includes specifications for a numeric representation of information regarding date and time of day and specifies representation of the formats of these numeric representations.

Dates are expressed as strings CCYY-MM-DD (century/year/month/day) with the use of hyphens optional. Right-hand truncation is permitted (e.g. “20” is used to represent the range of years 2000-2099, “2007” for the year 2007, “2007-12” for the month December 2007, and “2007-12-28” for the day 28 December 2007. Times are expressed as strings HH:MM:SS,FF (hours/minutes/seconds followed by a decimal fraction separated by a comma or decimal point) with the semicolons optional. Right-hand truncation is permitted (by omission, e.g. “13” for 1 o’clock in the afternoon) as well as left-hand truncation (replacing missing elements with

leading hyphens, e.g. -00:00 for the top of the hour). In combined date/time expressions, the two parts are separated by “T”, e.g. “2007-12-28T13:00”. It is also possible to express durations (e.g. “P2W” to represent two weeks) and ranges (<start point>/<end point>, e.g. “2007-12-28T13:00/2007-12-28T13:10”).

This standard is widely used in many domains and, in fact, is referenced as the basic expression mechanism for date and time in almost all of the standards summarised below, as well as in standards with a wider scope such as ISO 19115, OGC’s Geographic Markup Language GML and WMO Core metadata specification.

ISO 19108:2002

This International Standard defines the standard concepts needed to describe the temporal characteristics of geographic information as they are abstracted from the real world. It is part of the set of standards for geospatial information commonly referred to as the ISO 19100 series.

The standard addresses the temporal extent of geographic information, not the temporal information related to the lifecycle of the resource. It treats time as a dimension analogous to spatial dimensions, defining a geometry and topology for time, with points in time occupying a position in relation to a temporal reference system and intervals measured as distances in that reference system. Whenever dates or times are to be encoded, this standard refers to the format specified by ISO 8601, while it also makes it possible to use other types of reference systems, such as ordinal ones defining named periods.

W3CDTF (published 27 August 1998)

This specification, the World-Wide Web Consortium’s Date and Time Format, defines a profile of ISO 8601, restricting the supported formats to a small number “likely to satisfy most requirements” and used extensively in the Web environment, including the Semantic Web.

The main restrictions are:

- hyphens are mandatory to separate the elements of a date
- right-hand truncation is only allowed up to the year, e.g. “2007-12” and “2007”, not “20”
- left-hand truncation is not allowed, e.g. time cannot be expressed without a date
- durations and ranges are not included

TimeML and ISO/CD 24617-1

TimeML defines a formal specification language for events and temporal expressions. The main objective is to be able to tag news items and other current textual information.

It addresses two objectives, (a) the representation of events and temporal expressions and (b) the anchoring of ordering dependencies that may exist in a given text describing events in time. The basic temporal information is contained in TIMEX3 tags that allow for dates, times, durations and sets (e.g. repeated intervals) with dates and times expressed according to W3CDTF.

ISO/CD 24617-1 standard is currently under development in ISO TC37/SC4 and is scheduled to be published in 2009. It includes a formal specification of TimeML and additionally discussion semantic issues, mainly related to the way natural languages express temporal information using verb tense and aspect and words to express temporal relations.

ICS Geological Time Scale 2004

The Geologic Time Scale 2004 was completed by the International Commission on Stratigraphy in a multi-year project which summarised the history and status of boundary definitions of all geologic stages, compiled integrated stratigraphy (biologic, chemical, sea-level, magnetic, etc.) for each period, and assembled a numerical age scale from an array of astronomical tuning and radiometric ages.

4.2 Harmonisation of temporal expressions

A related issue to processing temporal information mentioned above occurs when information from various sources is brought together where the way that temporal information is expressed or the precision of this information varies, is the need to harmonise the information. This can involve the processing of ranges but also the conversion of reference systems (e.g. Julian to Gregorian calendar), expansion of named periods, aligning of precision (e.g. expansion of "1999" to "19990101/19991231") and estimates of uncertain dates.

The result of such harmonisation may then be presented as a timeline when various resources resulting from a search are presented as icons or other graphical representations along a timeline. An example of such presentation can be seen at http://e-culture.multimedian.nl/demo/search that gives access to cultural resources from many collections for which temporal information was pre-processed and harmonised to show related resources on a timeline at the bottom of the page.

Another area where harmonisation is relevant is where resources are related to geological periods. The work of the International Commission on Stratigraphy is relevant here, both in establishing the 2004 Geologic Timescale as well as in the GeoWhen project, aiming to reconcile all regional and legacy vocabularies into a consistent whole.

4.3 Processing ranges: reasoning and coalescing

A part of research into reasoning on temporal information is concerned with the processing of temporal ranges across resources. An example is the case where two resources contain measurements of the same phenomenon over different periods. Bringing these resources together, either in the same database or in a result set from a query, enables some way of interpreting the various ranges. This processing may look at fairly simple relations, e.g. which period is earlier and which is later, or more complex relations, for example determining whether ranges overlap fully or partially, are completely disjunctive, or are precisely sequential.

