

Multilingual timespan expressions

Ceri Binding, Hypermedia Research Group, University of South Wales
<http://hypermedia.research.southwales.ac.uk/>
ceri.binding@southwales.ac.uk



The problem

- Timespans are often expressed textually
- fuzzy, may be less committed than absolute dates
- comparison and interoperability of records is difficult, even where numbers are present

Timespan

Circa Fifth Century BC

Early 2nd Century

XV secolo d.C.

intorno a VI sec. d.C.

pymthegfed ganrif

Canol y15fed ganrif

1450-1460

1485-86

C. 1485

1540±9

AD400+

400 AD

Circa 1860s

intorno al decennio 1910

1930au

5th – 6th century AD

VIII-VII secolo a.C.

5ed 6ed ganrif

July 1855

Luglio 1855

Gorffennaf 1855

Georgian

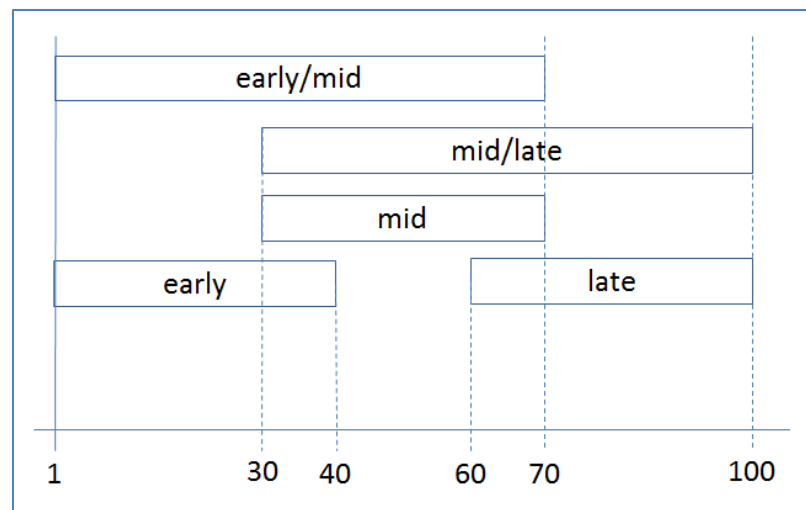
Victorian

Multilingual timespan expressions

English	Dutch	French	Spanish	Italian	German	Swedish	Welsh
11 th century	11de eeuw	11ème siècle	Siglo 11	11 ° secolo Secolo XI	11. Jahrhundert	11-talet	11eg ganrif
Early 11 th century	Begin 11e eeuw	début du 11ème siècle	principios del siglo XI	inizio dell'XI secolo	frühes 11. Jahrhundert	tidigt 11-talet	Dechrau'r 11eg ganrif
Mid 11 th century	Midden 11e eeuw	milieu du 11ème siècle	mediados del siglo XI	metà del 11 ° secolo	Mitte des 11. Jahrhunderts	mitten 11- talet	Canol yr 11eg ganrif
Late 11 th century	Eind 11e eeuw	fin du 11ème siècle	finales del siglo XI	fine dell'XI secolo	Ende des 11. Jahrhunderts	slutet av 1100- talet	Diwedd yr 11eg ganrif
Eleventh	Elfde	Onzième	Undécimo	undicesimo	Elfte	Elfte	unfed ar ddeg
First quarter	eerste kwartier	premier quart	primer cuarto	primo quarto	Erstes Viertel	första kvarten	chwarter cyntaf
First half	eerste helft	première moitié	primera mitad	prima metà	erste Hälfte	första halvan	hanner cyntaf
AD	Na Christus	après JC	dC	dC	n. Chr	e.Kr.	AD
BC	v.Chr	avant JC	aC	aC	Vor Christus	f.Kr.	CC (Cyn Crist)

Other associated issues

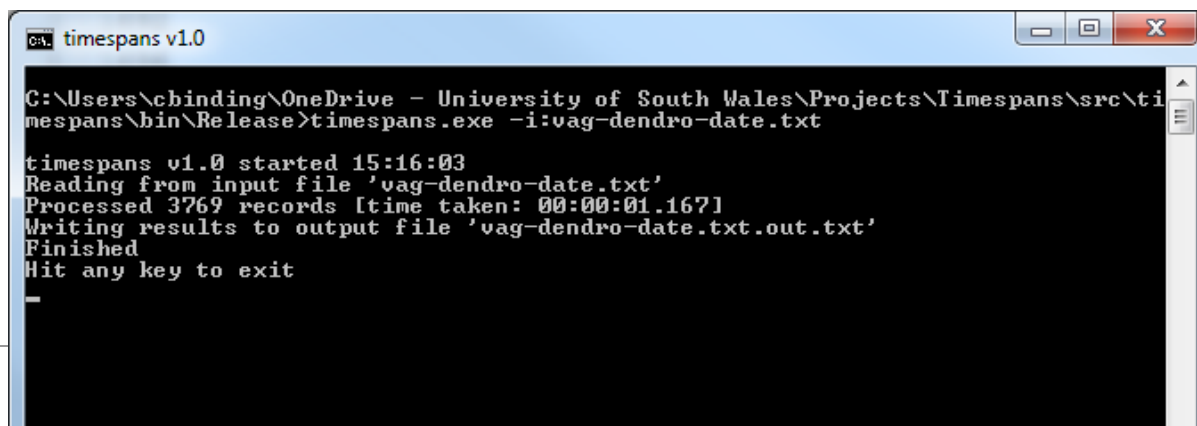
- Consistency of punctuation, grammar, whitespace etc.
- Prefixes: early / mid / late / around / approx. / circa / after etc. - what do they mean?
- Suffixes: AD / BC / CE / BCE / BP (and multilingual equivalents)
 - BP relative to 1950
- Complexity and variety of expressions:
 - *maybe(prefix), year/year span, maybe tolerance, maybe(suffix)*
 - *Named periods / ordinal named or numbered centuries / rulers*



Timespans application

- Open source console application (on GitHub)
 - <https://github.com/cbinding/timespans>
- Using regular expression matches and lookups
 - Small modular expressions are used to build more complex expressions
 - Language specific regular expression patterns for multilingual matching
 - Named period lookups on authority resources (e.g. heritage-standards periods list)

