Format Name	Short Description	General Application area	Strengths of the format	Weaknesses of the format	Selected Software Packages which read or write	Supported by Treebase	Supported by OpenTree of Life	Supported by MorphoBank	compatible with ontologies
Newick	simple text string using embedded parentheses to indicate the topology of a phylogenetic tree	representing phylogenies as text	simple; widely-used	multiple incompatible extensions exist; reliable parsing is difficult; reticulation unsupported; metadata not extensible	Mesquite, BioPython, BioPerl, DendroPy, Arbor, Geiger, PAUP, MyBayes, FigTree	No, unless embedded in nexus	Yes, but requires additional curation by user	No	No
Nexus	A plain text format for trees and alignments; can also contain commands for tree inference software	phylogenetics-specific data exchange; a single file can contain a matching tree and character matrix	can hold tree and matrix together in a single file and can embed character and state information; can hold other arbitrary data such as notes and weighting schemes	no formal schema that can be validated; multiple incompatible extensions exist; reliable parsing is difficult	Mesquite, BEAST, NCL (Nexus Class Library), DendroPy, MacClade, Nexus Data Editor, PAUP	Yes for upload and download	Yes, but requires additional curation by user	Yes	No
NeXML	rich, extensible XML representation of trees and character data	phylogenetic trees, matrices, and associated data	Standard schema and validation tools exist; easier to parse because of XML; many elements explicitly linked to CDAO ontology for machine-readability	few phlogenetic inference programs read / write NeXML yet; files can become large if trees or number annotations large	DendroPy, BioPython, BioPerl, Mesquite	Yes, for export	Yes	planned in 2014	Yes, can link metadata elements to ontologies
PhyloXML	XML format for tree and arbitrary annotations	tree hierarchy; arbitrary data can be attached to any node	can hold character data, branch lengths, and geolocations where a taxon is observed, etc (anything, really) on any node of the tree	lack of widespread schema and validation tools; in danger of being extended in different, inconsistent ways by the community	BioPython, Forester, Archaeopteryx	No	Not yet	No	
CDAO	ontology specifically for phylogenetics that describes trees and associated data	supports inferencing queries on trees & matrices	supports complex analytical queries over the source data	currently hard to get data into this specific format; supported by only a few tools	BioPython through CDAO input/output	Yes, via annotations in NeXML	Yes, via annotations in NeXML	No	Is an ontology
APE/ R data objects	phylotrees for R analysis	seminal data model for trees used by R-based analysis libraries	compact representation; native format for R computation	internal represntation within APE library (in R language)	APE, Geiger			No	
FASTA	genetic sequence fragments	sequence alignment			Genbank, SAM, ClustalW		Only as supplemental file	No	
BEAST	XML format with custom extensions for tree and assocated additional data	tree construction	advanced tree inferencing	custom Newick extension incompatible with many packages	Beast, FigTree, BEAUTi	No	Not yet	No	

Hennig / TNT	A text format to represent genetic sequences, tree hierarchy, character matrix	phylogenetics-specific data exchange; single-file support for comparative methods	can hold tree and matrix together in a single file and can embed character and state information; can hold other arbitrary data such as notes and weighting schemes	no validator, parsing can be difficult	TNT	No	No	Yes	
NexSON? Karen should we include this?	NeXML serialized as JSON	exchange of NeXML data using JSON (Javascript object notation)	a more compact representation of NeXML	see NeXML					
Phylip	format for multiple sequence alignment	represent individual sequences and their result after alignment, including gaps	human readable, well-known	limited in data that can be represented	RaXML, Phylip	No	As supplemental files only	No	
BayesTraits							No	No	
SDD	XML for the exchange of characters, states, descriptions of taxa and specimens, and authored polytomous keys	exchange of descriptive data	published schema allows for validation; computer readable	descriptive data only; not phylogenetic trees	Lucid, EDIT	No	No	Yes; for export on project data only currently	unknown
STK-XML	Paper describing new format in review with Biodiversity Data Journal. Relax NG schema available from checkout of code: http://bazaar.launchpad.net/~stk-developers/supertree-toolkit/stk/files/head:/schema/	phylogenetic trees and associated metadata (not inc. alignments/matrices). Designed with supetree-making in mind but otherwise still useful as a data exchange format	published schema allows for validation; computer readable	Not implemented by any other tools AFAIK	STK & STK 2	No	No	No	
XTG XML (just mentioning this because it exists & seems in scope)	Well documented format used by TreeGraph 2 http://treegraph.bioinfweb.info/Development/XTG	phylogenetic trees	published schema allows for validation; computer readable	Not implemented by any other tools AFAIK	TreeGraph 2	No	No	No	
REST interfaces	An online data resource that can be quieried directly. An alternative way to access source data without reading files	web-hosted database archives (e.g. Paleobiology Database, OpenTree, LifeMapper, Encylopedia of Life, etc.) use this technique to serve data to users	readily available to the whole community through the internet; huge archives are now available	PhyloWS standard has been developed, but not all data services are compliant yet. differently by each archive	A web browser or a scripting language (R, Python, Java,) is currently needed to access or write datasets	Yes, TreeBase offers a REST API to download archived content. PhyloWS compliant.	Yes, OpenTree offers a REST API to download archived content	Planned	No