Results of the research identify performance problems resulting from doing complex processing in on-line environments, for example to process result sets from distributed searches, and seem to point to the need for off-line pre-processing of as much information as possible to achieve acceptable response times.

4.4 Event description

Event description is currently a main topic in research environments such as the Semantic Web, but also in research that is concerned with the tagging of current news items. These approaches are primarily concerned with text-based material where event-related information, including temporal information, is to be tagged in-line so that subsequent machine-processing of the text can automatically derive and index this temporal information.

As the emphasis is on current news and other contemporary resources, all these developments refer to ISO 8601, or the simplified W3C profile, W3CDTF, as the main way to express date and time.

5 SURVEY DISTRIBUTION AND RESPONSE STATISTICS

The survey (copy included in Annex A) was sent out to 22 respondents on 29 October 2007. By the end of November 2007, 8 of the 22 invited respondents have not reacted, while two people explicitly declined the invitation.

One of the respondents, the director of Eurogeographics, asked and was given permission to forward the survey to this organisation's membership. This secondary distribution reached 51 organizations.
At the end of November, 30 responses were received with one response covering two application areas (seismology and groundwater). Two organizations submitted two responses (German Bundesamt für Kartographie und Geodäsie and the Swiss Office of Topography swisstopo), which brings the number of participating organizations to 28. Of the responses, 18 were from the membership of Eurogeographics (17 in geography, 1 in geology).

The organisations that participated in the survey are listed in Annex B.

The distribution of the responses over domains was as follows:

- Geography (including cadastre, land surveying, land registration, mapping, cartography, topography, geodetic surveying, GIS data and geospatial data capture): 18 responses
- Geology: 5 responses
- Meteorology: 4 responses
- Environment: 2 responses
- Research: 2 responses.

As a result of the secondary distribution to the members of Eurogeographics, the results are biased towards the domain of geography. In terms of the analysis of the results, this is not an issue as the responses are considered qualitatively and not quantitatively.

Countries represented in the responses are Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Iceland, Norway, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom and the USA.

6 ANALYSIS OF SURVEY RESULTS AND CONCLUSIONS

6.1 Lifecycle information - elements

In terms of temporal information related to the lifecycle of the information managed, the picture that emerges is fairly coherent with respect to the dates that are included in the INSPIRE Metadata Implementation Rules (creation, revision, publication) that are mentioned by many respondents across the domains:

Element	Total response (31)	Environment responses (2)	Geography responses (18)	Geology responses (5)	Meteorology responses (4)	Research responses (2)
Creation	24	2	16	3	2	1
Revision	18	2	14	2	0	0
Publication	14	2	6	4	1	1

Various respondents, spread out across all domains except for the pure researchers, also mentioned that information on the validity of resources is important, something that is not currently foreseen in INSPIRE. The only other lifecycle element mentioned more than once was the date for the version which could be interpreted as another way of indicating the revision of a resource.

Element	Total response (31)	Environment responses (2)	Geography responses (18)	Geology response s (5)	Meteorology responses (4)	Research responses (2)
Valid-from	11	2	6	1	2	0
Valid-to	8	2	4	1	1	0
Version	3	1	1	1	0	0

A number of other elements were mentioned once: ‘Status’, ‘Validation-start’, ‘Validation-end’, ‘Cycle’, ‘Last-updated’, ‘Business valid-from’ and ‘Business valid-to’. It seems that many of these are fairly close to elements in the two tables above.

6.2 Lifecycle information – format

In the majority of cases, the responses referenced either ISO8601 directly (11 out of 31 responses), or indirectly via reference to ISO19115 and 19139 (2), WMO recommendations (2), the TimeML standard (2), national standards that are based on ISO 19115 (2), or by explicit mentioning formats that are compatible with ISO 8601 (2). In four cases the response was that there was no standard approach with time and date information provided as free text, while in three cases temporal information was captured in a structured, non-ISO 8601 format (DD-MM-YYYY) for which a trivial conversion to ISO 8601 could be made.

6.3 Temporal coverage

When considering temporal information related to the data itself, which is covered by the INSPIRE temporal coverage metadata element, the picture is more differentiated. The various domains have different perspectives:

- Geography appears not to have sophisticated dating requirements for primary data other than the ability to register what the period of validity of the data is (e.g. recording during which period a certain geographical feature existed). This information is mainly related to

a timescale in the order of a couple of hundred years until now, with a granularity down to the day. Temporal information is rarely imprecise of unknown, although this is not always true for older data. Open ranges are common. This information is usually expressed using ISO 8601 or equivalent formats.

