ISO-8601:2019 and the Extended Date/Time Format (EDTF)

First the adoption of EDTF (Extended Date Time Format) within RDA (the Resource Description and Access format), its fine-tuning within the TC 154 WG 5, the publication of ISO 8601:2019 and finally its adoption by national standards bodies for date and time representation formats has a profound impact on metadata standards as it effects date structures and temporal search: approximate dates, uncertain dates, year dates exceeding 4 digits, seasons, extended intervals, unspecified elements and and and..

Edward C. Zimmermann

<edz@nonmonotonic.net>

StandICT.eu

Temporal Data Types

- Instants: something happened at an instant of time.
- Intervals: length of time (e.g. 1 month)
- Period: An anchored duration of time (e.g. Summer School Holidays in Bavaria 30.07.2016 – 12.09.2016)

Dates and Times

- The dates found in objects or descriptions in metadata in Libraries etc. don‘t always have dates like 20 July 2021 17:00:00 GMT (that‘s 7 PM in Munich) to describe the date of objects or references.

Instead:

- Circa 1940
- Summer 1974 (Journals)
- Christmas 1933 (not meaning 25 Dec.)
- Some Tues in August 1943 or 1944

Semantics depend upon context (Instant or Period)

- mid September, 1516
- 19 April at 9am (as could be found in a diary entry or letter)
- Time without time zone or unknown place?
- 7 Nov 1918 from ~7 PM (King of Bavaria overthrown).
- October in the early 18th Century
- Sometime between the 1790's and 1799
- Pearl Harbor 7 Dec 1941 7:53 AM.
- 15 April 1865 7:22 AM (Lincoln's death in Peterson House DC—across from Ford Theater)
- about 16:30 on Monday, 26 April 1937 (Guernica)
- 25th of December 3 BCE (accepted as factually wrong both in day and year)

Location and other information needed to normalize.

The basis of the International Date is the „Georgian Calendar“ introduced by Pope Georgory XIII in Oct. 1582

to replace the „Julian Calendar“ (est. By Caesar 46 BCE, 708 AUC „*ab urbe condita*“—founding of Rome)

Originally in the Julian Calendar the first of the year as 1 May. This shifted to 1 Jan.

Not universally accepted. Protestant and Eastern Orthodox countries did not implement.

Old Style Dates, New Style Dates. Julian adjusted to 1 Jan (N.S.) or not?

Holy Roman Empire Catholic:

- 1 Jan adopted 1544
- Gregorian Calendar adopted 1583

Holy Roman Empire Protestant

- 1 Jan adopted 1559
- Gregorian Calendar adopted 1700

Russia: 1 Jan in 1700 and Gregorian Calendar in 1918. British Empire (incl. USA) adopted both 1752– 11 days off. Wednesday, 2 September 1752 (Julian), was followed by Thursday, 14 September 1752 (Gregorian).

George Washington was born Feb. 11 1731. That is why it is celebrated on 22 Feb.

Date/Time without time zone

This talk was announced for:

- 2016-07-20T17:00Z (7 July 2016 5 PM UTC)
- 2016-07-20T19:00 (7 July 2016 7 PM in ???)
- 2017-07-20 (7 July 2016, ?? Where?)

From the perspective of the announcer it is well defined but how does it compare to the view of others?

Time Ontology in Linked Data

- https://www.w3.org/TR/2016/WD-owl-time-20160712/

OWL-Time supports some uncertain expressions by means of

- interval relations accounting for "before," "after," (sometime) "during," etc.
- It does not allow for approximate and vague expressions such as "circa 560 BCE" or "sometime in the early 1920's."

Time Resolution of Measureable Instants: Seconds?

The highest currently agreed upon ("interoperable") resolution of time is **1/9192631770 of a second** (from the state transition of Caesium-133).

1 second can have at most 9,192,631,770 sub-points

1 day can have at most 24*60*60*9,192,631,770 or **794243384928000 elements**.

In Babylon time (as today in the Jewish Calendar) 1 hour is divided into 1080 parts (“she”, called part or ***chelek*** (חלק) in the Jewish Calendar).

