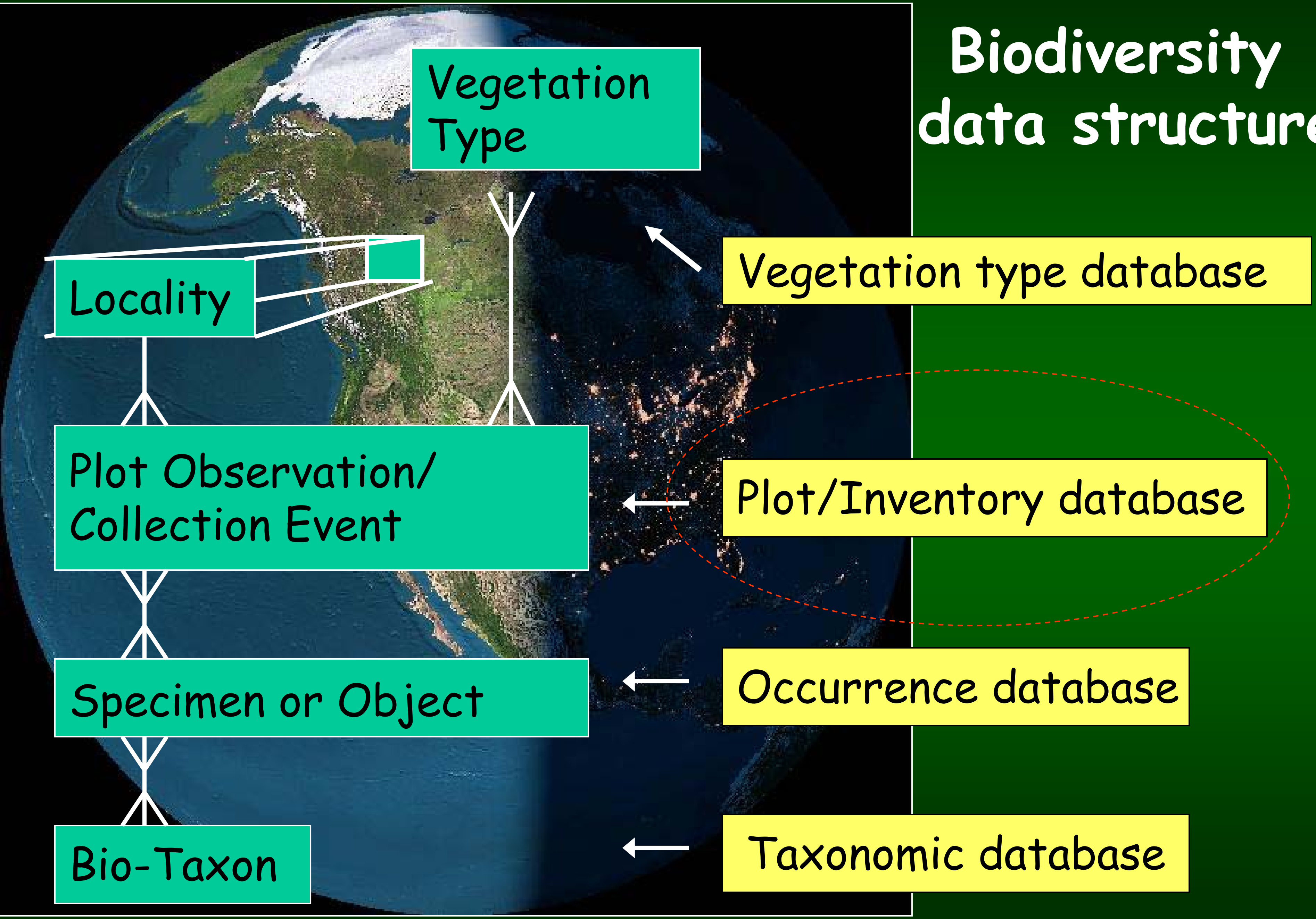
The VegBank Data Model

Biodiversity data structure



Vegetation Type

Locality

Plot Observation/ Collection Event

Specimen or Object

Bio-Taxon

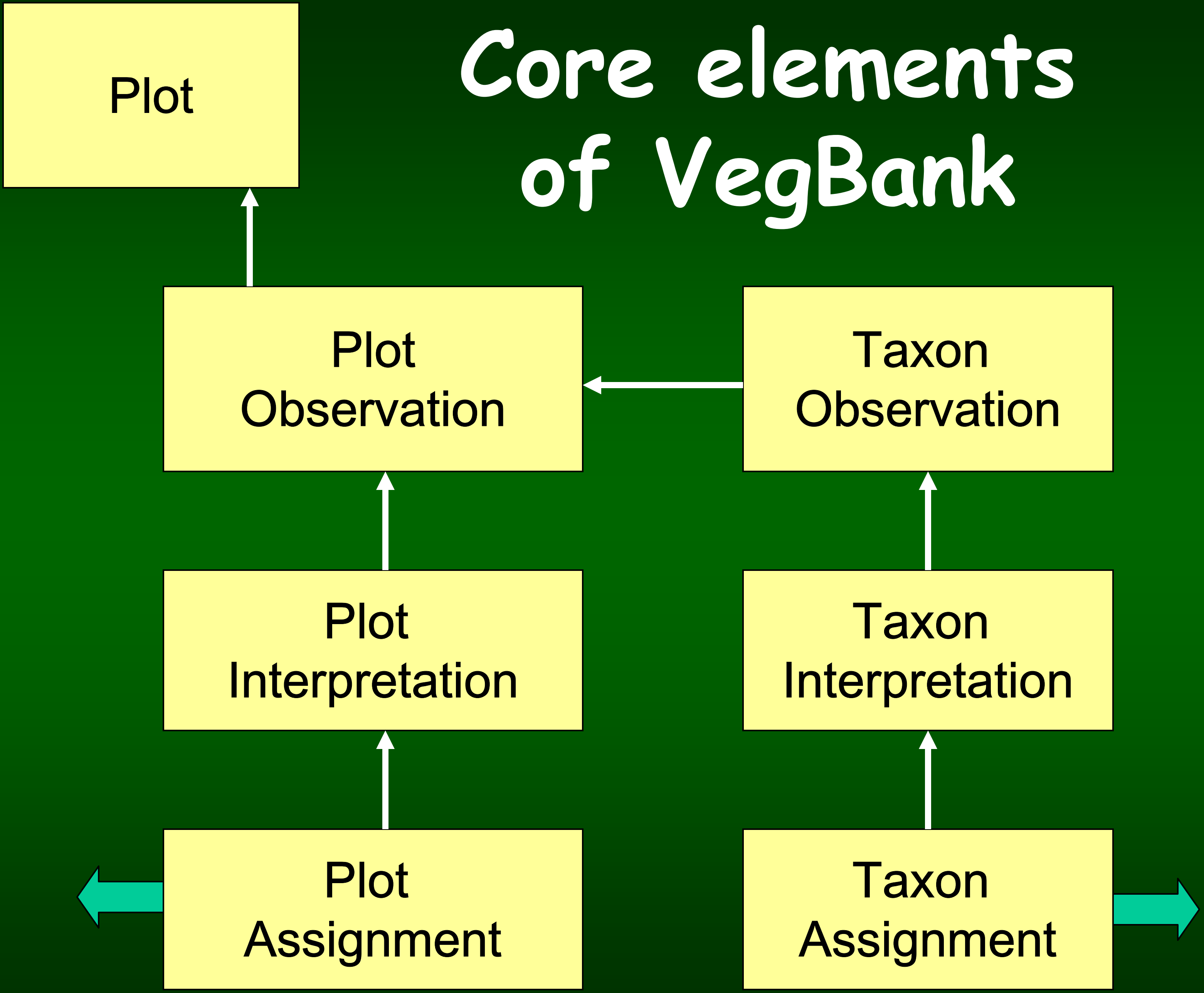
Vegetation type database

Plot/Inventory database

Occurrence database

Taxonomic database

Core elements of VegBank



VegBank consists of three integrated databases

1. The Plot Database
2. The Plant Database
3. The Community Database

Taxonomic database challenge: *Standardizing organisms and communities*

The problem:

Integration of data potentially representing *different* times, places, investigators and taxonomic standards.

The traditional solution:

A standard checklists of organisms.

Standard checklists for Taxa

Representative examples for higher plants in *North America / US*

USDA Plants	http://plants.usda.gov
ITIS	http://www.itis.usda.gov
NatureServe	http://www.natureserve.org
BONAP	http://www.bonap.org/
Flora North America	http://hua.huh.harvard.edu/FNA/

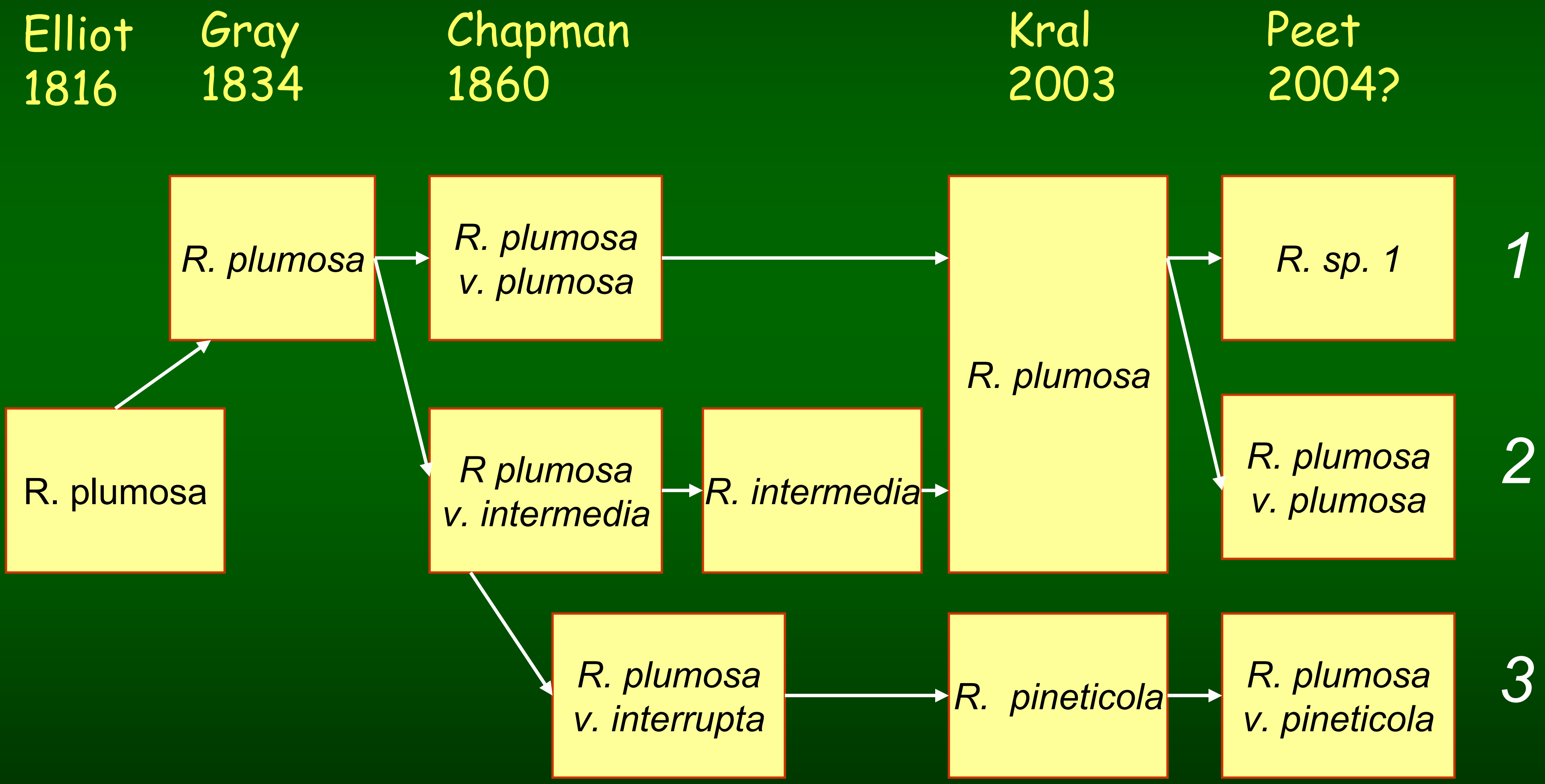
These are intended to be checklists wherein the taxa recognized perfectly partition all plants. The lists can be dynamic.