- Geology covers a very wide range of temporal coverage of the primary data. On the far end, there is a need to identify geological eras, going back billions of years; on the near end, recording of current events like earthquakes needs recording on a scale of milliseconds. Temporal information can often be unknown or imprecise, while there is also usually an intrinsic error rate in dating of geological events. Information is expressed using geological standards developed in the context of GML, the Geography Mark-up Language, which refers to ISO 8601 for time instants and periods and ISO/IEC 11404 for time durations and recurrence. For geological periods, controlled vocabularies are used, with the Geological Time Scale 2004 of the International Commission on Stratigraphy being mentioned as a candidate common vocabulary. It needs to be noted, though, that these types of vocabularies are not absolutely stable but experiences changes over time on the basis of ongoing research.
- Meteorology covers ranges of a couple of centuries in the past, through actual weather situations to forecast of a number of days. Specifically, relationships between meteorological events and observations and validity of forecasts are important. Granularity ranges from “sub-minute” to days and months. The temporal information is rarely unknown or imprecise. Most temporal information can be expressed using ISO 8601.
- The domain of environment uses information from all of the domains above and therefore has an aggregated set of requirement of historical and actual data, but also looks at longer-term forecasting up to decades and centuries. Granularity is in the order of hours, with open ranges. Temporal data can be expressed using ISO 8601.
- In the research domain, an important issue mentioned is the problem of reasoning across different ways of expressing temporal information, different granularities and the relations between overlapping ranges. An important standard for the expression of temporal information is TimeML, which allows the expression of time and date in context of events in XML documents. The expression of date and time is again compatible with ISO 8601, with the exception of BC dates, where ISO 8601-2004 uses a minus sign and TimeML uses the two leading characters “BC”. This is however not likely to be relevant for INSPIRE applications.

There is no clear trend in the use of temporal information for search access within the domains. In geography, there were 8 positive and 6 negative responses when asked whether end users can search on temporal aspects; in geology 4 positive and 1 negative; in meteorology 1 positive and 2 negative; in environment 1 positive and 1 negative. In some cases, all available temporal fields are used for end user searches, but in other cases temporal information is not stored as metadata and indexed for searching but used to drive navigation.

When asked what level of detail would be useful in a cross-domain environment, answer vary from “time” with granularity of seconds for meteorology, through “date” related to creation, update and validity of resources in the geographic domain, to “stratigraphic periods” for geology.

7 CRITERIA FOR THE USE OF TEMPORAL INFORMATION

One of the objectives of the study was to identify criteria which distinguish temporal information required for data evaluation or usage, and temporal information required for discovery. In the following sections, we will first look at the temporal elements that are defined in the Implementing Rules for INSPIRE metadata, and then analyse some user requirements of potential services that may be developed for discovery and for data evaluation on the basis of INSPIRE metadata.

7.1 Temporal information in INSPIRE

The Draft Implementing Rules for Metadata (Version 3) defines four metadata elements to express temporal information related to a resource:

- *Temporal extent*, the period covered by the content of the resource
- *Date of publication*: date when the resources became available or entered into force
- *Date of last revision*: date when the resource was last modified
- *Date of creation*: date when the resource was created if it has not been revised.

Furthermore, temporal information is defined for the management of the metadata in an element *Metadata date* specifying the date when the metadata record was created or updated.
Please note that in the survey results described in chapter 6 above it was indicated that also elements to specify the period of validity may be important for many resources. This is information that may be contained in the temporal extent element, although this is a bit of a stretch given the semantic definition of the element.

The five temporal metadata elements above can be grouped into three categories:

- The instance or period that the content of the resource is related to, for example containing a description or holding a set of measurements for a period of time: *Temporal extent*
- Dates related to the lifecycle of the resource: *Date of publication*, *Date of last revision* and *Date of creation*
- Date related to the lifecycle of the metadata: *Metadata date*

In the following paragraphs, we will consider how these elements could be related to possible INSPIRE services.

7.2 Use of temporal information for discovery

Looking at the use of temporal information for discovery, users may be interested in two aspects:

- A particular date or time period in the lifecycle of the resource. An example could be a search for resources that were published on a certain date or in a certain year. The metadata elements relevant for such a service are *Date of publication*, *Date of last revision* and *Date of creation*
- A particular date or time period during which an event took place which is described in the resource. An example could be a search for the statistics of rainfall in a particular time

period. The metadata element relevant for such a service is *Temporal extent*. The validity of the resource could also be relevant for such a service.

Another service that is not directly related to a user interface is the selection of metadata records, for example for a harvesting service at a portal or aggregator site that selects metadata records from other information providers to build an integrated index. Such a service would use the *Metadata data* element to determine if metadata records have changed since the last selection.