1/72 of 1 degree of celestial rotation (1/360 day = gesh/mu-eš) :

360 degrees * 72 she / 24 hours = 1080 she / hour

60 minutes = 1 hour ☾ 18 she / minute (Hebrew calendar 76 Regaim = 1 Chelek).

The smallest measuring unit was 3 1/3 seconds (1/18 minute).

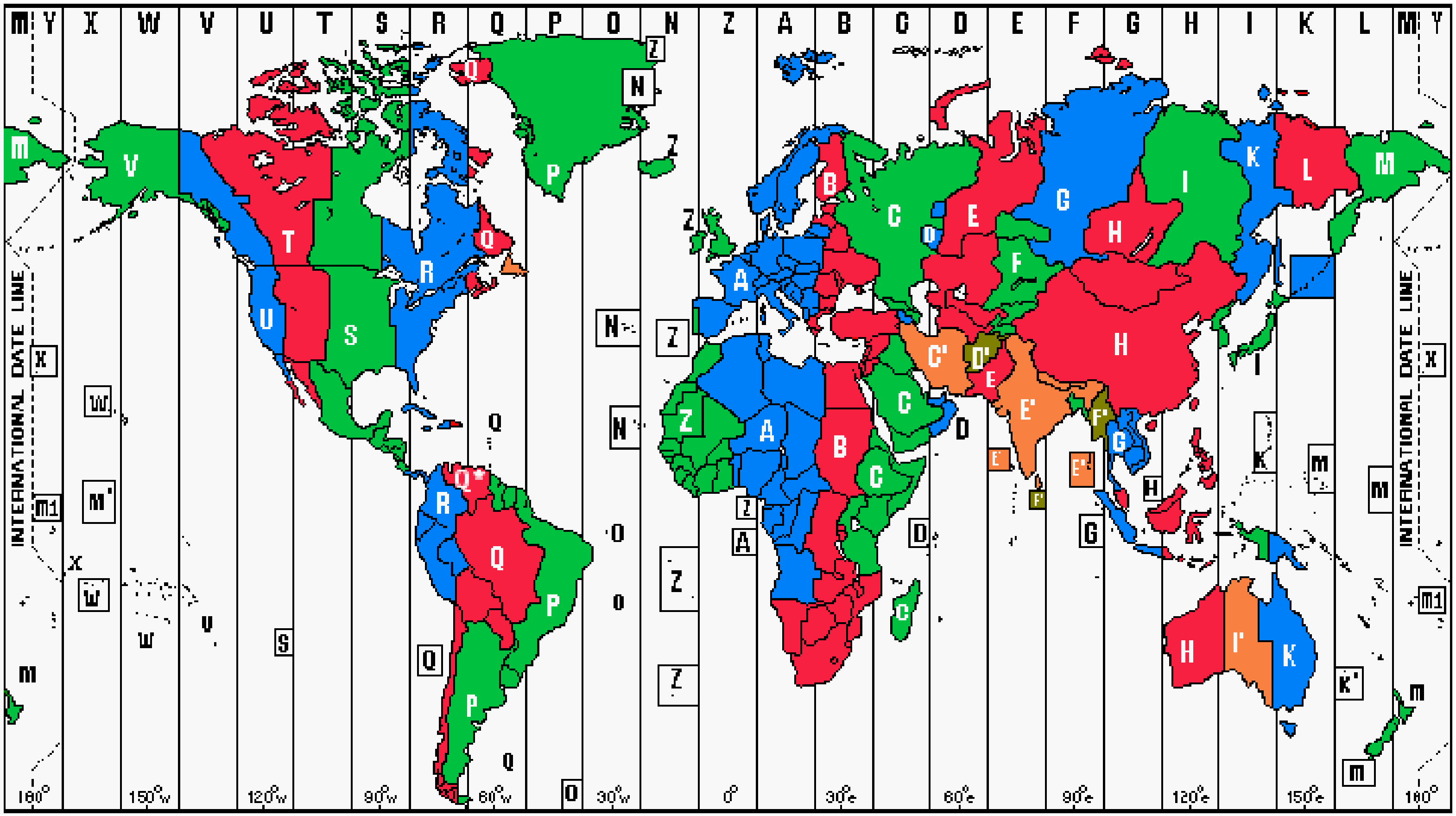
Date/Time Resolution

Many date types in engines use a resolution of seconds. Search is typically viewed as a numeric search where the date is converted into seconds since the Epoch (1 Jan 1970).