Regular expression example	Input	Output
<code>^(?<yearMin>\d+)(?:\s?-\s?)(?<yearMax>\d+)\$</code>	"1450-1485"	yearMin=1450, yearMax=1485
<code>^(?<decade>\d+0)\'?s\$</code>	"1850's"	decade=1850
<code>^(?:C(?:irca .)\s*)?(?<year>\d+)\$</code>	"C. 1485"	year=1485



```
timespans v1.0
C:\Users\cbinding\OneDrive - University of South Wales\Projects\Timespans\src\timespans\bin\Release>timespans.exe -i:vag-dendro-date.txt

timespans v1.0 started 15:16:03
Reading from input file 'vag-dendro-date.txt'
Processed 3769 records [time taken: 00:00:01.167]
Writing results to output file 'vag-dendro-date.txt.out.txt'
Finished
Hit any key to exit
```

Results for specific test cases

Type	Language	Input	Output	
			Start	End
Ordinal named or numbered century	English	Circa Fifth Century BC	-500	-401
		Early 2nd Century	101	140
	Italian	XV secolo d.C.	1401	1500
		intorno a VI sec. d.C.	501	600
	Welsh	pymthegfed ganrif	1401	1500
		Canol y 15fed ganrif	1430	1470
Year span	Any	1450-1460	1450	1460
		1485-86	1485	1486
Single year (with tolerance)	Any	C. 1485	1485	1485
		1540±9	1531	1549
		AD400+	400	400
		400 AD	400	400
Decade	English	Circa 1860s	1860	1869
	Italian	intorno al decennio 1910	1910	1919
	Welsh	1930au	1930	1939
Century span	English	5th – 6th century AD	401	600
	Italian	VIII-VII secolo a.C.	-800	-601
	Welsh	5ed 6ed ganrif	401	600
Month and year	English	July 1855	1855	1855
	Italian	Luglio 1855	1855	1855
	Welsh	Gorffennaf 1855	1855	1855
Season and year	English	Summer 1855	1855	1855
	Italian	Estate 1855	1855	1855
	Welsh	Haf 1855	1855	1855
Named periods (from lookup)	English	Georgian	1714	1837
		Victorian	1837	1901

Informal results from real datasets

Dataset	Correct	Incorrect	No Match	Total	Notes
VAG Cruck Database [Cruckfull2014.primary_date] https://doi.org/10.5284/1031497	87 (100%)	0	0	87	Used for testing during development - hence good results
VAG Dendrochronology Database [Ind_full2016_7.date] https://doi.org/10.5284/1043296	3730 (98.97%)	0	39	3769 (non-blank)	Non-matches were on additional punctuation or text (+ a grid ref!)
Staffordshire Hoard [appendix1a.Dates] https://doi.org/10.5284/1041576	46 (97.87%)	0	1	47 (non-blank)	Non-match on “c 10000 BC- c AD1799” Reason – 2 nd ‘circa’ not included in the matching pattern
Cottam B https://doi.org/10.5284/1034389	556 (51.34%)	0	527	1083	Non-matches were on unknown named periods: “Anglian”, “Anglo-Scandinavian”
Goldsmith St Exeter [Context dating PDF] https://doi.org/10.5284/1035174	116 (83.45%)	0	23	139	Non-matches were on additional punctuation or text e.g. “?12th century”, “after 1300”

Results – testing on multilingual data

- Italian data
 - European Open Science Cloud (EOSC) pilot project - Daniel Williams, USW researcher
 - Named Entity Recognition applied to annotation of archaeological reports
 - One strand - extracted list of Italian timespans from free text documents
 - Results were used for testing the application – see example results extract opposite
- Needs additional test data for English - and other languages
 - Dutch, French, Spanish, Italian, German, Swedish and Welsh supported – but to varying degrees due to lack of test data

secolo d.C.	0	0
settembre	0	0
tra IV e III secolo a.C.	-400	-201
tra IV e V sec. d.C.	301	500
tra V e IV secolo a.C.	-500	-301
tra VII e VI a.C.	-700	-501
tra VIII e IX secolo	701	900
tra XII e IX secolo a.C.	-1200	-801
tra lo 141 ed lo 91 a.C.	-141	-91
tra lo 1790 ed lo 1797	1790	1797
tra lo 1861 ed lo 1887	1861	1887
tra lo 1895 ed lo 1896	1895	1896
tra lo 1985 e lo 1990	1985	1990
tra lo 217 e lo 206 a.C.	-217	-206
tra lo 241 e lo 238 a.C.	-241	-238
tra lo I e lo VI sec.	1	600
tra lo III e lo IV sec. d.C.	201	400
tra lo III ed lo II	101	300
tra lo III ed lo II secolo a.C.	-300	-101
tra lo ed lo III a.C. V d.C.	0	0