Most taxon checklists fail to allow effective dataset integration

The reasons include:

- The user cannot reconstruct the database as viewed at an arbitrary time in the past,
- Taxonomic concepts are not defined (just lists),
- Multiple party perspectives on taxonomic concepts and names cannot be supported or reconciled.

Multiple concepts of *Rhynchospora plumosa s.l.*



Elliot 1816

Gray 1834

Chapman 1860

Kral 2003

Peet 2004?

R. plumosa

R. plumosa

R. plumosa v. plumosa

R plumosa v. intermedia

R. plumosa v. interrupta

R. intermedia

R. plumosa

R. pineticola

R. sp. 1

R. plumosa v. plumosa

R. plumosa v. pineticola

1

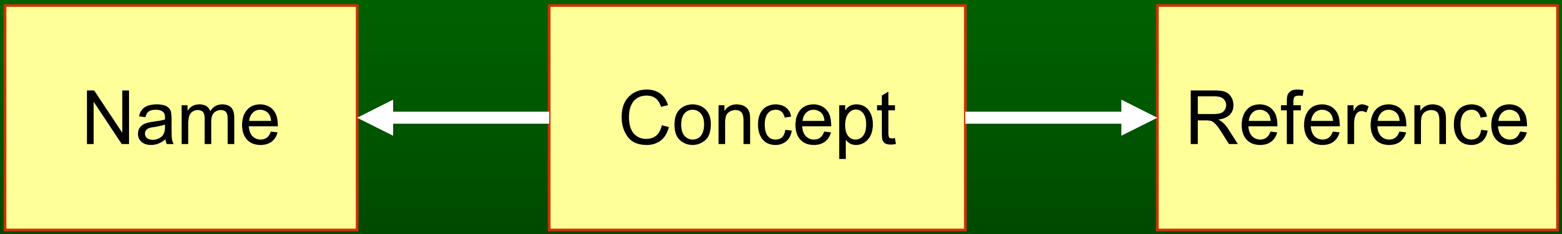
2

3

Taxonomic theory

A taxon concept represents a unique combination of a ***name*** and a ***reference***

"Taxon concept" roughly equivalent to "Potential taxon" & "assertion"

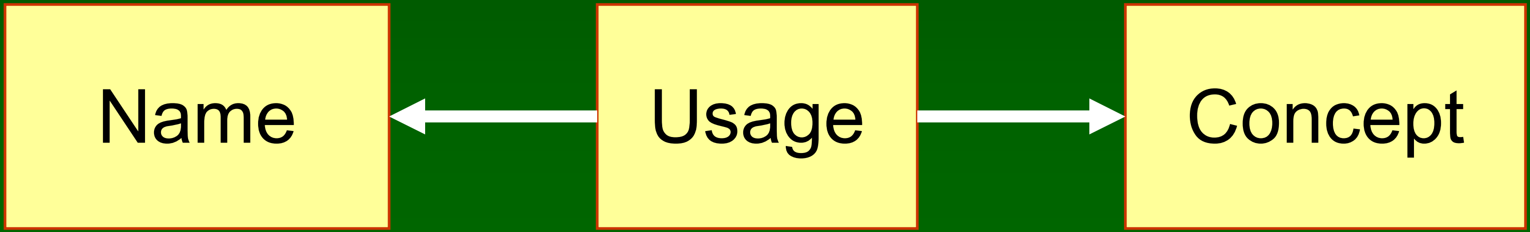


Name

Concept

Reference

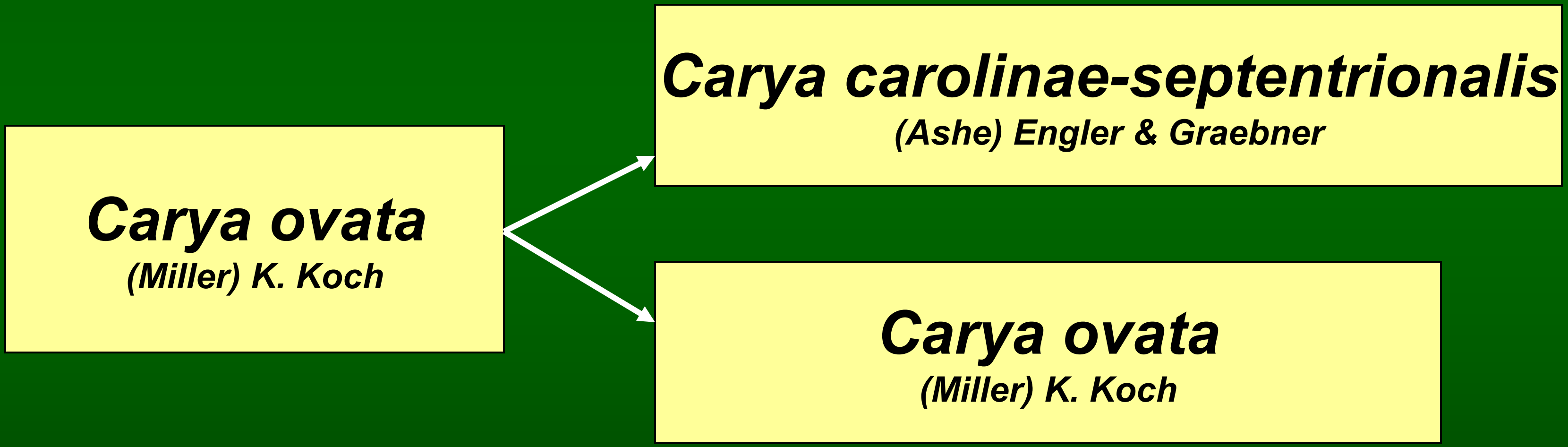
A *usage* represents an association of a *concept* with a *name.*



- Usage does not appear in the IOPI model, but instead is a special case of concept
- Usage can be used to apply multiple name systems to a concept
- Desirable for stability in recognized concepts

Three concepts of shagbark hickory

Splitting one species into two illustrates the ambiguity often associated with scientific names.



Carya ovata
(Miller) K. Koch

Carya carolinae-septentrionalis
(Ashe) Engler & Graebner

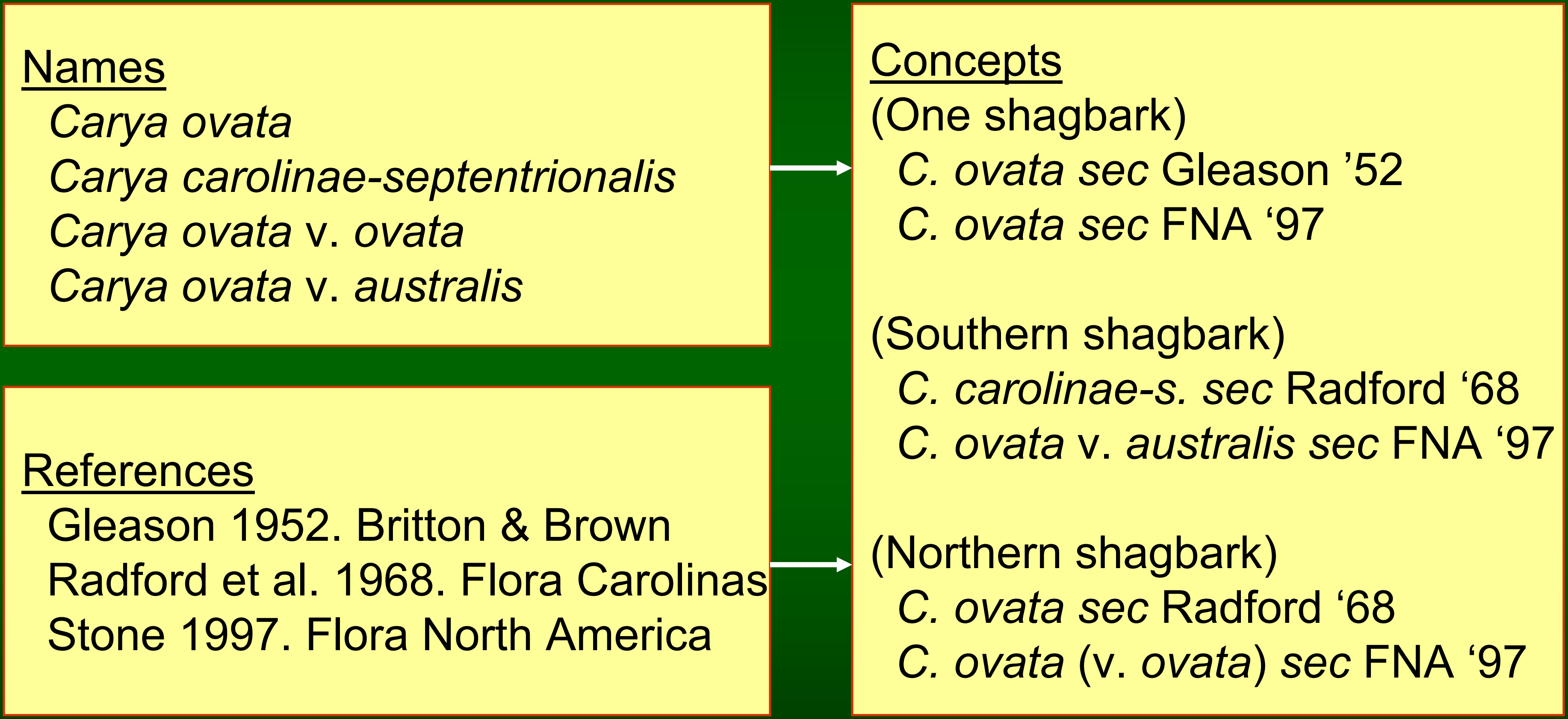
Carya ovata
(Miller) K. Koch

sec. Gleason 1952

sec. Radford et al. 1968

Six shagbark hickory concepts

Possible synonyms are listed together



Names

Carya ovata
Carya carolinae-septentrionalis
Carya ovata v. *ovata*
Carya ovata v. *australis*

References

Gleason 1952. Britton & Brown
Radford et al. 1968. Flora Carolinas
Stone 1997. Flora North America

Concepts

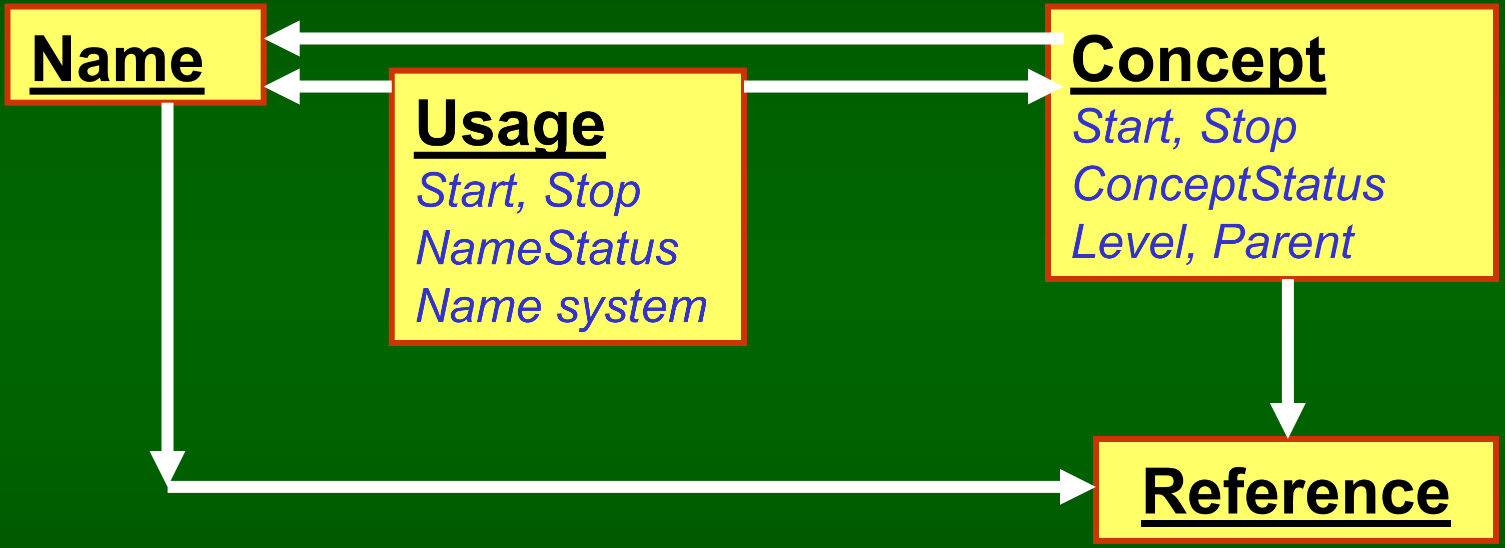
(One shagbark)
C. ovata sec Gleason ’52
C. ovata sec FNA ‘97