There are a number of approaches—e.g Lucene's precision-step and trie. Lucene uses milisecond precision and by default a step of 16 (same for double)—and assumes all times are UTC (and a day starts at 00:00:00 and ends at 23:59:59).

Speed of light is 299 792 458 m/s. DCF77 is Mainflingen near Frankfurt. 280.93 km. 1/1067 second. Around 1 ms(!).

Intl. Time Zones: Max Diff > 24hr



The report of an event in a location without time zone can match events reported on two different calendar days!

The International

- Time difference between UTC+14 and UTC−12 is that of **26 hours**.
- Time difference between UTC+13 and UTC−12 is of **25 hours**.
- Time difference between UTC+14 and UTC−11 is of **25 hours**.

- Kamchatka (Russia) uses UTC+12 (9 hours ahead of Moscow).
- Midway Atol, Jarvis Island and a number of uninhabited US Islands use UTC-11.
- Baker Island and Howland Island (uninhabited Oceania) use UTC-12 all year round
- Line Islands of Kiribati observe UTC+14 as standard time.
- Phoenix Islands of Kiribati, Tokelau, and Tonga observe UTC+13 all year round.
- Fiji and New Zealand observe UTC+13 in southern summers.
- Samoa observes UTC+13 as standard time and UTC+14 in southern summer.

Time Zone Conversions are not Date Independent

Time Zone also depends upon the year (Examples)

- In 2009, Samoa moved the International Date Line to the other side of its territory, which means that its time zone was changed from UTC−11 to UTC+13 and UTC+14 in the southern summer.
- In 1947 Hawaii went from GMT-10:30 to GMT-10. ***Pearl Harbor 7 Dec 1941 7:53 AM.***
- In 1912 Korea as a colony changed the time zone to GMT+09:00 to align with Japan Standard Time. On 5 August 2015, the North Korean government decided to return to UTC+08:30, effective 15 August 2015

NOTE: The name UTC was formally adopted in 1967 and GMT and UTC were the same until ’72. GMT is NOT LONGER a time zone.

Man on the Moon

- Neal Armstrong first set foot on the Moon?

  21 July 1969 02:56:15 GMT (Up until 1972 UTC and GMT was the same)
- The date is different when viewed from Los Angeles (GMT-8) than from Munich.
- “On July 20, 1969, the world watched as Neil Armstrong opened the hatch of the Eagle and stepped outside onto a barren lunar landscape. Buzz Aldrin followed him out the door and they became the first humans to ever walk on the moon. All the while, their third crewmate, Michael Collins, sat in a lonely orbit around moon in the Columbia command module.” -- http://www.space.com/26532-nasa-apollo-11-moon-landing-anniversary.html
- When talking about an event that took place in Munich around the same time as Armstrong set foot on the Moon yields a different day than someone in Los Angeles speaking of an event ….
- The event took place ON THE MOON so can‘t even talk about the time or date of the location of the event. Perhaps from the prespective of JPL in Pasadena? Cape in Florida? Washington DC? Some observatory?

What is the expression of a date as year mean?