ScAPA: Scottish Archaeological Periods & Ages

Partitioning the past: A periodisation for Scotland

Ceri Binding, Hypermedia Research Group, University of South Wales

<http://hypermedia.research.southwales.ac.uk/>

ceri.binding@southwales.ac.uk

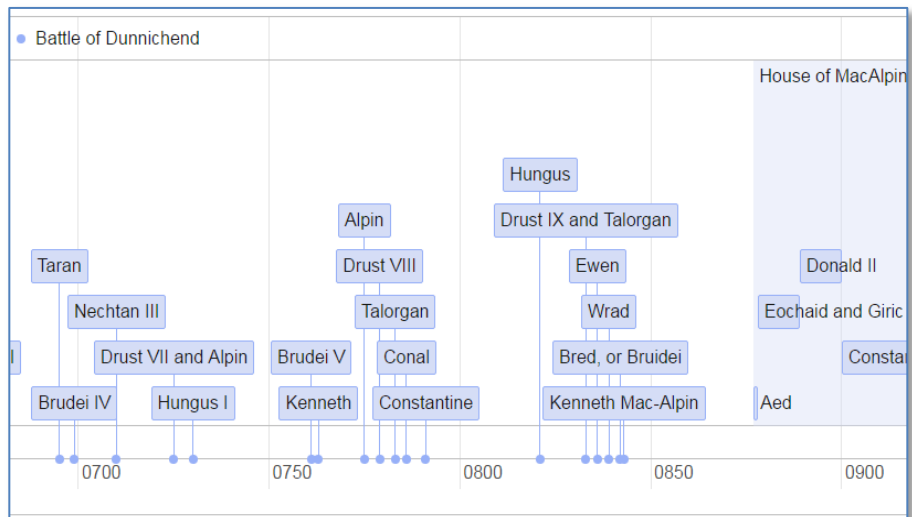
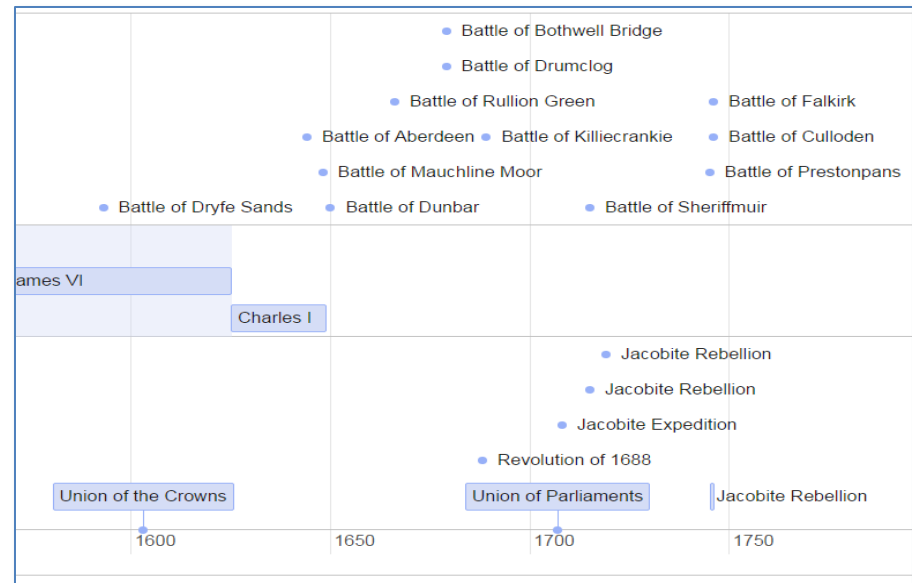
Chronology and Periodisation

■ Periodization

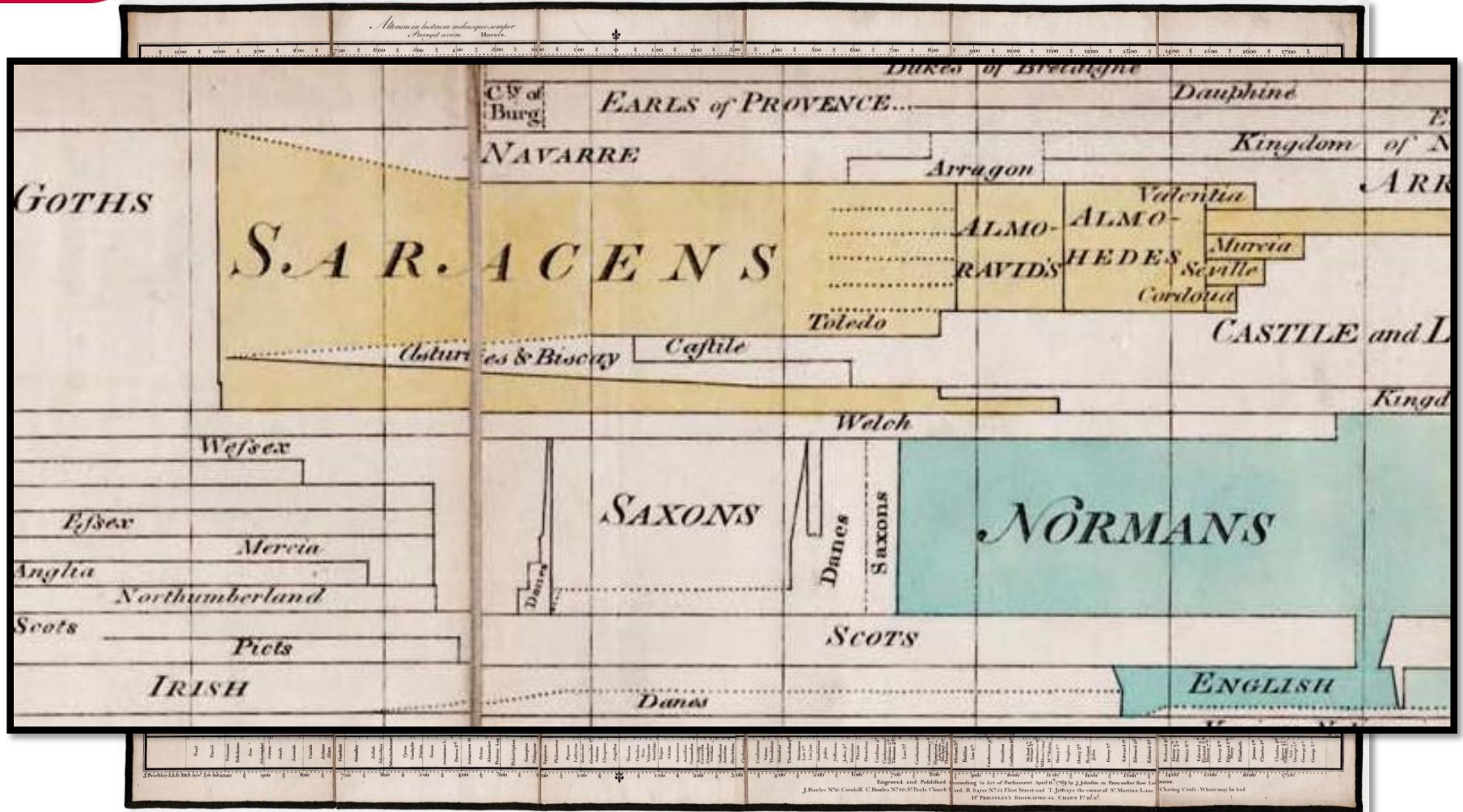
- Subdivision of time into discrete named blocks representing periods
- Utilises knowledge of past events (periods delimited by events)
- Produces an extensible framework providing relative context for other events