(Southern shagbark)
C. carolinae-s. sec Radford ‘68
C. ovata v. *australis sec* FNA ‘97

(Northern shagbark)
C. ovata sec Radford ‘68
C. ovata (v. *ovata*) *sec* FNA ‘97

Data relationships

VegBank taxonomic data model



Name

Usage
Start, Stop
NameStatus
Name system

Concept
Start, Stop
ConceptStatus
Level, Parent

Reference

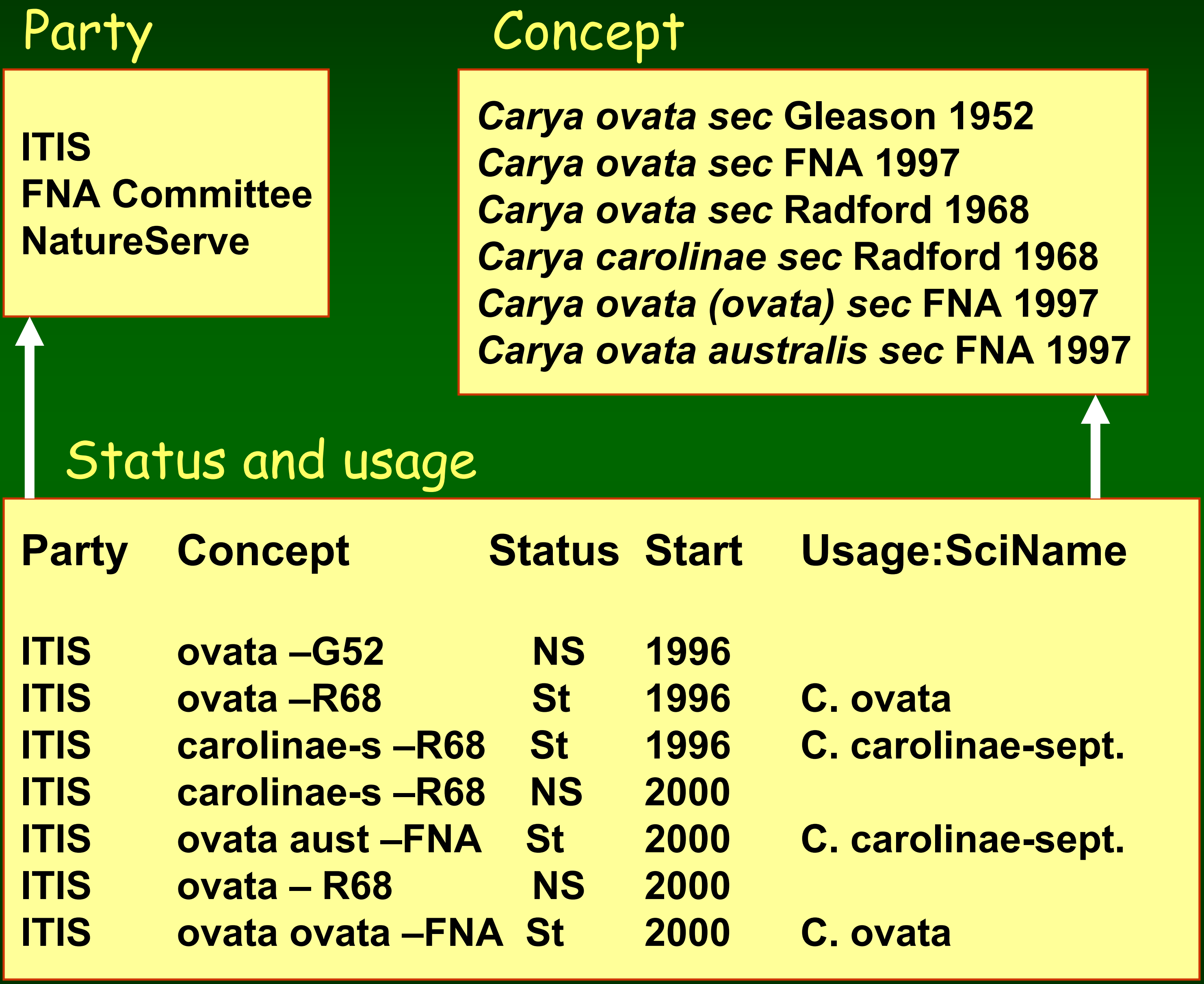
Single party, dynamic perspective

Party Perspective

The Party Perspective on a concept includes:

- Status – Standard, Nonstandard, Undetermined
- Correlation with other concepts – Equal, Greater, Lesser, Overlap, Undetermined.
- Lineage – Predecessor and Successor concepts.
- Start & Stop dates for tracking changes

Application of Party Perspective



Party

ITIS
FNA Committee
NatureServe

Concept

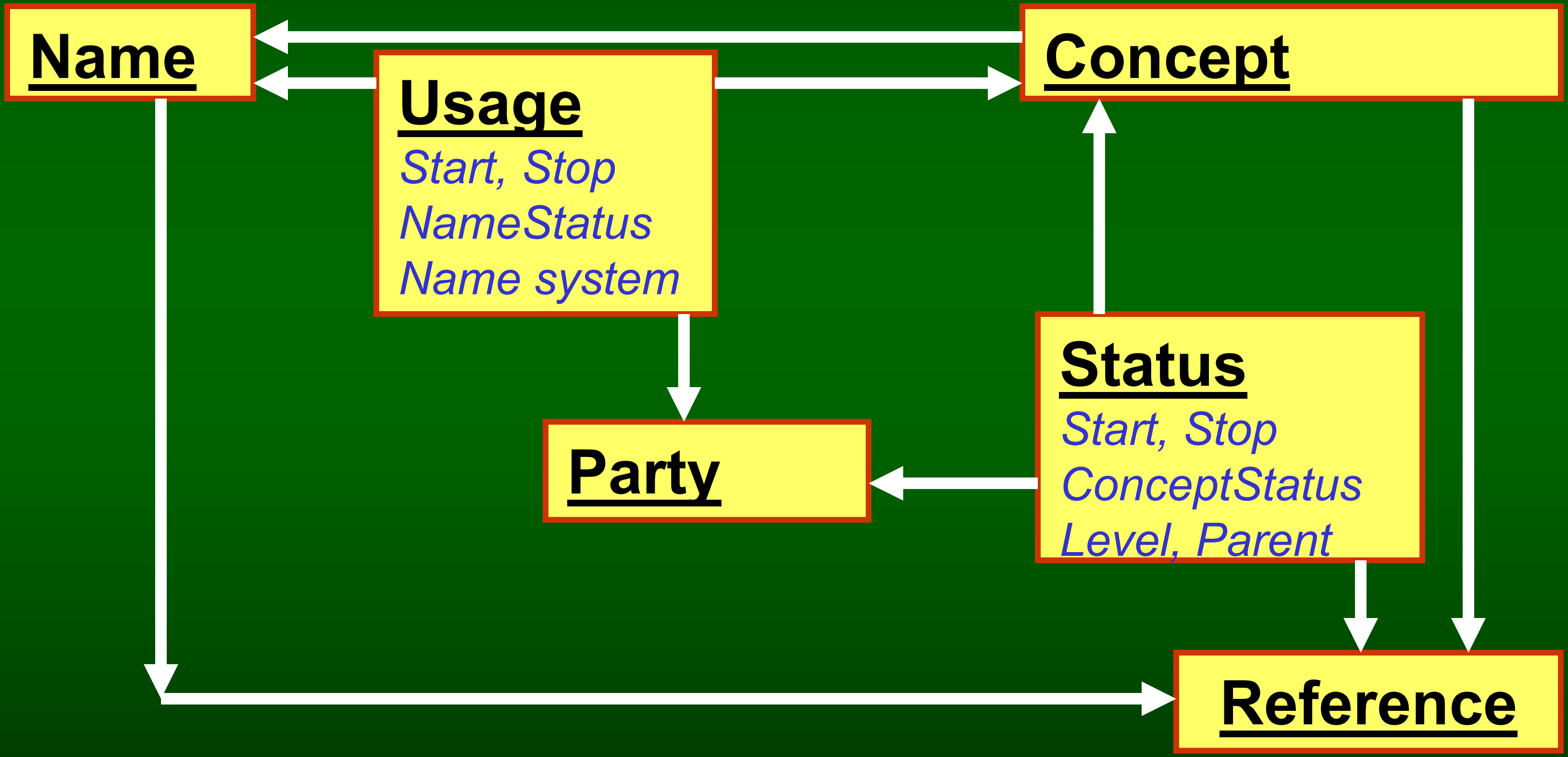
Carya ovata sec Gleason 1952
Carya ovata sec FNA 1997
Carya ovata sec Radford 1968
Carya carolinae sec Radford 1968
Carya ovata (ovata) sec FNA 1997
Carya ovata australis sec FNA 1997

Status and usage

Party	Concept	Status	Start	Usage:SciName
ITIS	ovata –G52	NS	1996	
ITIS	ovata –R68	St	1996	C. ovata
ITIS	carolinae-s –R68	St	1996	C. carolinae-sept.
ITIS	carolinae-s –R68	NS	2000	
ITIS	ovata aust –FNA	St	2000	C. carolinae-sept.
ITIS	ovata – R68	NS	2000	
ITIS	ovata ovata –FNA	St	2000	C. ovata