- The date expression "1964" does not express the "whole year" but rather a measurement point with precision of year. Abe Lincoln, for example, was not murdered the whole day of 14 April 1865 but shot at an instant within that day. Its typically given that he was shot at 22:15 and died the next day at 7:22.
- Was he shot the "whole" of the minute?
- Did he die a whole minute long? Or, again, was it not both instants--- this time measured, resp. expressed, to minutes? In comparing two measures of two different objects with differing precisions---- here its dates but it could be temperature, length, mass or whatever--- one must normalize to a common precision and this ultimately is typically the coarser of the two. Imagine, for example, two people that measure the weight of same specimen using different scales. One person using a good quality analytic lab balance- -- for example readable to 0.01mg with 0.05mg repeatability--- may read 1.0146g while the other using a much cruder scale readable to grams may read only 1 gram. 1.0146g is not a greater measure than the 1g. Both have measured the same sample but at different precisions. The weights must then be "the same". Measuring another specimen the first reads on their analytic balance 1.0370g. The second person reads 1 gram. At the precision of the lab balance the second sample is heavier than the first but at the precision of the crude scale the two weigh the same. Turing now to dates the date "1964-04" if measured in year precision would yield "1964". In month precision "1964-04" is before "1964-06" but in year precision they are "the same" just as two babies born on the same day but at different times share the same birthday.

Intervals

- [sidetrack: The interesting problem is that an abstract time interval is not really a set of discrete points of undividable precision but a fuzzy continuum with measurements always at some precision. If we judge time considering precision we have a clear well defined element count for any interval and have a well defined set of points. The highest currently agreed upon ("interoperable") resolution of time is 1/9192631770 of a second (from the state transition of Caesium-133). If one second can have at most 9,192,631,770 sub-points then 1 day can have at most 24*60*60*9,192,631,770 or 794243384928000 elements-- by current computational standards a relatively small number (fits in 64-bit ints).] Does it matter if its defined? For, at least my, typical questions I don't think so. For any time point (expressed in any precision) I can answer the question: "Is it within the given interval?" Does "1960/1964-04" contain - "1963"? Yes - "1964"? Yes (since the precision of comparison can't ever be finer than the the precision of the least precise measure). - "1964-04-12"? Yes - "1964-05"? No Other typical questions such as: - Do the intervals overlap? - Is one interval included in the other? - Are the two intervals exclusive to one another (and, if so, their order)? similarly follow.... With a higher level of (measurement) precision, of course, it might turn out that the date expressed as "1964" in the comparison was really "1964-05- 12T12:30" and this NOT in the interval but is it really the same comparison? The state of knowledge, afteral, is different. The same observation, of course, holds for intervals with the same precision for start and end when we increase precision. Increasing the precision--- just as increasing knowledge--- can sometimes lead to new and different conclusions....]

Precision

Readability, Repeatability, Knowledge

- 20 July 2016 19:00:00 Munich
- 20 July 2016 19:00 Munich
- 20 July 2016 7 PM Munich
- 20 July 2016 Munich
- 20 July 2016
- Late July 2016
- July 2016
- 2nd Quarter 2016
- Summer 2016
- 2016
- 2010s (decade)
- 21th century

Granularity is NOT driven by some integer representation—bits.

Aprox. Vs Uncertain Dates

They are not the same.

- There are questionable dates--- “not sure”.
- And there are approx. ones..

When did the Egyptian pharaoh Tutankhamun live?

Following, the Torah there are a number of calculations for the date of the Great Flood. Any of these are not approximate but questionable. Chabad, for example, put the date of start of the Great Flood at 1656 since creation and, in their model, 2105 BCE. Others put it at 5000 BCE, 2304 BCE and a number of other dates.. Each calculated with a clear and well developed methodology.. Going from Creation forward.. from Exodus backwards.. etc. etc.. These days--- save the +-1 year--- are questionable but not approximate. Did the Great Flood even occur?

25th of December 3 BCE is generally given as the birthday for a Yehoshua ben Yosef. That date is questionable. Its the accepted date but its also generally accepted to be wrong. Among the followers of Christian faiths the year 2 BCE, 3 BCE, 4 BCE and 6 BCE and even 1 CE have their adherents and alongside 25 Dec. the 6th of April is a popular date. It is also not even accepted if such a person ever existed-- whence can have no date of birth. Some consider the figure to be a myth while others see multiple Yehoshuas within a number of Messianic sects.

So how did we get here?