■ Chronology

- The science of locating historical periods and events relative to each other.
- Accurate dates may be unknown, but can arrange periods in relative chronological order
- Can be graphically illustrated using a timeline



Regional differences



A new Chart of history – Joseph Priestley (1769)

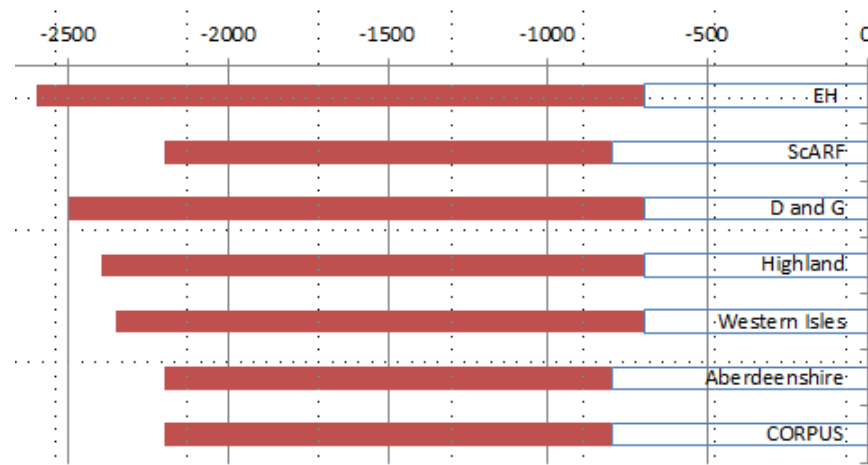
https://en.wikipedia.org/wiki/A_New_Chart_of_History

Q. When/where was the Bronze Age?

Identifier	Label(s)	Source	Spatial Coverage	Start	End
ark:/99152/p0gigrs6qb2	Bronze Age	Portable Antiquities Scheme	United Kingdom	2350 BC	801 BC
ark:/99152/p0kh9ds7q8m	Bronze Age	HeritageData.org (Historic England Periods List)	United Kingdom	2600 BC	700 BC
ark:/99152/p0zj6g8rbvk	Bronze Age	ARENA Portal	United Kingdom	2500 BC	700 BC

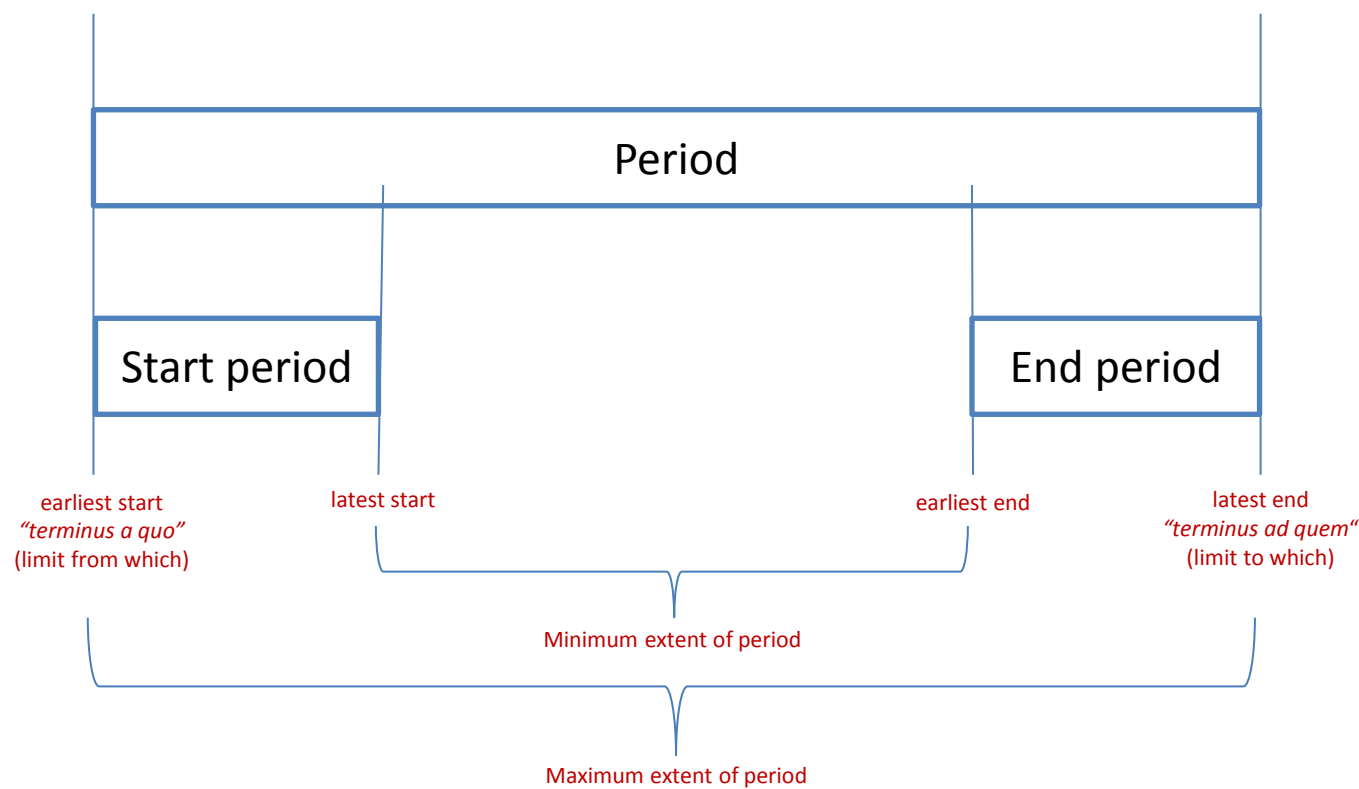
“Refers to the period and culture associated with the second age in the Three Age system developed by Christian Jürgensen Thomsen in 1836. It is characterized by the widespread use of bronze, which is an alloy of copper and tin, lead, antimony, or arsenic, in the manufacture of tools and weapons. It developed at different times in different parts of the world, from around 3,500 BCE in Greece and China to around 1,400 BCE in several areas of Europe. There was not a Bronze Age in the Americas, since the local Stone Age cultures were introduced directly to Iron Age technologies by European explorers.” <http://vocab.getty.edu/aat/300019275>