7.3 Use of temporal information for data evaluation and use

The way temporal information is used for data evaluation and use is, even more than in the case of discovery services, dependent on exactly what the data evaluation or use service tries to achieve. It needs to be noted that the INSPIRE metadata, as it is effectively a definition of a common subset of domain-specific metadata that can be delivered by the various related domains, can only provide a subset of services that can be delivered within a single domain. Again, we can think of two sets of services in this area:

- A service could be constructed that analyses information from resources discovered in various data sets that were created, modified or published in a certain period. This could be used to determine how much knowledge there was in that period of a certain phenomenon. Such a service would use the temporal elements related to the lifecycle of the resources.
- Another service could be imagined that combines the content of various resources from different sources, for example adding meteorological information to geological events that took place at a particular time. Such a service would use the element *Temporal extent* to enable the combination of the various types of data.

It needs to be noted, however, that for most of the data evaluation services there is a need for the content of the resource to be available for processing. In most cases, metadata about the resource will not be expressive enough to perform complex processing related to the content.

7.4 Specific issues for large data volumes

A number of issues can be identified that are relevant for situations with large data volumes that may be relevant in the INSPIRE infrastructure. In practice, the way these issues will be encountered and how they can be resolved depends for a large part on what the INSPIRE-based services try to achieve.

- Creation of metadata

  Not all collections of relevance to INSPIRE will have metadata that is fit for indexing. Some services rely on navigation to get a human user to a relevant resource rather than offering search services. In all cases, where metadata needs to be created, a decision needs to be taken whether to do this manually or automatically. There will be a trade-off: manual creation of metadata is a resource-intensive process while automatic generation may not be able to generate all of the necessary contextual detail.

- Conversion of metadata to conform to common standards

  Driven by the needs to share information in the INSPIRE infrastructure, data providers that use non-standard approaches may consider to convert to more standard approaches (e.g. using ISO 19108, ISO 8601). For large data volumes, this may be an expensive

option, especially where expression of time and date information has been done in an uncontrolled way (e.g. as free text).

- Distributed search

  In approaches where user searches are sent out across a number of systems in real time, performance is going to be dependent on the number of distributed servers that are included, the performance characteristics of each of those servers and the number of hits that can be expected overall. Pre-processing search arguments to match the preferences of individual search systems, or post-processing the results may degrade the response times to unacceptable values. As a consequence, such a distributed approach can only be expected to work in a fairly homogeneous environment.

- Harvesting and aggregation

  Aggregation of information related to a specific application area has been done in other domains (e.g. in the Open Archives Initiative for scientific information, in e-Government and in the cultural domain) that provide portal access to distributed resources though a central index. Harmonisation of temporal information can be done offline, for example as part of the ingest process when harvested metadata is pre-processed before building the indexes.

7.5 Other relevant issues

Two more relevant issues should be mentioned here:

- Granularity and precision
- Quality and reliability

Below, a number of observations are included that may be considered as part of further guidelines for content contributors in the INSPIRE infrastructure.

7.5.1 Granularity and precision

In any database or data set, managers of content collections will naturally select the appropriate granularity and precision of temporal information they include and maintain, just as they will decide on exactly how they describe other aspects of the resources they manage.

Issues may arise when descriptions of resources from different collections are brought together, either physically through metadata harvesting or logically in a distributed search service. Some of these issues related to the reasoning and coalescing of ranges and harmonisation of temporal expressions are described in Sections 4.2. and 4.3.

Resolution of these issues may well require substantial efforts to produce acceptable results when using such aggregated or combined collections. The extent to which such issues can be resolved satisfactorily depends on the particular resources and user expectations, and it will be important to foresee substantial efforts when building actual services as well as to carefully manage user expectations, making sure that users understand that there is a trade-off between depth (accessing individual collections using a full set of functionalities) and breadth (accessing a wider set of collections with limited functionalities).

7.5.2 Quality and reliability

It goes almost without saying that the usefulness of a service and a data collection is proportional to its quality and reliability. Especially when a service is based on a set of collections produced and managed by many independent content providers, the management of quality can be particularly complex.

In building services on the basis of the INSPIRE infrastructure, it may well be considered useful to include the specification of service level agreements in the development of services. In the case of temporal information, in addition to the semantic and syntactic aspects defined by the Implementation Rules, agreements as to the precision of the temporal information as well as the procedures and guarantees related to quality assurance may be considered.

After all, user acceptance and provider satisfaction of services built on the basis of the INSPIRE infrastructure will depend critically on whether relevant information can be found and searches do not result in a great amount of irrelevant results.

8 RECOMMENDATIONS

Based on the research, the analysis of the responses to the survey and further contacts with experts in the area of practical harmonization of date and time information, the following recommendations are proposed:

8.1.1 Recommendation 1

In terms of resource lifecycle information the elements most mentioned are: date of creation (24 out of 31 responses), date of revision (18), date of publication (14), date valid-from (11) and date valid-to (8).Whereas this result validates the INSPIRE specification for the first three of those (creation, revision, publication), it appears that the dates related to the validity of the resources are important across the domains surveyed.