ISO 8601:2004 was not expressive enough

- A lot of date metadata was not „clean“
- **Rebecca Guenther** of the Library of Congress (world‘s leading authority on digital preservation metadata) asked Ray Dennenburg to help develop a new date language.
- I had developed my own date superset for IB.
- I knew Ray from Z39.50, SRU and other standard efforts.
- Ray, myself and a number of interested geeks joined the show and EDTF was born.
- Widely accepted by the digital preservation community it was adopted by a number of other standards especially RDA (Resource Description and Access) metadata standard for cultural heritage (adopted by, among others, the German library).

- EDTF responded to a need for a date/time string to be more expressive than ISO 8601:2004.
- Defined features to be supported in a date/time string, features considered useful for a wide variety of applications.

EDTF was originally defined as a **Profile of** / **extension to** ISO 8601:2004

Profile of / extension to ISO 8601

- **Profile**
  - 8601 describes a number of date/time features. Some are redundant and/or not very useful; to reduce the scope for error and the complexity of software, it seems worthwhile to restrict the supported formats to a smaller set.

- **Extension**
  - On the other hand, there are a number of date and time format conventions in common use that are not included in ISO 8601; it seemed worthwhile to normalize these.

EDTF → ISO-8601:2019

Predecessors:

- ISO 2014:1976 (all-numeric dates)
- ISO 2015:1976 (week numbering)
- ISO 2711:1973 (ordinal date numbering)
- ISO 3307:1975 (representations of time of the day)
- ISO 4031:1978 (time differentials)

These standards were all superseded by the first ISO 8601, ISO 8601:1988.

And subsequently ISO 8601 has been updated multiple times:

- ISO 8601:1988
- ISO 8601:1988/Cor 1:1991
- ISO 8601:2000
- ISO 8601:2004

EDTF Three Levels: 0, 1, and 2

- Level 0: A profile of 8601
- Level 1: simple extensions
- Level 2: more complex extensions

Level 0 (profile of 8601)

- Date
- Date and Time
- Interval

Date

- Any of the following:
  - year, month, and day (e.g. 2001-02-03)
  - year and month (e.g. 2008-12)
  - year (e.g. 2008)
- extended form only (hyphens)
- Four-digit year only

Date and Time

- One of the following three forms
  - ❑ 2001-02-03T09:30:01
  - ❑ 2004-01-01T10:10:10Z
  - ❑ 2004-01-01T10:10:10+05:00

- I.e. date ‘T’ time   followed by either nothing, “Z” or zone offset

- Zone-offset: '+' or '-' followed by a 2-digit hour, followed optionally by a colon and the 2-digit minutes.

Interval

Normal Interval Rules

- date / date
  - No times
  - No durations

- Examples:
  - 1964/2008
  - 2004-06/2006-08
  - 2004-02-01/2005-02-08

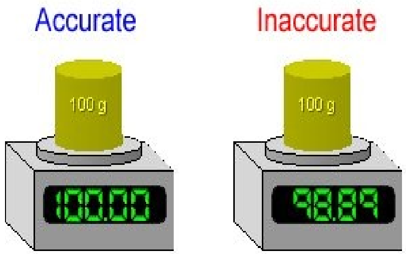
- Start and end date may have different precisions:
  - 2004-02-01/2005-02
  - 2004-02-01/2005
  - 2005/2006-02

Level 1 (Simple Extensions)

- Uncertain
- Approximate
- Unspecified
- Extended Interval
- Year Exceeding Four Digits
- Season

Uncertain

- 1984?
- 2004-06?
- 2004-06-11?



Approximate

- 1984~
- 2004-06~
- 2004-06-11~

Unspecified

- Replacement character(s) rather than appended (X originally u).
  - 199X
    some unspecified year in the 1990s. (year precision)
  - 19XX
    some unspecified year in the 1900s. (year precision)
  - 1999-XX
    some month in 1999 (month precision)
  - 1999-01-XX
    some day in January 1999 (day precision)
  - 1999-XX-XX
    some day in 1999 (day precision)

Extended Interval

- unknown/2006
  beginning unknown

- 2004-06-01/unknown
  end unknown

- 2004-01-01/open
  no end date

- 1984~/2004-06
  beginning approximately 1984 and ending June 2004

- 1984-06-02?/2004-08-08~
  beginning uncertain, end approximate

Year Exceeding Four Digits

- ‘Y’ at the beginning of the string means it is a year, with no limit on number of digits.