- Answer - It depends (of course)...
- Problem - we use the term in indexing
- When? Multiple opinions
- Where? Regional differences
- Can be visualised on parallel timelines



Modelling of time periods

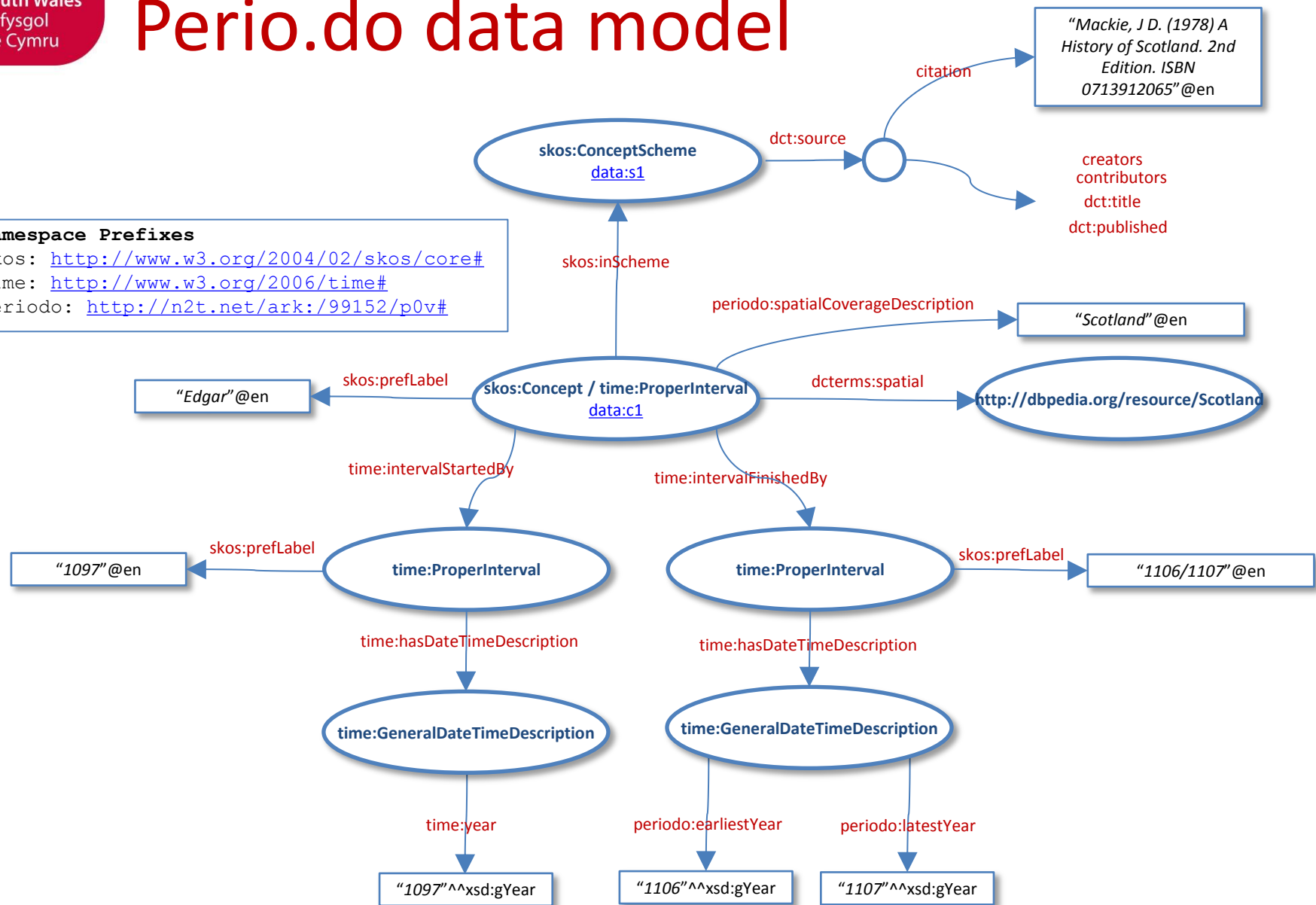
- Start/end periods can represent uncertainty, or a gradual increase/decline
- Periods also have spatial / cultural context – they are not just timespans



Perio.do data model

Namespace Prefixes

skos: <http://www.w3.org/2004/02/skos/core#>
time: <http://www.w3.org/2006/time#>
perio: <http://n2t.net/ark:/99152/p0v#>



Periodo

A gazetteer of period definitions for linking and visualizing data.

Periods

Viewing 26 - 50 of 327

Show periods at a time.

[Previous](#)

[1](#) [2](#) [3](#) [...](#) [13](#) [14](#)

[Next](#)

Label	▲ Earliest start	Latest stop
Early Mesolithic	-10000	-7001
Mesolithic	-10000	-4001
Mesolithic	-10000	-4001
Early Prehistoric	-10000	-1001
Early Mesolithic	-10000	-7001
Early Mesolithic	-10000	-7001
Early Mesolithic	-10000	-7001
Mesolithic	-10000	-4001
Prehistoric	-10000	300

Mesolithic

In collection: [Shetland Periods List. 2018.](#)

Original label	Mesolithic (eng-latn)
Spatial coverage des...	Shetlands
Spatial coverage	Scotland
Start	-10000 (ISO value: -10000)
Stop	-4501 (ISO value: -4501)

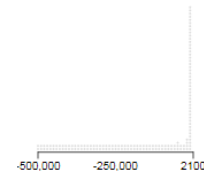
Mesolithic

In collection: [Highland periods list. 2018.](#)

Original label	Mesolithic (eng-latn)
Alternate labels	ME (eng-latn)
Spatial coverage des...	Highlands
Spatial coverage	Scotland
Start	-8000 (ISO value: -8000)

Filters

Time range



☒ Hide outliers?