Data relationships

VegBank taxonomic data model



Name

Usage
Start, Stop
NameStatus
Name system

Concept

Status
Start, Stop
ConceptStatus
Level, Parent

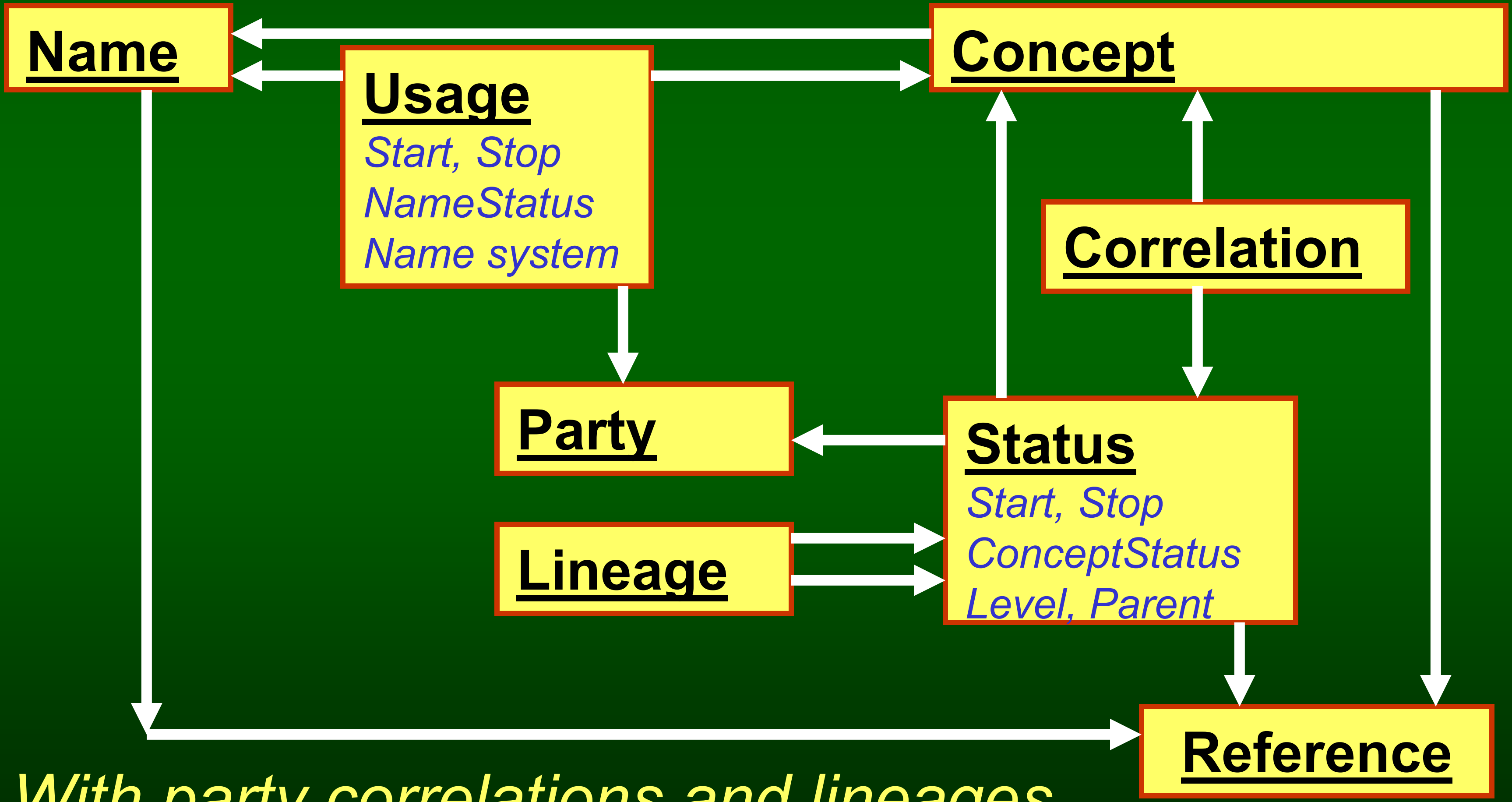
Party

Reference

Multiple parties, dynamic perspectives

Data relationships

VegBank taxonomic data model

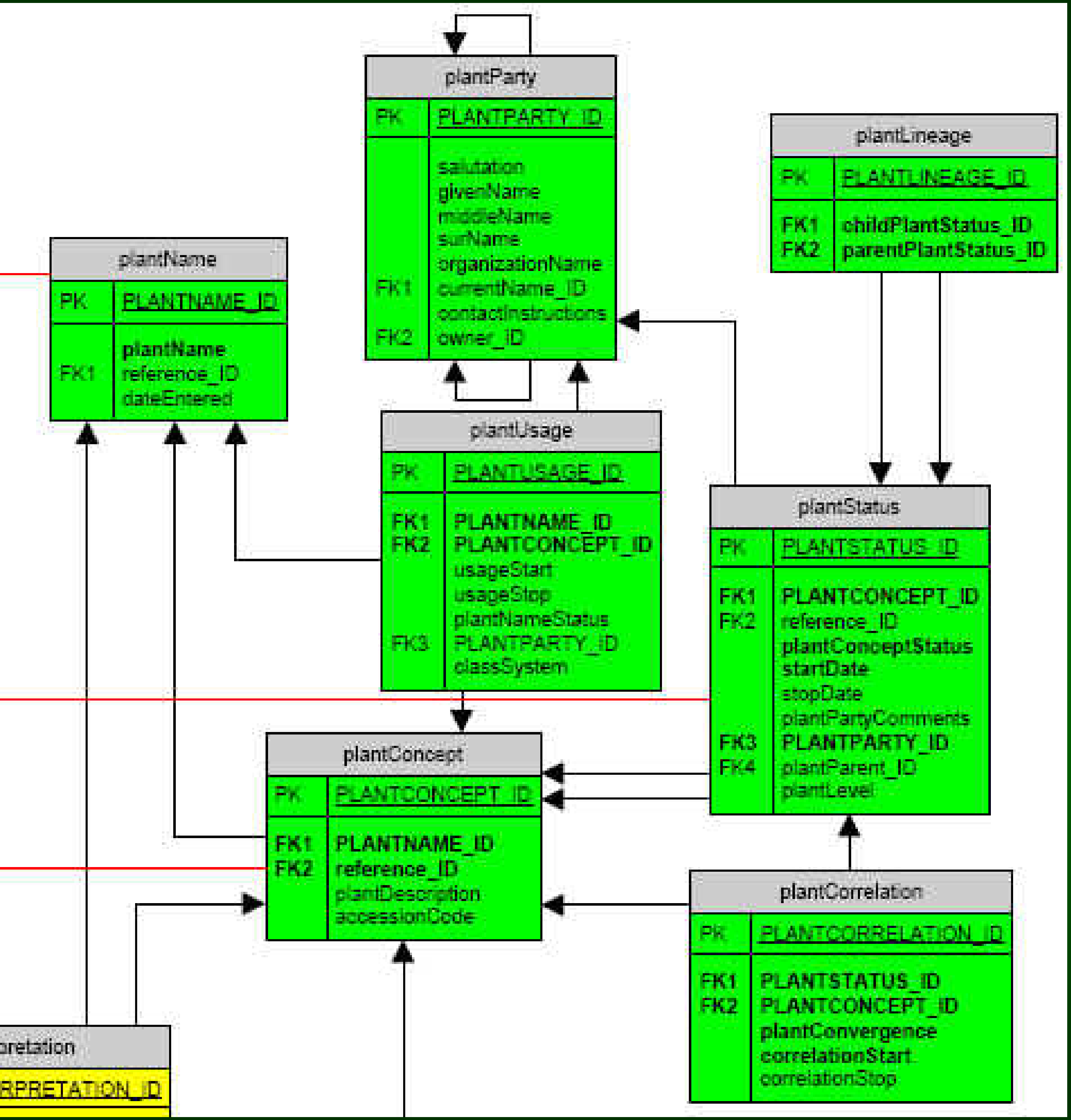


With party correlations and lineages

Intended functionality

- Organisms are labeled by reference to concept (name-reference combination),
- Party perspectives on concepts and names can be dynamic, but remain perfectly archived,
- User can select which party perspective to follow,
- Different names systems are supported,
- Enhanced stability in recognized concepts by separating name assignment and rank from concept.

Plant Taxa



plantParty	
PK	PLANTPARTY_ID
	salutation givenName middleName surName organizationName
FK1	currentName_ID
	contactInstructions
FK2	owner_ID

plantLineage	
PK	PLANTLINEAGE_ID
FK1	childPlantStatus_ID
FK2	parentPlantStatus_ID

plantName	
PK	PLANTNAME_ID
	plantName
FK1	reference_ID
	dateEntered

plantUsage	
PK	PLANTUSAGE_ID
FK1	PLANTNAME_ID
FK2	PLANTCONCEPT_ID
	usageStart usageStop plantNameStatus
FK3	PLANTPARTY_ID
	classSystem

plantStatus	
PK	PLANTSTATUS_ID
FK1	PLANTCONCEPT_ID
FK2	reference_ID
	plantConceptStatus startDate stopDate plantPartyComments
FK3	PLANTPARTY_ID
FK4	plantParent_ID
	plantLevel

plantConcept	
PK	PLANTCONCEPT_ID
FK1	PLANTNAME_ID
FK2	reference_ID
	plantDescription accessionCode

plantCorrelation	
PK	PLANTCORRELATION_ID
FK1	PLANTSTATUS_ID
FK2	PLANTCONCEPT_ID
	plantConvergence correlationStart correlationStop

- Name
- (Reference)
- Concept
- Status
- Correlation
- Lineage
- Usage
- Party

State of Taxon Concept Development

1. TDWG, IOPI, & SEEK
2. VegBank
3. Collaborators

- NatureServe Biotics4
- USDA PLANTS & ITIS

VegBank taxon data content

Prototype populated with USDA PLANTS lists and synonyms = weak concepts.

Contract with NatureServe and John Kartesz

- Develop reference-based concepts for 14000 by July 2004 of the ~32000 vascular plant taxa at species level and below
- List of unambiguous taxa (~6000?)
- Treatment of most ambiguous taxa
- Demonstration mapping to FNA
- A few demosntration groups in depth

Concept workbench

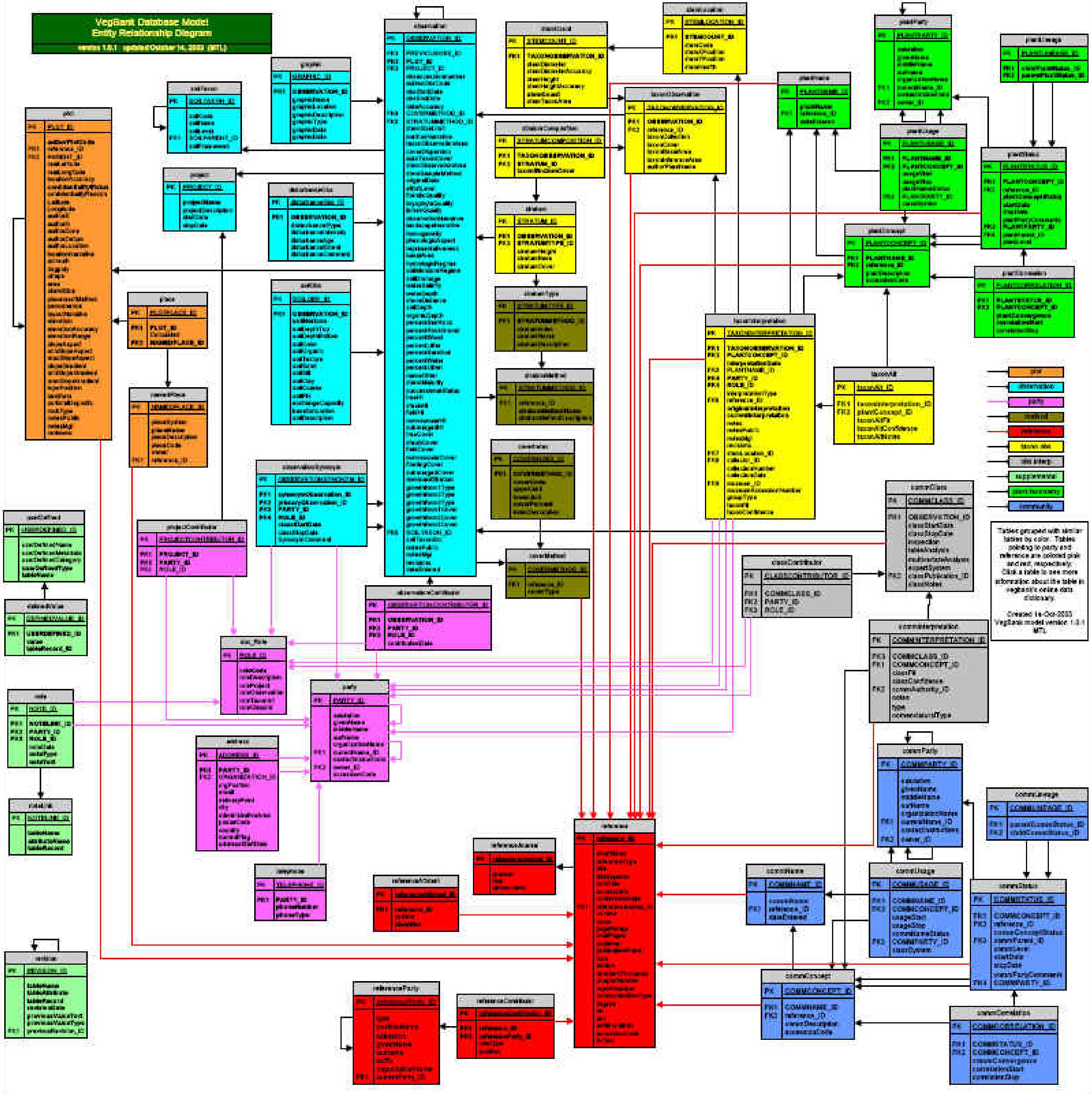
- Concept workbench for both plant concepts and community concepts is planned.