In practical cases, it needs to be analysed if this validity information is about the lifecycle of the resource, e.g. when this information indicates that the resource can only be used from a certain date, or that the information will be replaced by a new version at a specific point in time in the future, or if the validity information is really about the content of the resource, e.g. because a geographic feature started or ceased to exist at a certain time (such as the border between the Czech and Slovak Republics).

Therefore, it is recommended that:

INSPIRE consider adding temporal elements for validity of resources to the core metadata set, and issues a recommendation to use the 'Temporal coverage' element whenever the validity information is related to the validity of the content of the resource.

8.1.2 Recommendation 2

From the survey it appears that in most cases ISO 8601 (or a similar format) is used as the basis for expression of points in time in the calendar. Integrating information from the various domains will be rather straightforward if all information in the INSPIRE environment is exchanged in this ISO format.

Therefore, it is recommended that:

INSPIRE make it mandatory that points in time are expressed (or converted to) ISO 8601, with the only exception being the expression of geological timescales.

8.1.3 Recommendation 3

For expression of time periods, for example where a dataset is based on a measurement that covers an extended period of time, standard mechanisms are available. ISO 19108 defines an element TM_Period defined by two TM_Instants that mark its beginning and end; Dublin Core has defined a structured value expression called DCMI Period and ISO 8601 has an encoding for periods of the form <start date>/<end date>. Neither ISO 19115 nor Dublin Core (the two encoding options for INSPIRE metadata) use the ISO8601 mechanism to express periods but allow more expressive ways to define periods (e.g. defining temporal reference systems, other ways of encoding date and time or naming periods)

Therefore it is recommended that:

INSPIRE mandate the expression of periods of time using encoding of ISO19108 TM_Period if metadata is exchanged according to ISO 19115/19119 and using encoding of DCMI Period if the exchange is according to Dublin Core.

8.1.4 Recommendation 4

The various domains register date and time information with different granularities. Given the type of resources described by the various domains, it is unlikely that this will lead to problems in most cases (e.g. the timescale of geological eras does not interfere with date information relevant for geographic resources). It does appear to be useful for the future development of INSPIRE-based services and products to avoid losing granularity and precision in all temporal information exchanged.

Therefore, it is recommended that:

INSPIRE issue a recommendation that the highest available precision is preserved in the exchange or conversion of temporal information

8.1.5 Recommendation 5

For geological information, the use of controlled vocabularies is important to indicate geological periods, in order to be able to cross-reference information from different sources.

Therefore, it is recommended that:

INSPIRE encourage the use of standard vocabulary for geological timescales, such as the Geological Time Scale 2004 developed by the International Commission on Stratigraphy.

8.1.6 Recommendation 6

Even if different information providers use the same standard for expressing temporal information, metadata instance values may not always be directly comparable. This could for example be caused by a different way of expressing temporal information (e.g. one provider may say “2003” while another says “2003-01-01 through 2003-12-31” for the same period), or by different precisions (e.g. one provide may say a resource was published in “2003” while another says more precisely it was published in “2003-11”). It depends on the types of services that will be built on the basis of the INSPIRE metadata how issues of reasoning over the temporal information and of matching queries to resources can be resolved in order to return useful results from user queries.

If the services are based on online distributed searches, only fairly basic reasoning can be done, e.g. determining how a period given as a search argument is related to periods that are associated with resources (e.g. whether they overlap or not), while if metadata is harvested from multiple providers into a central index or portal, offline harmonization of temporal information can be done, e.g. mapping all temporal information to a standard timeline, which would allow more specific reasoning to be done (how well search arguments fit with periods associated with resources).

Therefore, it is recommended that:

INSPIRE determine on a case-by-case basis how issues can be solved with respect to reasoning and matching in actual services that will be based on INSPIRE metadata.

9 RESOURCES WITH ANNOTATIONS

[1] Jun-ichi Akahani et al. Towards the Real-World Semantic Web — Web Search based on Spatial and Temporal Metadata.

http://www.ntt.co.jp/tr/0306/files/ntr200306071.pdf

Article investigating the way that spatial and temporal aspects of "real-world information" can be used in Web search engines, proposing the use of ontologies for spatial and temporal characteristics that can then be used at search time to match query arguments (e.g. "now") to (static) time intervals.

[2] James F. Allen and George Ferguson, Actions and Events in Interval Temporal Logic, July 1994

ftp://ftp.cs.rochester.edu/pub/papers/ai/94.tr521.Actions_and_events_in_interval_temporal_logic.ps.Z

Article presenting a representation of events and action based on interval temporal logic motivated by work in natural language semantics and discourse, temporal logic, and AI planning and plan recognition.