  - y170000002
    the year 170000002

  - y-170000002
    the year -170000002

Season

- 2001-21
  Spring, 2001

- 2003-22
  Summer, 2003

- 2000-23
  Autumn, 2000

- 2010-24
  Winter, 2010

Level 2: more complex extensions

- 2004?-06-11
  uncertain year; month, day known

- 2004-06~-11
  year and month approximate; day known

- 2004-06-(11)~
  day approximate; year, month known

- 2004-(06)~-11
  month approximate, year and day known

- 15uu-12-25
  December 25 sometime during the 1500s

Year Exceeding Four Digits (L2): Exponential Form

- y17e7
  the year 170000000 (contrast with y170000000 of level 1)

- y-17e7
  the year -170000000

- y17101e4p3
  Some year between 171000000 and 171999999, estimated to be 171010000 ('p3' indicates a precision of 3 significant digits.)

ISO 8601:2019

Of particular significance the new edition (which is effectively a new standard) contains:

— a concept of precision

— uncertain or approximate dates, or dates with portions unspecified;

— extended time intervals;

— divisions of a year;

— sets and choices of calendar dates;

— grouped time scale units;

— repeat rules for recurring time intervals; and

— date and time arithmetic.

It is now in two parts. Part 1 specifies basic rules. It functionally more or less corresponds the the previous standard but has been heavily rewritten (>80%). Part 2 (wholly new) specifies extensions, including community profiles to specify conformance levels for use of which rules.

ISO 8601 “Part 2”

- ISO / TC 154 WG 5
  - Convened Spring 2015
  - Ray Denenburg of the Library of Congress is currently representing the EDTF community in the WG.
  - ISO elected to adopt EDTF features and extend 8601 to two parts
    - ISO 8601 Part 1 & Part 2 Update agreed (8 June 2016) for CD Ballot.
  - Preliminary drafts were made available to EDTF community - http://www.loc.gov/standards/datetime/iso-tc154-wg5_n0038_iso_wd_8601-1_2016-02-16.pdf

    - http://www.loc.gov/standards/datetime/iso-tc154-wg5_n0039_iso_wd_8601-2_2016-02-16.pdf

ISO 8601:2019 Part 2 finally published in 2019 (and since 2020 being ratified by many national bodies)

In ISO-8601:2019 extensions there are two levels defined: 1 and 2 (in contrast to EDTF's 3 given that the standard has been advanced to support EDTF features).

Example of Level-1 vs Level-2:

Unspecified (X) – in earlier versions of EDTF the character 'u'
The value is not stated. The point in time may be unspecified because it did not occur yet, because it is classified, unknown or for any other reason.

Level 1: 'X' may be substituted for the right-most digits, e.g. day, day and month, …
Level 2: 'X' may be used as a replacement for any character in the string.

Seasons (Year Divisions)

Level 1:
The values 21, 22, 23, 24 may be used to signify ' Spring', 'Summer', 'Autumn', 'Winter', respectively. Example: 2020-21 for Spring 2020

Level 2:
These values are from the perspective of the “reporter” and don’t respect their location. But since Summer is the Southern Hemisphere maps to Winter in the Northern we have in Level-2 additionally:

25-28 = Spring - Northern Hemisphere, Summer- Northern Hemisphere, Autumn - Northern Hemisphere, Winter - Northern Hemisphere
29-32 = Spring – Southern Hemisphere, Summer– Southern Hemisphere, Autumn – Southern Hemisphere, Winter - Southern Hemisphere
33-36 = Quarter 1, Quarter 2, Quarter 3, Quarter 4 (3 months each)
37-39 = Quadrimester 1, Quadrimester 2, Quadrimester 3 (4 months each)

Decade (available in Level-2 only)
Format: YYY

Example: 196 (1960-1969)
Example: 196~ (approximate decade)

Note: the expression 196 expresses what is commonly called the “1960s”.