The VegBank ERD

- Available at http://vegbank.org
- Click tables for data dictionary and constrained vocabulary

VegBank Database Model
Entity Relationship Diagram

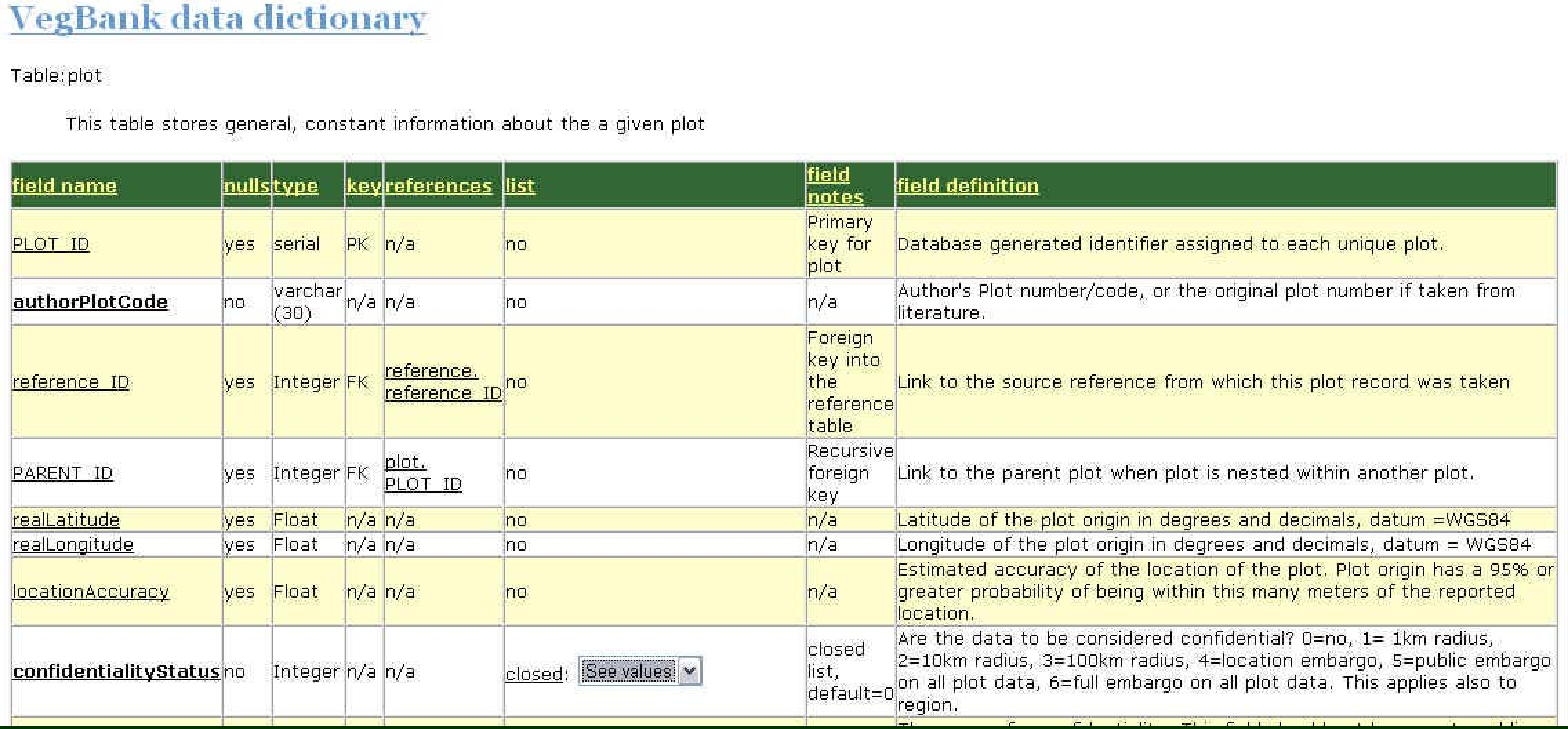
version 1.0.1 updated October 14, 2003 (MTL)



Tables grouped with similar tables by color. Tables pointing to party and reference are colored pink and red, respectively. Click a table to see more information about the table in vegbank's online data dictionary.

Created 14-Oct-2003
VegBank model version 1.0.1
MTL

The data dictionary provides critical information such as field types, field definitions, and constrained vocabularies.



VegBank data dictionary

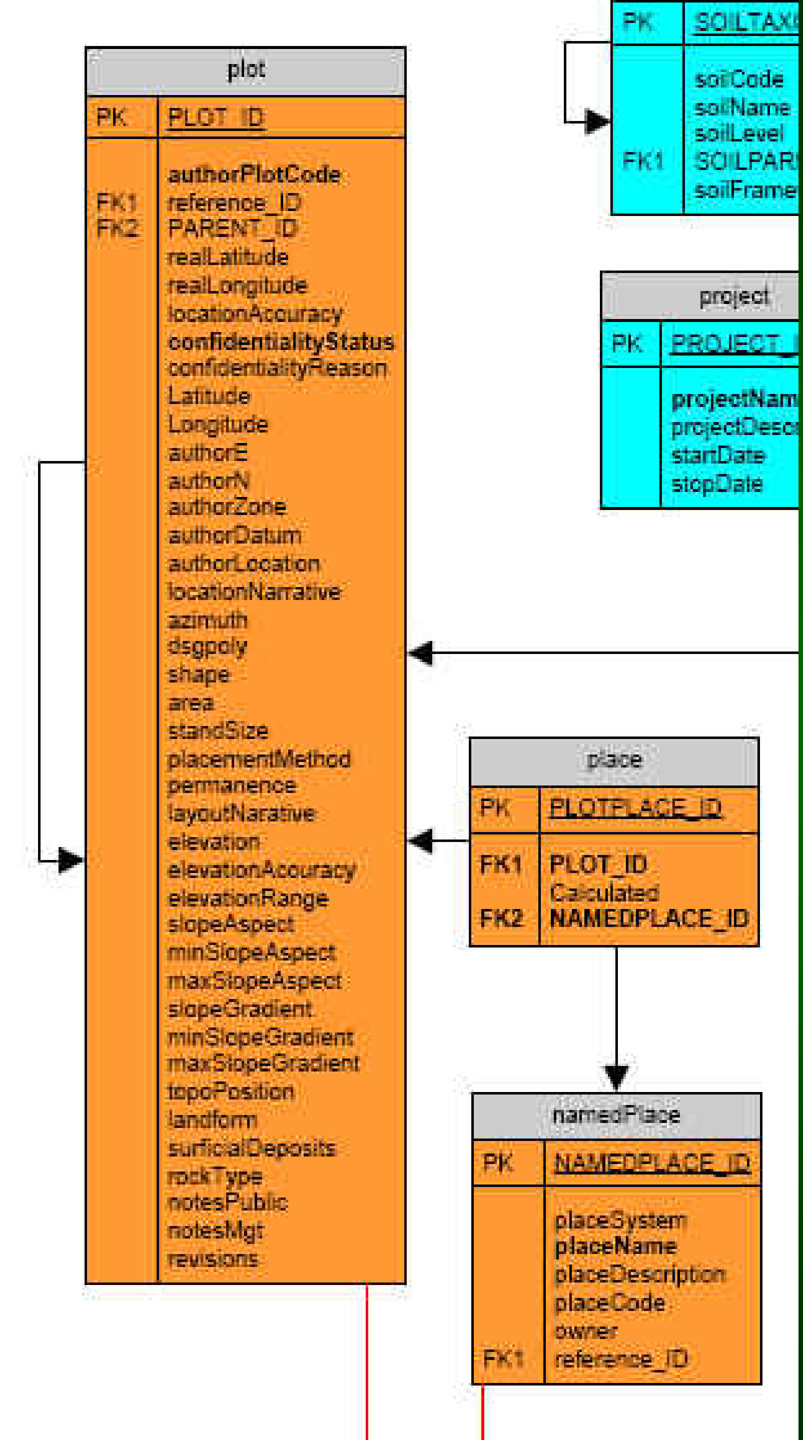
Table: plot

This table stores general, constant information about the a given plot

field name	nulls	type	key	references	list	field notes	field definition
PLOT_ID	yes	serial	PK	n/a	no	Primary key for plot	Database generated identifier assigned to each unique plot.
authorPlotCode	no	varchar (30)	n/a	n/a	no	n/a	Author's Plot number/code, or the original plot number if taken from literature.
reference_ID	yes	Integer	FK	reference. reference_ID	no	Foreign key into the reference table	Link to the source reference from which this plot record was taken
PARENT_ID	yes	Integer	FK	plot. PLOT_ID	no	Recursive foreign key	Link to the parent plot when plot is nested within another plot.
realLatitude	yes	Float	n/a	n/a	no	n/a	Latitude of the plot origin in degrees and decimals, datum =WGS84
realLongitude	yes	Float	n/a	n/a	no	n/a	Longitude of the plot origin in degrees and decimals, datum = WGS84
locationAccuracy	yes	Float	n/a	n/a	no	n/a	Estimated accuracy of the location of the plot. Plot origin has a 95% or greater probability of being within this many meters of the reported location.
confidentialityStatus	no	Integer	n/a	n/a	closed: See values ▾	closed list, default=0	Are the data to be considered confidential? 0=no, 1= 1km radius, 2=10km radius, 3=100km radius, 4=location embargo, 5=public embargo on all plot data, 6=full embargo on all plot data. This applies also to region.

Example plot metadata

- Project attributes
- Plot parties
- Observation date
- Cover & stratum methods
- Plot selection
- Plot layout
- Site data
- Geographic data



plot

PK	PLOT_ID
	authorPlotCode
FK1	reference_ID
FK2	PARENT_ID
	realLatitude
	realLongitude
	locationAccuracy
	confidentialityStatus
	confidentialityReason
	Latitude
	Longitude
	authorE
	authorN
	authorZone
	authorDatum
	authorLocation
	locationNarrative
	azimuth
	dsgpoly
	shape
	area
	standSize
	placementMethod
	permanence
	layoutNarative
	elevation
	elevationAccuracy
	elevationRange
	slopeAspect
	minSlopeAspect
	maxSlopeAspect
	slopeGradient
	minSlopeGradient
	maxSlopeGradient
	topoPosition
	landform
	surficialDeposits
	rockType
	notesPublic
	notesMgt
	revisions