Text

Match string

Source

[Reset](#)

- 65 [PKHER PeriodLUT. 2017.](#)
- 40 [Historic Environment Scotland periods list. 2018.](#)
- 35 [Leanne and Angela. DG HER listing. September 2017.](#)
- 33 [Aberdeenshire lists. 2017.](#)
- 32 [East and Midlothian. 2017.](#)
- 29 [Shetland Periods List. 2018.](#)
- 27 [Mackie, J D. A History of Scotland. 1978.](#)
- 26 [Stirling and Clackmannanshire HER periods list. 2018.](#)
- 26 [Highland periods list. 2018.](#)

Language

[Reset](#)

- 327 [English](#)

Spatial coverage

[Reset](#)

- 67 [Scotland](#)
- 65 [Perth and Kinross](#)
- 35 [Dumfries and Galloway](#)
- 33 [Aberdeenshire](#)
- 32 [East and Mid Lothian](#)
- 29 [Shetlands](#)
- 26 [Stirling and Clackmannanshire](#)

Existing Linked Data period lists

Heritage Data

Linked Data Vocabularies for Cultural Heritage

[Scheme List](#)
[Concept Search](#)
[SPARQL Query](#)
[About The Project](#)

http://purl.org/heritagedata/schemes/eh_period/concepts/MD ([QR Code](#))

Property	Value
rdf:type	skos:Concept
cc:license	http://creativecommons.org/licenses/by/3.0
cc:attributionURL	http://www.english-heritage.org.uk
cc:attributionName	English Heritage
skos:topConceptOf	Historic England Periods
skos:inScheme	Historic England Periods
skos:prefLabel	MEDIEVAL
skos:scopeNote	The Medieval period or Middle Ages begins with the Norman invasion and ends with the dissolution of the monasteries.
skos:notation	MD
dct:publisher	http://www.english-heritage.org.uk
dct:identifier	http://purl.org/heritagedata/schemes/eh_period/concepts/MD
dct:issued	2013-12-06T16:44:37

RDF downloads ([N-Triples](#) [Turtle](#) [JSON](#) [XML](#))

Historic England

Heritage Data

Linked Data Vocabularies for Cultural Heritage

[Scheme List](#)
[Concept Search](#)
[SPARQL Query](#)
[About The Project](#)

<http://purl.org/heritagedata/schemes/11/concepts/505128> ([QR Code](#))

Property	Value
rdf:type	skos:Concept
cc:license	http://reference.data.gov.uk/id/open-government-licence
cc:attributionURL	http://www.rcahmw.gov.uk
cc:attributionName	RCAHMW
skos:inScheme	PERIOD (WALES)
skos:prefLabel	CANOLOESOL [cy]
skos:prefLabel	MEDIEVAL
skos:scopeNote	From the establishment of Norman control over large parts of Wales in 1086 to the Act of Union between England and Wales in 1536.
dct:publisher	http://www.rcahmw.gov.uk
dct:created	05/08/2010
dct:modified	05/08/2010
dct:identifier	http://purl.org/heritagedata/schemes/11/concepts/505128
dct:issued	2017-01-20T16:02:49
dct:issued	2015-09-04T15:27:52

RDF downloads ([N-Triples](#) [Turtle](#) [JSON](#) [XML](#))

RCAHMW

ScAPA data

- Compiled to spreadsheet then converted to SKOS RDF using SKOS Play <http://labs.sparna.fr/skos-play/>

	A	B	C	D	E	F	G	H	I
7	PREFIX	aat	http://vocab.getty.edu/aat/						
8	PREFIX	tgn	http://vocab.getty.edu/tgn/						
9	dct:identifier	http://purl.org/heritagedata/schemes/scapa							
10	dct:title@en	ScAPA : Scottish Archaeological Periods & Ages							
11	skos:prefLabel@en	ScAPA : Scottish Archaeological Periods & Ages							
12	dct:description@en	List of archaeological periods & ages for cultural heritage use in Scotland							
13	dct:coverage	tgn:7002444							
14	dct:publisher	https://www.historicenvironment.scot/							
15	dct:contributor	orgs:HES							
16	dct:created^^xsd:date	17/01/2018							
17	dct:modified^^xsd:date	17/01/2018							
18	dct:issued^^xsd:date	17/01/2018							
19	dct:license	lic:open-government-licence							
20	dct:spatial	tgn:7002444							
21									
22	NOTE: composite data compiled from separate spreadsheets received from Scottish HERs. Mappings to external concepts by C Binding								
23									
24									
25	URI	dct:identifier	:notation@en(separator=	dct:license	skos:prefLabel@gd	os:	skos:prefLabel@en	skos:altLabel@en(separator=";	skos:broader
26	scapa:PA	scapa:PA	PA	lic:open-government-licence	(todo)		Palaeolithic	Paleolithic	scapa:EPR
27	scapa:LPA	scapa:LPA	LPA	lic:open-government-licence	(todo)		Lower Palaeolithic	Lower Paleolithic	scapa:PA
28	scapa:MPA	scapa:MPA	MPA	lic:open-government-licence	(todo)		Middle Palaeolithic	Middle Paleolithic	scapa:PA
29	scapa:UPA	scapa:UPA	UPA	lic:open-government-licence	(todo)		Upper Palaeolithic	Upper Paleolithic	scapa:PA
30	scapa:PR	scapa:PR	PR	lic:open-government-licence	Ro-eachdraidheil		Prehistoric	Prehistory; Prehistoric uncertain	scapa:period
31	scapa:EPR	scapa:EPR	EPR	lic:open-government-licence	Tràth Ro-eachdraidheil		Early Prehistoric	Early Prehistory	scapa:PR
32	scapa:LAT	scapa:LAT	LAT	lic:open-government-licence	(todo)		Late Prehistoric	Late Prehistory	scapa:PR
33	scapa:ME	scapa:ME	ME	lic:open-government-licence	(todo)		Mesolithic		scapa:EPR
34	scapa:EME	scapa:EME	EME	lic:open-government-licence	(todo)		Early Mesolithic		scapa:ME
35	scapa:LME	scapa:LME	LME	lic:open-government-licence	(todo)		Late Mesolithic		scapa:ME

SKOS Play – verification

SKOS testing tool

[Home](#) [en](#) ▼

Ambiguous Notation

References : FAIL (1)

Cyclic Hierarchical Relations :
OK

Disconnected Concept Clusters
: OK

Disjoint Labels Violation : OK

Empty Labels : OK

Hierarchical Redundancy : OK

HTTP URI Scheme Violation :
OK

Incomplete Language Coverage
: OK

Inconsistent Preferred Labels :
OK

Missing Labels : FAIL (36)

Mapping Relations Misuse :
FAIL (46)

No Common Languages : OK

Orphan Concepts : OK

Omitted or Invalid Language
Tags : OK

Verified **scapa-composite-data.rdf** at 01/02/2018 17:07:38

NEW TEST

25 rules checked in 1.80 seconds - 8 rules failed.