[3] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)

http://www.ec-gis.org/inspire/directive/l_10820070425en00010014.pdf

Legal basis for activities to establish the infrastructure, which gives the Member States the obligation to "provide descriptions of available spatial data sets and services in the form of metadata" and states that implementing rules should address the aspect of "information on the temporal dimension of the data"

[4] Draft Implementing Rules for Metadata (Version 3). European Commission, 2007-10-26.

http://www.ec-gis.org/inspire/reports/ImplementingRules/INSPIRE_Metadata_ImplementingRule_v3_20071026.pdf

Commission's Draft proposal for Implementing Rules (IRs) on metadata as required by the INSPIRE Directive.

[5] Dyreson, Curtis E. Temporal Coalescing with Now, Granularity, and Incomplete Information. ACM SIGMOD International Conference on Management of Data , 2003

http://portal.acm.org/citation.cfm?id=872779&dl=ACM&coll=portal

Article presenting a strategy for coalescing temporal extents in a combination of offline processing of fixed time information and online processing of information that can only be evaluated at query time (such as "now"). The article concludes that the partial coalescing strategy seems to be much faster than coalescing at run-time and can handle a great variety of time values.

[6] Federal Geographic Data Committee. Content Standard for Digital Geospatial Metadata. FGDC-STD-001-1998.

http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf

US standard providing a common set of terminology and definitions for the documentation of digital geospatial data. The standard establishes the names of data elements and compound elements (groups of data elements) to be used for these purposes, the definitions of these compound elements and data elements, and information about the values that are to be provided for the data elements. It includes temporal keywords using temporal keyword thesauri, and various types of time periods (single, multiple, range).

[7] GeoWhen Database

http://www.stratigraphy.org/geowhen/

Project aiming to reconcile the international stratigraphic standards with many of the regional and archaic naming schemes that appear in the literature. It combines the most recent ICS time scale with information on the estimated placement and length of other stages in order to put approximate dates on all of the stratigraphic intervals into a single self-consistent whole.

[8] F.M. Gradstein and J.G. Ogg, Geologic Time Scale 2004 – Why, How, and Where next!

http://www.stratigraphy.org/scale04.pdf

Article summarizing key features of the Geologic Time Scale 2004 and discussing how it was constructed and how it can be improved.

[9] Gary Houston. ISO 8601:1988 Date/Time Representations. 1993-01-31.

http://hydracen.com/dx/iso8601.htm

Article describing the ISO standard for numerical date/time interchange formats, ISO 8601:1988 including examples of complete, abbreviated and truncated formats and expression of periods of time.

[10] ISO 8601:2004. Data element and interchange formats – Information interchange – Representation of dates and times.

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40874

International Standard including specifications for a numeric representation of information regarding date and time of day as well as specifying representation of the formats of these numeric representations.

[11] ISO 11404:2007. Information technology – General-Purpose Datatypes (GPD)

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39479

International Standard specifying the nomenclature and shared semantics for a collection of datatypes commonly occurring in programming languages and software interfaces, including temporal information.

[12] ISO 19108:2002. Geographic information – temporal schema.

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26013

International Standard defining the standard concepts needed to describe the temporal characteristics of geographic information as they are abstracted from the real world.

[13] ISO 19115:2003. Geographic information – metadata.

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26020

International Standard providing the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.

[14] ISO/TS 19139:2007. Geographic information – Metadata – XML schema implementation

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32557

International Standard defining Geographic MetaData XML (gmd) encoding, an XML Schema implementation derived from ISO 19115.

[15] ISO/CD 24617-1. Language resource management — Semantic annotation framework (SemAF) — Part 1: Time and events

http://lirics.loria.fr/doc_pub/SemAFCD24617-1Rev12.pdf

Semantic Web-related standard under development aiming to provide a standard, consisting of basic concepts, metamodel and others, for the annotation language ISO-TimeML.

[16] Christian S. Jensen et al. The Consensus Glossary of Temporal Database Concepts - February 1998 Version.

http://www.cs.uky.edu/~dekhtyar/685/papers/Temporal.DB.glossary.pdf

Article containing definitions of a wide range of concepts specific to and widely used within temporal databases. In addition to providing definitions, the document also includes explanations of concepts as well as discussions of the adopted names.

[17] Markus Kuhn. A summary of the international standard date and time notation.

http://www.cl.cam.ac.uk/~mgk25/iso-time.html

Overview for authors of Web pages and software engineers explaining ISO 8601.

[18] Kiyong Lee, James Pustejovsky, Branimir Boguraev: Towards an International Standard for Annotating Temporal Information

http://www.korea.ac.kr/~klee/download/English/200608_iso-timeml_Beijing2006.pdf

Paper introducing project that aims at establishing a semantically coherent framework for temporal annotation as an international standard. The established framework provides a formal specification language called ISO-TimeML for temporal information mark-up and includes a specific set of guidelines for such mark-up.

[19] MultimediaN N9C Eculture project

http://e-culture.multimedian.nl/demo/search

Project with the objective to develop a set of e-culture demonstrators providing multimedia access to distributed collections of cultural heritage objects. The demonstrators are intended to show various levels of syntactic and semantic interoperability between collections and various types of personalized and context-dependent presentation generation.