PK	SOILTAX
	soilCode
	soilName
	soilLevel
FK1	SOILPAR
	soilFrame

project

PK	PROJECT_
	projectNam
	projectDesc
	startDate
	stopDate

place

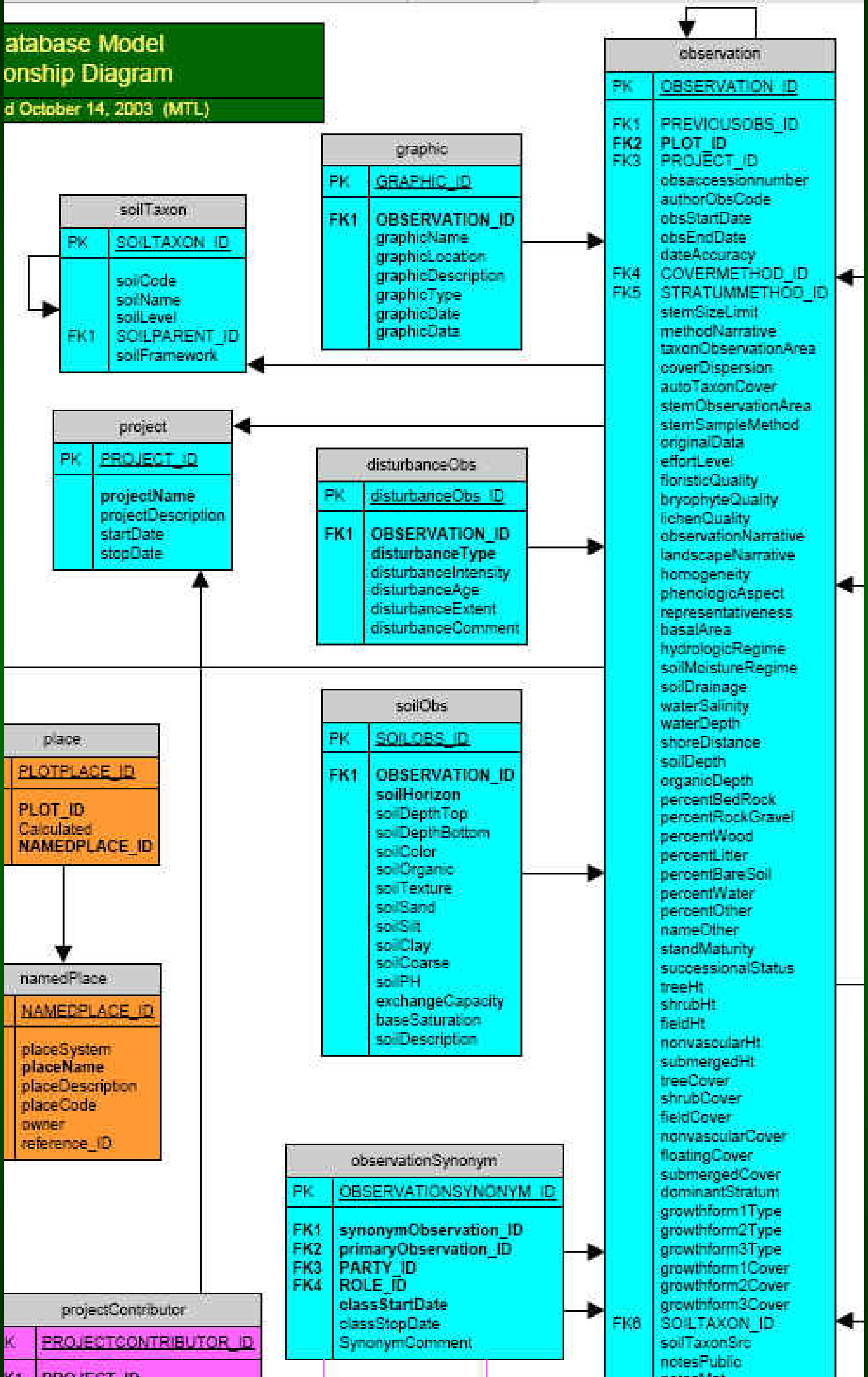
PK	PLOTPLACE_ID
FK1	PLOT_ID
	Calculated
FK2	NAMEDPLACE_ID

namedPlace

PK	NAMEDPLACE_ID
	placeSystem
	placeName
	placeDescription
	placeCode
	owner
FK1	reference_ID

Plot

- Place
- Named Place



atabase Model
onship Diagram
d October 14, 2003 (MTL)

graphic
PK GRAPHIC_ID
FK1 OBSERVATION_ID
graphicName
graphicLocation
graphicDescription
graphicType
graphicDate
graphicData

soilTaxon
PK SOILTAXON_ID
soilCode
soilName
soilLevel
FK1 SOILPARENT_ID
soilFramework

project
PK PROJECT_ID
projectName
projectDescription
startDate
stopDate

disturbanceObs
PK disturbanceObs_ID
FK1 OBSERVATION_ID
disturbanceType
disturbanceIntensity
disturbanceAge
disturbanceExtent
disturbanceComment

soilObs
PK SOILOBS_ID
FK1 OBSERVATION_ID
soilHorizon
soilDepthTop
soilDepthBottom
soilColor
soilOrganic
soilTexture
soilSand
soilSilt
soilClay
soilCoarse
soilPH
exchangeCapacity
baseSaturation
soilDescription

place
PLOTPLACE_ID
PLOT_ID
Calculated
NAMEDPLACE_ID

namedPlace
NAMEDPLACE_ID
placeSystem
placeName
placeDescription
placeCode
owner
reference_ID

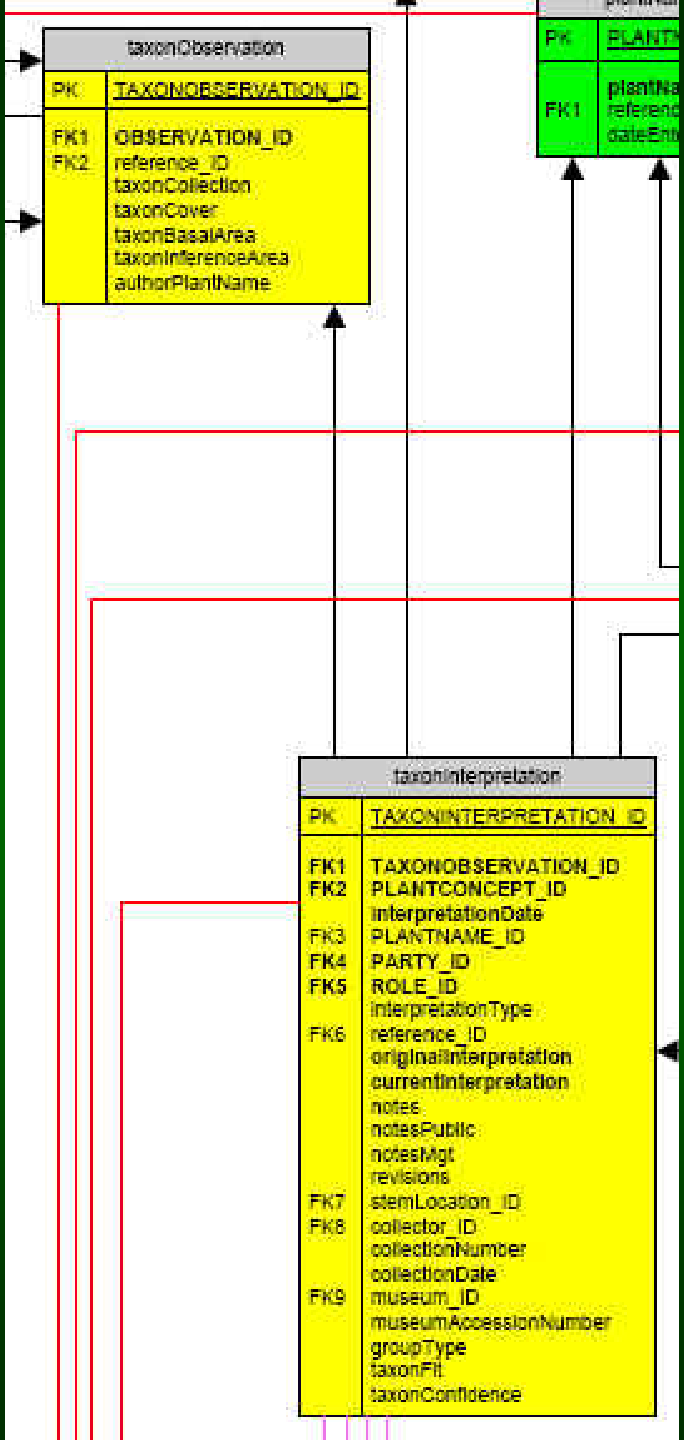
observationSynonym
PK OBSERVATIONSYNONYM_ID
FK1 synonymObservation_ID
FK2 primaryObservation_ID
FK3 PARTY_ID
FK4 ROLE_ID
classStartDate
classStopDate
SynonymComment

projectContributor
K PROJECTCONTRIBUTOR_ID
K1 PROJECT_ID

observation
PK OBSERVATION_ID
FK1 PREVIOUSOBS_ID
FK2 PLOT_ID
FK3 PROJECT_ID
obsaccessionnumber
authorObsCode
obsStartDate
obsEndDate
dateAccuracy
FK4 COVERMETHOD_ID
FK5 STRATUMMETHOD_ID
stemSizeLimit
methodNarrative
taxonObservationArea
coverDispersion
autoTaxonCover
stemObservationArea
stemSampleMethod
originalData
effortLevel
floristicQuality
bryophyteQuality
lichenQuality
observationNarrative
landscapeNarrative
homogeneity
phenologicAspect
representativeness
basalArea
hydrologicRegime
soilMoistureRegime
soilDrainage
waterSalinity
waterDepth
shoreDistance
soilDepth
organicDepth
percentBedRock
percentRockGravel
percentWood
percentLitter
percentBareSoil
percentWater
percentOther
nameOther
standMaturity
successionalStatus
treeHt
shrubHt
fieldHt
nonvascularHt
submergedHt
treeCover
shrubCover
fieldCover
nonvascularCover
floatingCover
submergedCover
dominantStratum
growthform1Type
growthform2Type
growthform3Type
growthform1Cover
growthform2Cover
growthform3Cover
FK6 SOILTAXON_ID
soilTaxonSrc
notesPublic

Observation

- Project
- Disturbance Obs
- Soil Obs
- Graphic
- Observation Synonym
- Cover method



taxonObservation

PK	TAXONOBSERVATION_ID
FK1 FK2	OBSERVATION_ID reference_ID taxonCollection taxonCover taxonBasalArea taxonInferenceArea authorPlantName

PK	PLANTN
FK1	plantNa referenc dateEnt

taxonInterpretation

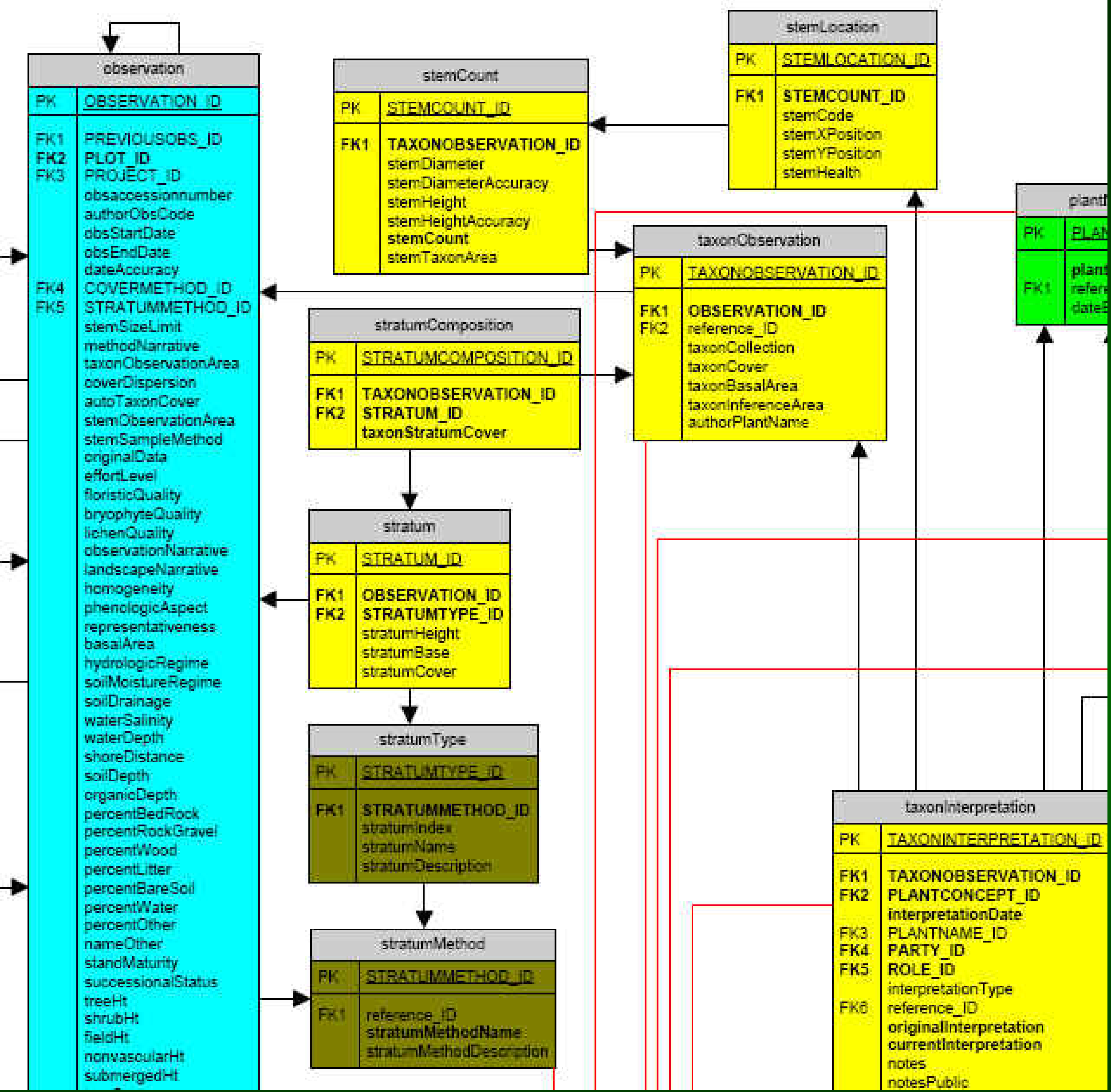
PK	TAXONINTERPRETATION_ID
FK1 FK2 FK3 FK4 FK5 FK6 FK7 FK8 FK9	TAXONOBSERVATION_ID PLANTCONCEPT_ID interpretationDate PLANTNAME_ID PARTY_ID ROLE_ID interpretationType reference_ID originalInterpretation currentInterpretation notes notesPublic notesMgt revisions stemLocation_ID collector_ID collectionNumber collectionDate museum_ID museumAccessionNumber groupType taxonFit taxonConfidence