177 Concepts in 0 Collections and 2 Concept Schemes.

[Print this report](#)

Rule	Description	Result
anr - Ambiguous Notation References	Finds concepts with multiple or identical notations within the same concept scheme.	FAIL (1)
<i>Resource: http://purl.org/heritagedata/schemes/scapa/concepts/AN, Notations: AN, ANG</i>		
chr - Cyclic Hierarchical Relations	Finds concepts that are hierarchically related to each other.	OK
dcc - Disconnected Concept Clusters	Finds sets of concepts that are isolated from the rest of the vocabulary.	OK
dlv - Disjoint Labels Violation	Finds resources with identical entries for different label types.	OK

- <http://labs.sparna.fr/skos-play/>

SKOS Play – alphabetical listing

- Generated interactive listing
- Includes hierarchical tree
- Includes multilingual translations

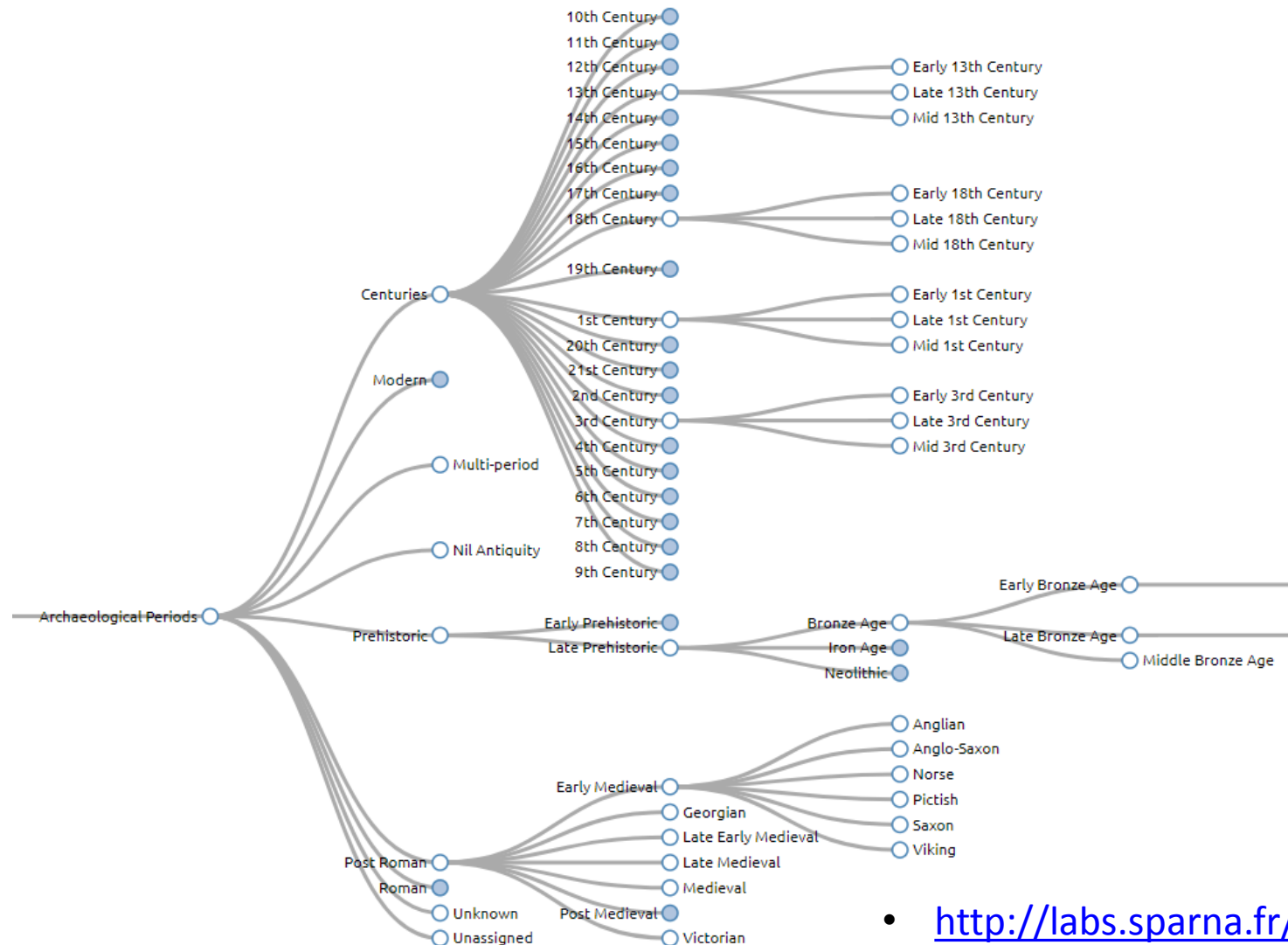
S

- **Saxon**
GD : (todo)
CC : 5A
SN : (todo)
BT : **Early Medieval**
TT : **Archaeological Periods**
- ~~Second Century~~
USE : **2nd Century**
- **Second World War**
GD : An Dara Cogadh
CC : WW2
note : CFB: Scope note from EH list - may need to be revised
UF : *2nd World War*
UF : *WW2*
UF : *WWII*
UF : *World War 2*
UF : *World War II*
SN : Used to record buildings, defensive monuments and sites dating to, and associated with, the Second World War. For other types of building, such as houses, built during this period use MID 20TH CENTURY.
BT : **Modern**
TT : **Archaeological Periods**
- ~~Seventeenth Century~~
USE : **17th Century**
- ~~Seventh Century~~
USE : **7th Century**
- **Severan**
GD : (todo)
CC : ROSE
note : CFB: temp notation
SN : (todo)
BT : **Roman**
TT : **Archaeological Periods**
- ~~Sixteenth Century~~