It does not refer to an event within the interval defined from 1 Jan 1960 to 31 Dec 1969 but refers to a singular event viewed with decade precision.

RFC-3339

The RFC is the IETF standard for “Date and Time on the Internet: Timestamps”.

Is a profile (subset) of ISO 8601:1988. Its only “extension” is that it permits the space separator for time, e.g. 2021-10-09 01:23:00Z

Many products that claim ISO-8601 (typically without an edition listed) don’t but are more or less based on this RFC.

RFC-2822/RFC-822 Dates

The RFC is the IETF standard for “Internet Message Format”. It is another beast.. and well.. not really the domain of this talk.. But our library….

W3 XML Schema Built-in types?

xs:date and xs:datetime are inadequate even for ISO 8601 level 0 core.

2001-02-03 is a valid xs:date value, but not 20010203

(This is a choice that W3C made when defining xs:date: the hyphenated form was chosen and the non-hyphenated form excluded.)

2001-02-03T09:30:01 is xs:dateTime

(W3c demands colons in between hours, seconds, etc.)

xs:date does not allow just the year and month (without the day), or just the year, or a date range (ISO 8601 supports all of these).

- SPARQL
- SQL
- CQL

Date Schema?

- <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified"attributeFormDefault="unqualified">
- <!--
- ************************ edtfSimpleType edtfSimpleType is the union of three simple types - xsDate, xs:dateTime, and edtfRegularExpressions. ("union" means that any string conforming to any one of the types in the union will validate.) xs:date and xs:dateTime are built-in W3Cschema types. edtfRegularExpressions is a set of four regular expressions which are described below. So any string that conforms to one of the two built-in types or any of the four regular expressions will validate.
- -->
- <xs:simpleType name="edtfSimpleType">
- <xs:union memberTypes="xs:date xs:dateTime edtfRegularExpressions"/>
- </xs:simpleType>
- <!-- ******** edft -->
- <xs:simpleType name="edtfRegularExpressions">
- <xs:restriction base="xs:string">
- <!--
- The following pattern is for year (yyyy) or year-month (yyyy-mm) The last or last two digits of year may be '?' meaning "one year in that range but not sure which year", for example 19?? means some year from 1990 to 1999. Similarly month may be '??' so that 2004-?? "means some month in 2004". And the entire string may end with '?' or '~' for "uncertain" or "approximate". Hyphen must separate year and month.
- -->
- <xs:pattern value="\d{2}(\d{2}|\?\?|\d(\d|\?))(-(\d{2}|\?\?))?~?\??"/>
- <!--
- The following pattern is for yearMonthDay - yyyymmdd, where 'dd' may be '??' so '200412??' means "some day during the month of 12/2004". The whole string may be followed by '?' or '~' to mean "questionable" or "approximate". Hypens are not allowed for this pattern.
- -->
- <xs:pattern value="\d{6}(\d{2}|\?\?)~?\??"/>
- <!--
- The following pattern is for date and time with T separator:'yyyymmddThhmmss'. Hypens in date and colons in time not allowed for this pattern.
- -->
- <xs:pattern value="\d{8}T\d{6}"/>
- <!--
- The following pattern is for a date range. in years: 'yyyy/yyyy'; or year/month: yyyy-mm/yyyy-mm. Beginning or end of range value may be 'unknown'. End of range value may be 'open'. Hypens mandatory when month is present.
- -->
- <xs:pattern value="((\d{4}(-\d{2})?)|unknown)/((\d{4}(-\d{2})?)|unknown|open)"/>
- <!-- -->
- </xs:restriction>
- </xs:simpleType>
- </xs:schema>

EDTF is still freely available:

- EDTF specification at http://www.loc.gov/standards/datetime/
- Listserv / Archive http://www.loc.gov/standards/datetime/listserv.html

And, of course, ISO 8601:2019 is availble from ISO (€€€)

The repo https://github.com/re-Isearch/date:

contains a fork of the date parser from the re-Isearch project (an open source search engine) designed/intended for use in other projects. Among the changes it uses std:string instead of the re-Isearch's own STRING class.

The above repository also contains the latest generic version of this presentation