Taxon Observation

- Importance values
- Author name

Taxon Interpretation

- Which taxon
- Who decided and why
- Stem or collective
- Voucher information



	observation
PK	OBSERVATION_ID
FK1 FK2 FK3	PREVIOUSOBS_ID PLOT_ID PROJECT_ID
	obsaccessionnumber authorObsCode obsStartDate obsEndDate dateAccuracy
FK4 FK5	COVERMETHOD_ID STRATUMMETHOD_ID
	stemSizeLimit methodNarrative taxonObservationArea coverDispersion autoTaxonCover stemObservationArea stemSampleMethod originalData effortLevel floristicQuality bryophyteQuality lichenQuality observationNarrative landscapeNarrative homogeneity phenologicAspect representativeness basalArea hydrologicRegime soilMoistureRegime soilDrainage waterSalinity waterDepth shoreDistance soilDepth organicDepth percentBedRock percentRockGravel percentWood percentLitter percentBareSoil percentWater percentOther nameOther standMaturity successionalStatus treeHt shrubHt fieldHt nonvascularHt submergedHt

	stemCount
PK	STEMCOUNT_ID
FK1	TAXONOBSERVATION_ID stemDiameter stemDiameterAccuracy stemHeight stemHeightAccuracy stemCount stemTaxonArea

	stemLocation
PK	STEMLOCATION_ID
FK1	STEMCOUNT_ID stemCode stemXPosition stemYPosition stemHealth

	plant
PK	PLAN
FK1	plant refere dateS

	taxonObservation
PK	TAXONOBSERVATION_ID
FK1 FK2	OBSERVATION_ID reference_ID taxonCollection taxonCover taxonBasalArea taxonInferenceArea authorPlantName

	stratumComposition
PK	STRATUMCOMPOSITION_ID
FK1 FK2	TAXONOBSERVATION_ID STRATUM_ID taxonStratumCover

	stratum
PK	STRATUM_ID
FK1 FK2	OBSERVATION_ID STRATUMTYPE_ID stratumHeight stratumBase stratumCover

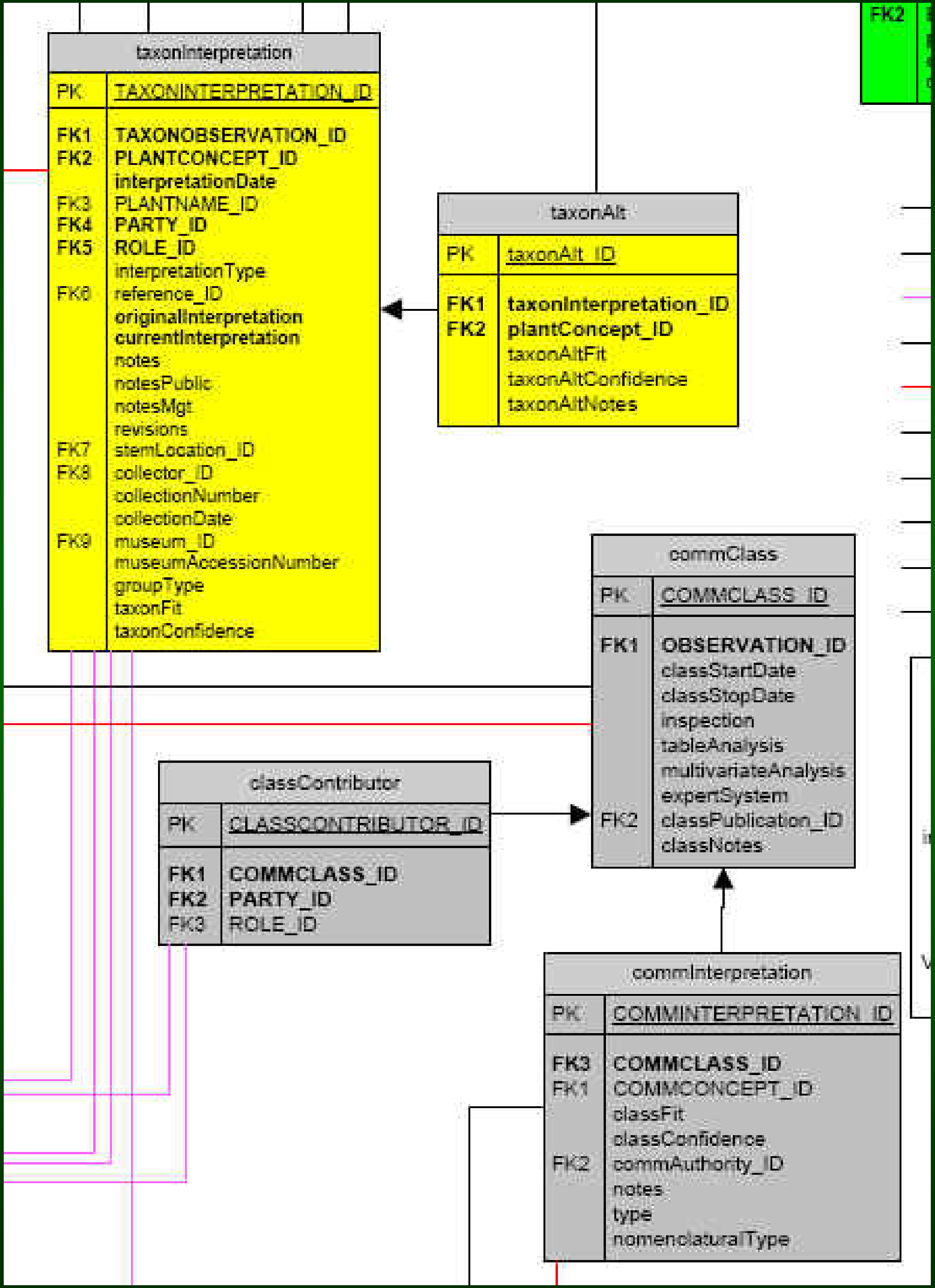
	stratumType
PK	STRATUMTYPE_ID
FK1	STRATUMMETHOD_ID stratumIndex stratumName stratumDescription

	stratumMethod
PK	STRATUMMETHOD_ID
FK1	reference_ID stratumMethodName stratumMethodDescription

	taxonInterpretation
PK	TAXONINTERPRETATION_ID
FK1 FK2	TAXONOBSERVATION_ID PLANTCONCEPT_ID interpretationDate
FK3 FK4 FK5	PLANTNAME_ID PARTY_ID ROLE_ID interpretationType
FK6	reference_ID originalInterpretation currentInterpretation notes notesPublic

Stems & Strata

- Stratum method
- Stratum type
- Stratum
- Stratum comp.
- Taxon observ.
- Stem count
- Stem location



	taxonInterpretation
PK	TAXONINTERPRETATION_ID
FK1	TAXONOBSERVATION_ID
FK2	PLANTCONCEPT_ID
	interpretationDate
FK3	PLANTNAME_ID
FK4	PARTY_ID
FK5	ROLE_ID
	interpretationType
FK6	reference_ID
	originalInterpretation
	currentInterpretation
	notes
	notesPublic
	notesMgt
	revisions
FK7	stemLocation_ID
FK8	collector_ID
	collectionNumber
	collectionDate
FK9	museum_ID
	museumAccessionNumber
	groupType
	taxonFit
	taxonConfidence

	taxonAlt
PK	taxonAlt_ID
FK1	taxonInterpretation_ID
FK2	plantConcept_ID
	taxonAltFit
	taxonAltConfidence
	taxonAltNotes

	commClass
PK	COMMCLASS_ID
FK1	OBSERVATION_ID
	classStartDate
	classStopDate
	inspection
	tableAnalysis
	multivariateAnalysis
	expertSystem
FK2	classPublication_ID
	classNotes

	classContributor
PK	CLASSCONTRIBUTOR_ID
FK1	COMMCLASS_ID
FK2	PARTY_ID
FK3	ROLE_ID

	commInterpretation
PK	COMMINTERPRETATION_ID
FK3	COMMCLASS_ID
FK1	COMMCONCEPT_ID
	classFit
	classConfidence
FK2	commAuthority_ID
	notes
	type
	nomenclaturalType

Interpretation continued

Plants

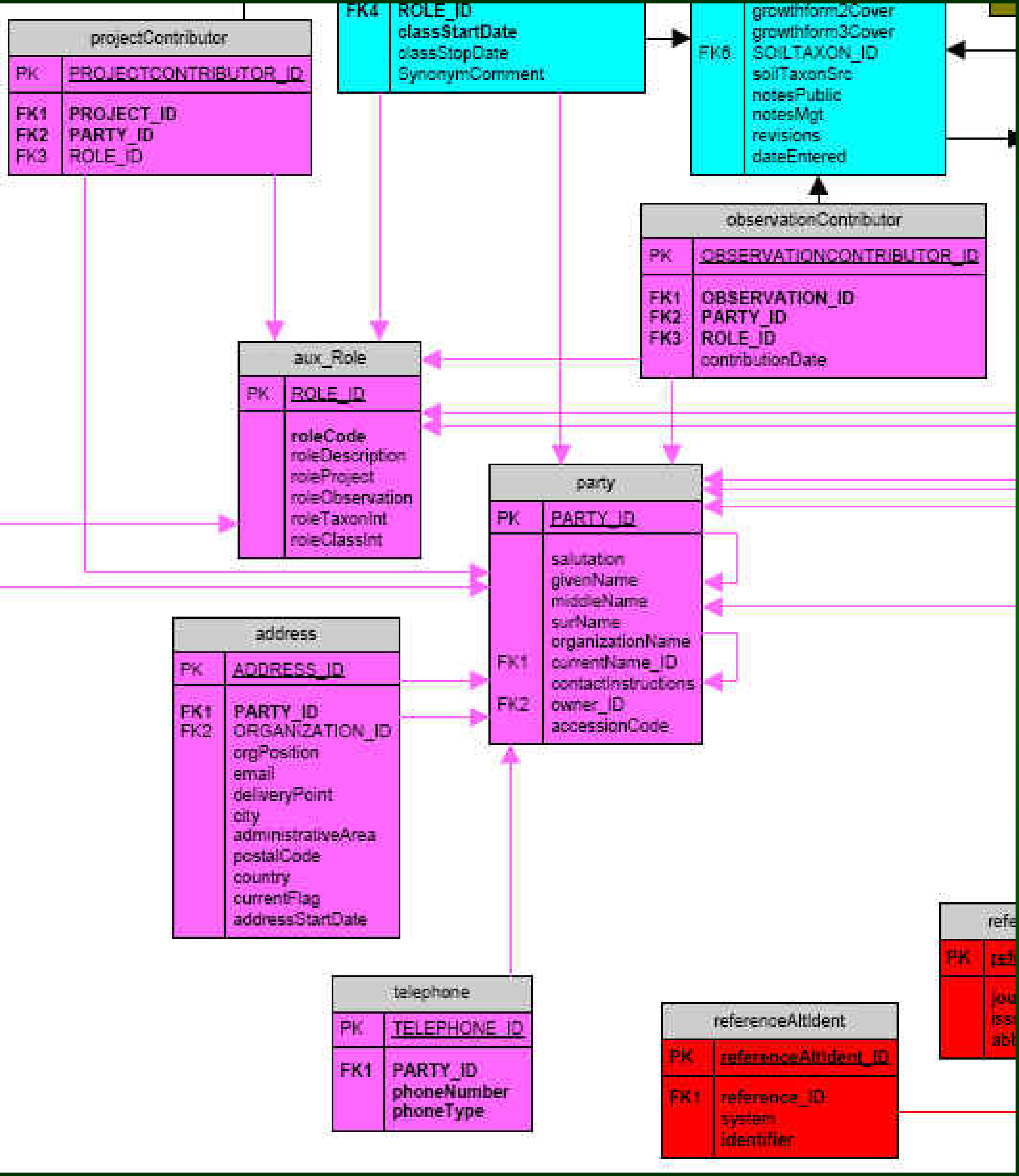
- Taxon Interpretation
- Taxon Alt

Communities

- Class
- Interpretation

Problematic taxa of ecological datasets

- *Carex sp.*
- Crustose lichen
- Hairy sedge #6.
- *Sporobolus sp. #1*
- *Picea glauca – engelmannii* complex
- *Potentilla simplex* or *P. canadensis*
- *Carya ovata sec.* Gleason 1952