1 2 3 4 5 6 7 8 9 A B C D E F G H I
L M N P R S T U V W

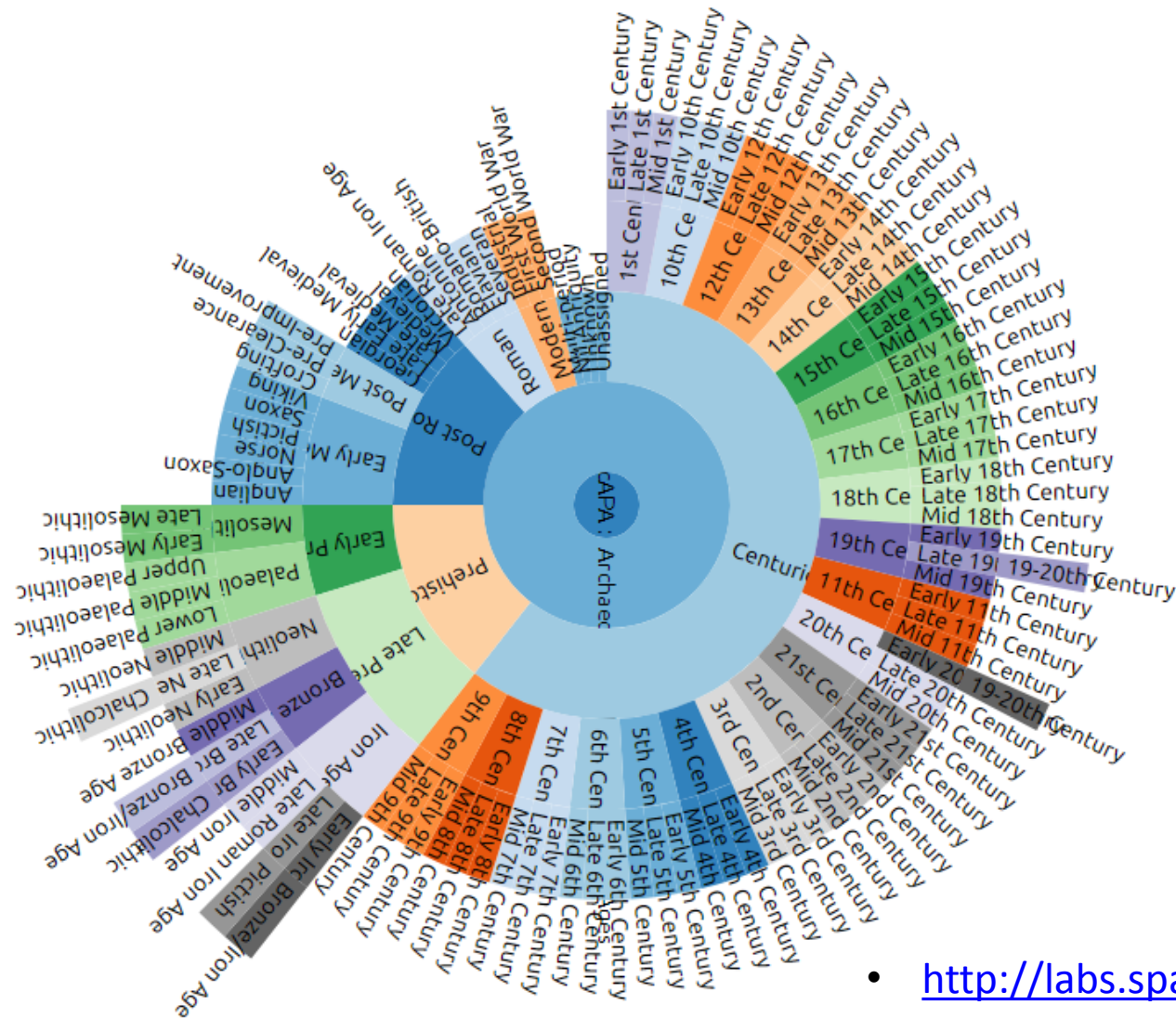
- <http://labs.sparna.fr/skos-play/>

SKOS Play – tree visualisation



- <http://labs.sparna.fr/skos-play/>

SKOS Play – circle visualisation



- <http://labs.sparna.fr/skos-play/>

SKOS Play – drop down list control

vi|

Flavian
★ *(todo)*

Victorian
★ *Dating to the reign of Queen Victoria*

Viking
The period of initial settlement following on from a period of Viking (raiding) activities and initial contact.
[Source: ScARF Medieval document]
★ *(todo)*

- <http://labs.sparna.fr/skos-play/>

Mappings to external resources

- Mappings to
Getty AAT
concepts and HE
Periods list

source concept			target concept
Palaeolithic	Art and Architecture Thesaurus	close	Paleolithic
	Historic England Periods	close	PALAEOLITHIC
Pictish	Art and Architecture Thesaurus	close	Pictish
Post Medieval	Historic England Periods	close	POST MEDIEVAL
Prehistoric	Art and Architecture Thesaurus	close	prehistoric
	Historic England Periods	close	PREHISTORIC

Mappings - R

source concept			target concept
Roman	Art and Architecture Thesaurus	close	Imperial (Roman)
	Historic England	close	ROMAN

Mappings - 1 Mappings - 2 Mappings - 3 Mappings - 4 Mappings - 5 Mappings - 6
 Mappings - 7 Mappings - 8 Mappings - 9 Mappings - A Mappings - B Mappings - C
 Mappings - E Mappings - F Mappings - G Mappings - I Mappings - L Mappings - M
 Mappings - N Mappings - P Mappings - R Mappings - S Mappings - U Mappings - V

Next steps

- Update scope notes, data structures and Gaelic terms in accordance with feedback
- Once data is finalised, publish the periods on <http://www.heritagedata.org/>
- Submit a patch for inclusion of ScAPA data in Perio.do main 'canonical' dataset

ScAPA: Scottish Archaeological Periods & Ages

Partitioning the past: A periodisation for Scotland

Ceri Binding, Hypermedia Research Group, University of South Wales

<http://hypermedia.research.southwales.ac.uk/>

ceri.binding@southwales.ac.uk