OWL	knowledge exchange format for ontologies / linked datasets	generaql knowledge representation as ontology; not specific to phylogenetics	supports powerful inferencing	hard to adapt source data into correct ontological relationships	Protege and other ontology editing tools		N/A	No	
Shapefile	Geospatial points, distributions, boundaries	Used to represent species observations, habitat extent	supported by most GIS systems		ArcInfo, QGIS, ESRI,LifeMapper		N/A	No	
TopoJSON, GeoJSON	emerging formats for geospatial data	geospatial, object positions	simple to parse; human-readable; supported by new web-based visualizations	verbose; inefficient for transfer of very large geo-data	D3; Vega (visualization), GDAL	No	N/A	No	No
KML (Keyhole Markup Language)	geospatial data (Google Earth format)	species occurrences and observation points	human readable, simple	little or no support for the format in phylogenetics software	Google Earth	No	N/A	No	No
JSON	Javascript Object Notation	similar to XML, general purpose, human-readable way to represent any structured data	increasingly being used in web-based software systems; human-readible and easily parsable	like CSV, JSON conveys arbitrary data; no semantics or detailed specfiication	many interactive web services included Paleobiology database, LifeMapper, OpenTreeOfLife	No	Yes, internally. Can be communicated through REST interfaces	Used internally	
CSV (comma separated values)	context-free format used for character matrices and other supporting data	used recently in R-based analysis, visualization, and other packages	easy to use and exchange	no specific semantics or meaning to the fields; open for mis-interpretation, cannot be direclty read into many widely used phylogenetics programs	supported for import/export by phylogenetics packages in R	Specimen or gene fragment metadata can be uploaded and downloaded in tab-separated text (closely allied with CSV)	N/A	Yes (for upload of taxonomy, not matrices)	

Format Name	Additional Comments	Literature citation
Newick	original and still most widely used phylogenetic tree format; conveys tree topology unambiguously, but no standard for annotations or additional attributes besides node names and branch lengths	
Nexus	tree hierarchy is generally represented using an embedded Newick string; cross reference tables used for taxon names, morphological values	Maddison, D.R., D.L. Swofford, and W.P. Maddison, NEXUS: an extendible file format for systematic information. Systematic Biology, 1997. 46: p. 590-621.
NeXML		Vos, R.A., et al., NeXML: rich, extensible, and verifiable representation of comparative data and metadata. Systematic Biology, 2012. 61(4): p. 675-89.
PhyloXML		Han, M.V. and C.M. Zmasek, phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics, 2009. 10: p. 356.
CDAO		Prosdocimi, F., et al., Initial Implementation of a Comparative Data Analysis Ontology. Evolutionary Bioinformatics, 2009. 5: p. 47-66.
APE/ R data objects		
FASTA		
BEAST		Drummond, A.J. and A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol, 2007. 7: p. 214.

Hennig / TNT		
NexSON? Karen should we include this?		
Phylip	Some versions place very restrictive limits on taxon label length	
BayesTraits		
SDD		(none: see http://www.keytonature.eu/wiki/Structured_Descriptive_Data_XML_FAQ)
STK-XML	Trees from thousands of previously published arthropod publications are due to be released in this format as part of the ongoing Arthropod Supertree BBSRC project http://www.bbsrc.ac.uk/pa/grants/AwardDetails.aspx?FundingReference=BB%2fK006754%2f1	
XTG XML (just mentioning this because it exists & seems in scope)	Just mentioning it for thoroughness-sake	
REST interfaces	BioCatalogue.org is a site that indexes many available REST services	

OWL	Ontologies offer a powerful means to represent knowledge concepts and relationships; they also require careful control of definitions and vocabulary. It is generally considered difficult to convert the knowledge a phylogenetic dataset contains into an ontology	
Shapefile	GIS = Geospatial Information Systems; commercial and free products for maps and geospatial information	
TopoJSON, GeoJSON	emerging alternative to historic GIS formats (i.e. shapefiles) now used by new web-based visualization frameworks.	
KML (Keyhole Markup Language)		
JSON	REST = Representational State Transfer; REST is a style of handshaking / communication between a client program and a server publishing its data using the web	
CSV (comma separated values)	universal exchange format devoid of helpful semantic meanings associated with data elements.	