FK4 ROLE_ID
classStartDate
classStopDate
SynonymComment

growthform2Cover
growthform3Cover
FK6 SOILTAXON_ID
soilTaxonSrc
notesPublic
notesMgt
revisions
dateEntered

projectContributor
PK PROJECTCONTRIBUTOR_ID
FK1 PROJECT_ID
FK2 PARTY_ID
FK3 ROLE_ID

observationContributor
PK OBSERVATIONCONTRIBUTOR_ID
FK1 OBSERVATION_ID
FK2 PARTY_ID
FK3 ROLE_ID
contributionDate

aux_Role
PK ROLE_ID
roleCode
roleDescription
roleProject
roleObservation
roleTaxonInt
roleClassInt

party
PK PARTY_ID
salutation
givenName
middleName
surName
organizationName
FK1 currentName_ID
contactInstructions
FK2 owner_ID
accessionCode

address
PK ADDRESS_ID
FK1 PARTY_ID
FK2 ORGANIZATION_ID
orgPosition
email
deliveryPoint
city
administrativeArea
postalCode
country
currentFlag
addressStartDate

telephone
PK TELEPHONE_ID
FK1 PARTY_ID
phoneNumber
phoneType

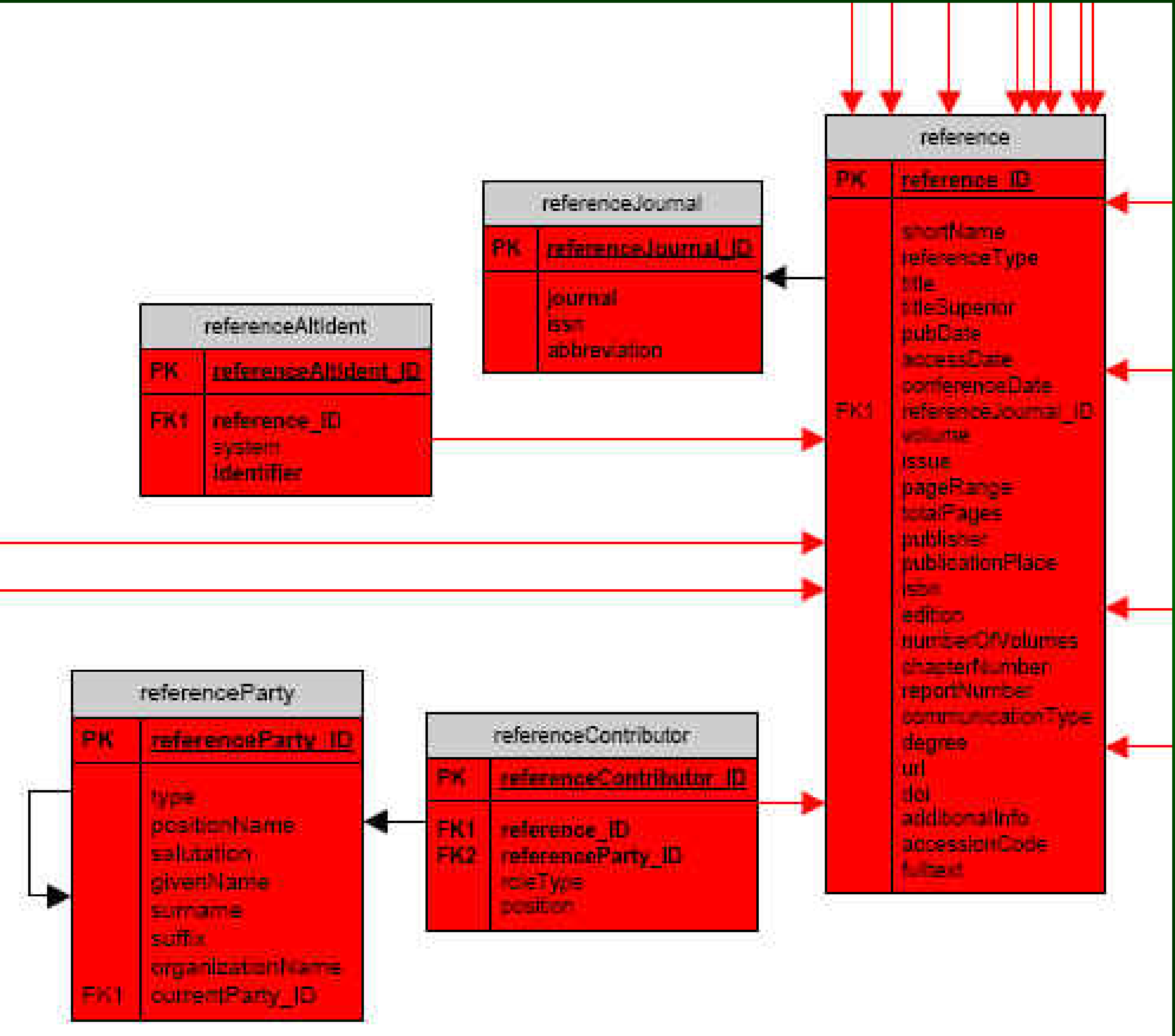
referenceAltIdent
PK referenceAltIdent_ID
FK1 reference_ID
system
identifier

refe
PK ref
jou
iss
abb

Party

- Project Contr.
- Obs Contr.
- Role
- Address
- Telephone

References



reference

PK reference_ID

FK1 referenceJournal_ID

shortName
referenceType
title
titleSuperior
pubDate
accessDate
conferenceDate
referenceJournal_ID
volume
issue
pageRange
totalPages
publisher
publicationPlace
isbn
edition
numberOfVolumes
chapterNumber
reportNumber
communicationType
degree
url
doi
additionalInfo
accessionCode
fulltext

referenceJournal

PK referenceJournal_ID

journal
issn
abbreviation

referenceAltIdent

PK referenceAltIdent_ID

FK1 reference_ID
system
identifier

referenceParty

PK referenceParty_ID

type
positionName
salutation
givenName
surname
suffix
organizationName
FK1 currentParty_ID

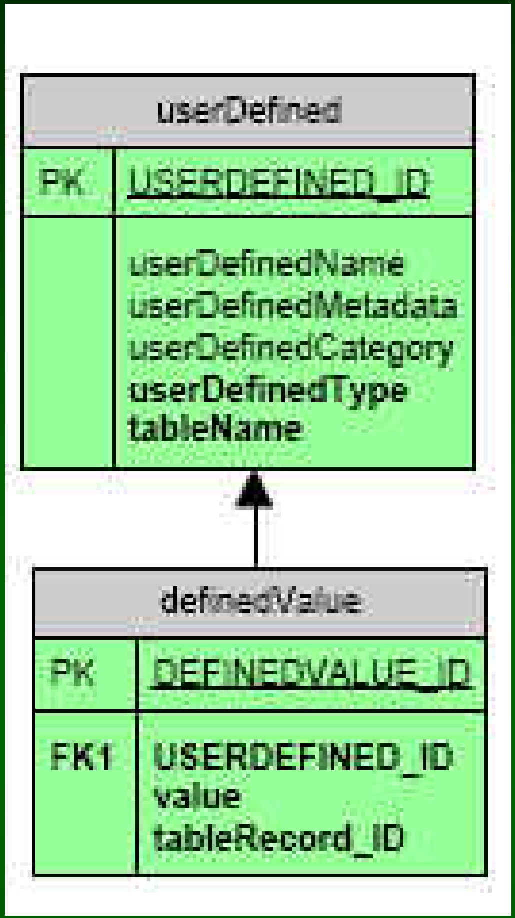
referenceContributor

PK referenceContributor_ID

FK1 reference_ID
FK2 referenceParty_ID
roleType
position

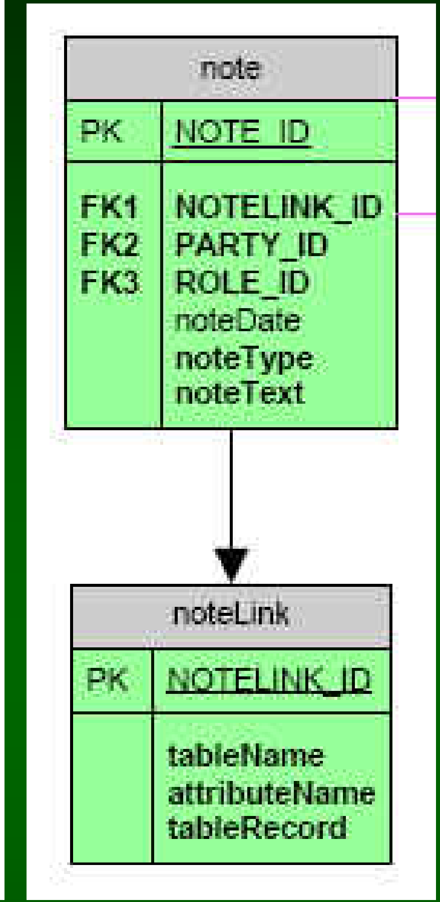
Utilities

- User defined
- Notes
- Revisions



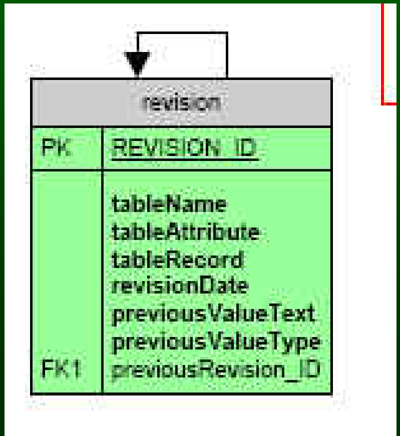
	userDefined
PK	USERDEFINED_ID
	userDefinedName userDefinedMetadata userDefinedCategory userDefinedType tableName

	definedValue
PK	DEFINEDVALUE_ID
FK1	USERDEFINED_ID value tableRecord_ID



	note
PK	NOTE_ID
FK1 FK2 FK3	NOTELINK_ID PARTY_ID ROLE_ID noteDate noteType noteText

	noteLink
PK	NOTELINK_ID
	tableName attributeName tableRecord



	revision
PK	REVISION_ID
FK1	tableName tableAttribute tableRecord revisionDate previousValueText previousValueType previousRevision_ID

Intellectual Property issues

- Rare species
- Private lands
- Working datasets – not yet complete
- Ongoing research
- Citation
- Annotation

Connectivity & Collaboration

- Loaders for popular plot databases
- Data exchange standards for plots
- Data exchange standards for taxa
- Refresh activities among VegBank, Biotics, and ITIS/PLANTS.
- Distributed VegBank systems
- Deep links into VegBank

Possible VegBank nodes

- US - ESA
- New Zealand
- Canada
- Amazon collaboration
- Europe
- South Africa

Tools for semantic mediation & data discovery: Science Environment for Ecological Knowledge

To improve how researchers can

1) gain global access to ecological data and information,

2) rapidly locate and utilize distributed computational services, and

3) capture, reproduce, and extend the analysis process itself.

The SEEK project

- Standard data structures.
- Public data archives (deposit, withdraw, cite).
- Standard exchange formats.
- Standard protocols.
- Tools for semantic mediation & data discovery.