[20] Stefano Nativi and Gil Ross. Temporal Extension metadata. Draft Version 0.1 (2006 September 16).

http://sdi.jrc.it/Research/publications/Temporal%20Extension%20Metadata%20Position%20Paper%2001.pdf

Article investigating some critical issues related to temporal information, such as relative or uncertain knowledge of temporal information, moving reference systems (e.g. “now”), ordinal relationships and named periods (e.g. “winter”), and wide varieties in granularity.

[21] Doug Nebert, Current standards, specifications, and practices of relevance to SDI development, FGDC Secretariat, June 2005.

http://www.anzlic.org.au/get/2413322725.ppt

Presentation outlining architectural components of Spatial Data Infrastructures, data standards and access and services.

[22] Open Archives Initiative.

http://www.openarchives.org/

International Initiative developing and promoting interoperability standards that aim to facilitate the efficient dissemination of content. OAI has its roots in the open access and institutional repository movements.

[23] OpenGIS Catalogue Services Specification

http://portal.opengeospatial.org/files/?artifact_id=20555

Document specifying the interfaces between clients and catalogue services, through the presentation of abstract and implementation-specific models. Catalogue services support the ability to publish and search collections of descriptive information (metadata) for data, services, and related information objects. Metadata in catalogues represent resource characteristics that can be queried and presented for evaluation and further processing by both humans and software. Catalogue services are required to support the discovery and binding to registered information resources within an information community.

[24] OpenGIS Geography Mark-up Language (GML) Encoding Standard version 3.2.1

http://portal.opengeospatial.org/files/?artifact_id=20509

Standard defining an XML grammar written in XML Schema for the description of application schemas as well as the transport and storage of geographic information.

[25] Jussi Rasinmäki. Modelling spatio-temporal environmental data.

http://www.agile-secretariat.org/Conference/mallorca2002/proceedings/dia25/Session_4/s4_Rasinmaki.pdf

Article presenting a conceptual model for environmental data with special emphasis on the ability to store spatio-temporal references of the data. Other aspects of the model are the ability to handle hierarchical data and semantics of the measurements.

[26] Relation between EN ISO 19115 and 19119 and the elements of the INSPIRE draft metadata implementing rules (informative). European Commission, 2007-10-09.

http://www.ec-gis.org/inspire/reports/ImplementingRules/metadata/MD_IR_and_ISO.pdf

Document aiming to demonstrate how EN ISO 19115 and EN ISO 19119 can be used to provide the metadata for INSPIRE.

[27] Frank Schilder and Christopher Habel. From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages. 2001. In: Proceedings of ACL'01 workshop on temporal and spatial information processing.

http://www.cs.brandeis.edu/~jamesp/arda/time/readings/schilder_habel.pdf

Article presenting a semantic tagging system that extracts temporal information from news messages

[28] Temporal Extension Metadata Implementation Model. Draft Version 0.5 (2006 September 16).

http://sdi.jrc.it/Research/publications/TempExtMetadataImplemModel_05.pdf

Specification of a model for expression of temporal metadata, defining the concepts of Instant, Interval, Composite Interval and Periodic Interval. The document states that the model can be implemented by ISO 19108 and recommends the use of ISO 8601 as the primary temporal reference system for use with geographic information.

[29] TimeML 1.2.1. A Formal Specification Language for Events and Temporal Expressions.

http://www.timeml.org/site/publications/timeMLdocs/timeml_1.2.1.html

Specification of the formal specification language for events and temporal expressions, TimeML, using the TIMEX3 tag to mark up explicit temporal expressions such as times, dates, durations etc.

[30] W3C Semantic Web Activity

http://www.w3.org/2001/sw/

Collaborative effort led by W3C and based on the Resource Description Framework (RDF), aiming to provide a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.

[31] W3C Time Ontology in OWL. Editor's Draft 6 September 2006.

http://www.w3.org/2001/sw/BestPractices/OEP/Time-Ontology

Document presenting an ontology of temporal concepts, OWL-Time for describing the temporal content of Web pages and the temporal properties of Web services. The ontology provides a vocabulary for expressing facts about topological relations among instants and intervals, together with information about durations, and about date and time information.

[32] WMO Core Metadata Profile of the ISO Metadata Standard, Version 0.3, July 2006

http://wis.wmo.int/2006/metadata/WMOCoreTextVer0.3.doc

Table providing an overview of the WMO Community Core Metadata Profile, a profile of ISO 19115:2003 with a number of deviations (adding elements for specific keywords, irregular geographic shapes, descriptions of satellite orbit and meteorological stations, and additional codes, e.g. for "reference" under CI_DateTypeCode and for "author" under CI_RoleCode, and for character sets, maintenance frequency, medium names, progress codes, scope codes, spatial representation types and topic categories).

[33] Misha Wolf and Charles Wicksteed. Date and Time Formats. 1998-08-27.

http://www.w3.org/TR/NOTE-datetime

Specification of the W3CDTF format which is a profile of ISO 8601, restricting the supported formats to a small number "likely to satisfy most requirements" and used extensively in the Web environment, including the Semantic Web.

[34] Alexander Zipf and Sven Krüger, TGML - Extending GML by Temporal Constructs - A Proposal for a Spatiotemporal Framework in XML.

http://smartkom.dfki.de/Vortraege/EML/TGML_Extending_GML.pdf

Article presenting a temporal XML schema to be used with the GML schema to realise a powerful spatio-temporal XML schema.

ANNEX A: QUESTIONNAIRE

Name:

Organization:

E-mail:

Your domain: E.g. geology, meteorology, geography, environment, land management, computer science/research etc.

Please answer the following questions for resource collections that you manage or investigate. Wherever appropriate or useful, please include URLs to documentation or online services.

1. Which date and time information concerning resource management are relevant in relation the information resources (and related services, if any) you manage or investigate? E.g. date of creation, revision, publication, valid-from, valid-to etc.

Your response:

2. Which kind of information with temporal dimension do you manage or investigate? E.g. geological processes, weather, climate, geography, natural hazards, water or air quality etc.

Your response:

3. What timescales does the temporal information relevant for these resources cover? E.g. contemporary timescale in the order of years or less, historical timescale measured in centuries, geological timescales measured in millions of years and named eras, etc. (including forecasts)

Your response:

4. Does the temporal information have uncertain, unknown or indeterminate aspects? Often, sometime, rarely, never? E.g. open-ended ranges, imprecise dates ("sometime during the last century"), relative times ("now", "in the next three days")

Your response:

5. How is the temporal information expressed, and specifically which standards or specifications are used for the expression of the temporal information, both for points in time and ranges? E.g. ISO 8601, W3CDTF, TimeML, controlled vocabularies for named time ordinal era terms

Your response:

6. Are temporal aspects searchable by users? If so, what search terms or expressions can be used? E.g. find all resources published before a certain date, find all resources that are relevant for a certain period

Your response:

7. If the resources you manage or investigate were to be integrated with information from other domains (e.g. to deliver cross-domain discovery services), what level of detail of temporal information would you consider useful for non-experts?

Your response:

Thank you for your time and contribution.

ANNEX B: RESPONDENTS

Belgium:

- General Administration of Patrimonial Documentation
- Institut Géographique National, Belgique (IGNB)

Croatia:

- State Geodetic Administration of Republic of Croatia

Czech Republic:

- Czech Office for Surveying, Mapping and Cadastre

Denmark:

- Kort & Matrikelstyrelsen

Finland:

- National Land Survey of Finland
- Industrial Ontologies Group, University of Jyväskylä

France:

- Bureau de recherches géologiques et minières (BRGM)

Germany:

- Bundesamt für Kartographie und Geodäsie (BKG)
- Bundesanstalt für Geowissenschaften und Rohstoffe
- Umweltbundesamt

Iceland:

- The National Land Survey of Iceland

Norway:

- Norwegian forest and landscape institute

Portugal:

- Instituto Geográfico Português

Romania:

- National Agency for Cadastre and Land Registration

Slovakia:

- Geodetic and Cartographic Institute Bratislava (GCCA)

Slovenia:

- Surveying and Mapping Authority of the Republic of Slovenia

Spain:

- Spanish Directorate General for Cadastre

Sweden:

- Landmäteriet

Switzerland:

- Office of Topography swisstopo

United Kingdom:

- British Geological Survey
- Met Office UK
- Ordnance Survey – GB

USA:

- National Geophysical Data Center (NGDC)
- Thomson Corp.

European Organisations:

- ECOMET
- European Environment Agency, Europe
- PRIMET

European Commission

EUR 23209 EN – Joint Research Centre – Institute for Environment and Sustainability
Title: Temporal Metadata for Discovery
Author: Makx Dekkers
Editor: Massimo Craglia
Luxembourg: Office for Official Publications of the European Communities
2008 – 31 pp. – 21 x 29.7 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593

Abstract
This report contains the findings of a study on temporal metadata for discovery commissioned by the Spatial Data Infrastructures Unit of the European Commission Joint Research Centre to AMI s.r.l. of Luxembourg. The study was commissioned in 2007 to investigate issues of expression and exchange of temporal information related to resources of importance to the INSPIRE Directive (2007/2/EC) of the European Parliament and of the Council establishing an Infrastructure for Spatial Information in the European Community. The objectives of the study were to review existing standards and practices in different thematic communities and make recommendations as input to the preparation of the Implementing Rules for Metadata required by the INSPIRE Directive.

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

Publications Office
Publications.eu.int