

The VegBank taxonomic datamodel

Robert K. Peet

Sponsored by:

The Ecological Society of America
US National Science Foundation

Produced at:

The National Center for Ecological Analysis and
Synthesis

1. The Intended User Groups

Specifically

- "Ecologists" who record species composition of vegetation.
- "Ecologists" who use records of vegetation composition.

In general

- Creators and distributors of data wherein organisms need to be labeled.
- Users and consolidators of data wherein organisms are labeled.

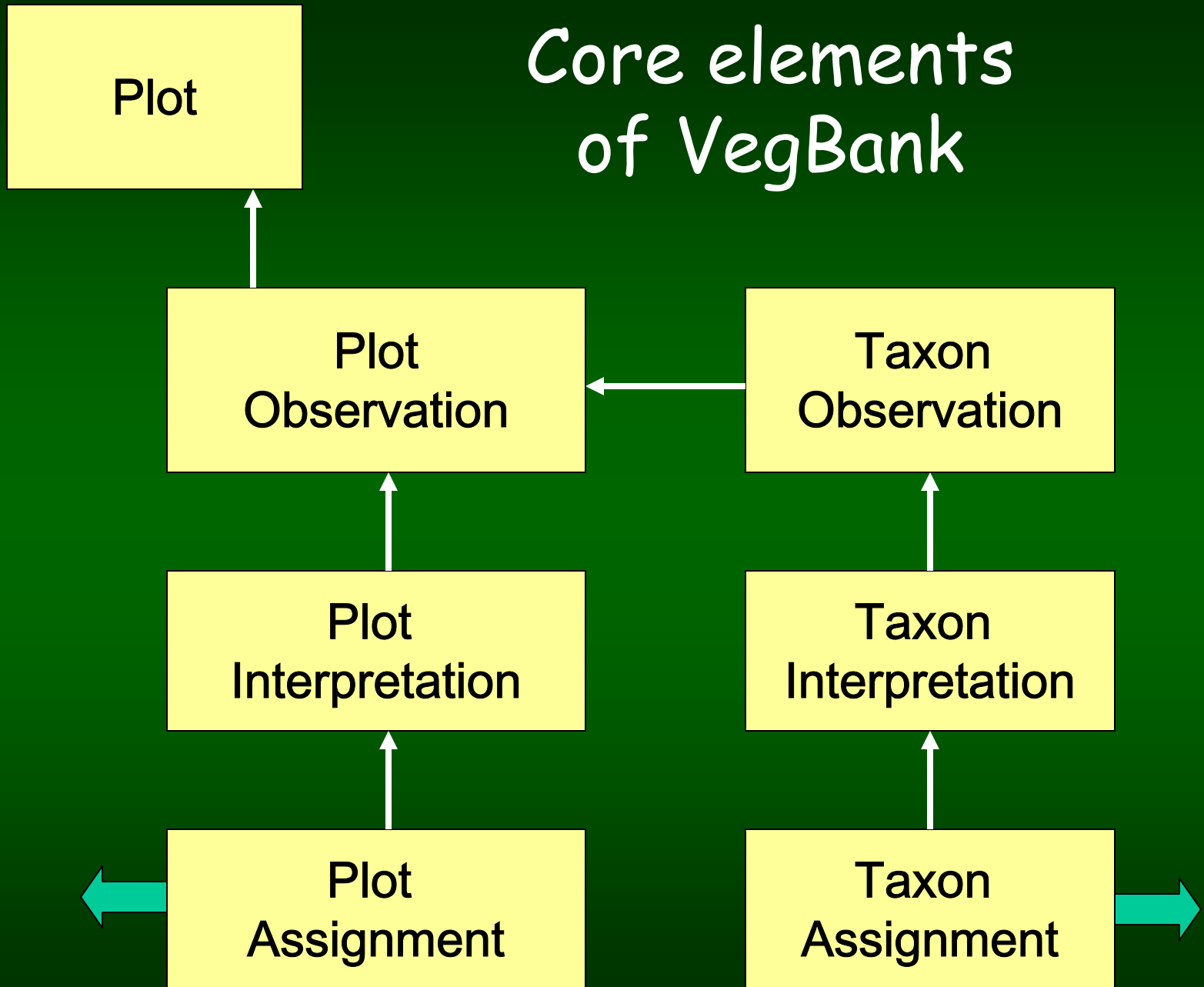
VegBank

- The ESA Vegetation Panel is currently developing *VegBank* (www.vegbank.org) as a public vegetation plot archive
- *VegBank* is expected to function for vegetation plot data in a manner analogous to *GenBank*.
- Primary data will be deposited for reference, novel synthesis, and reanalysis, particularly for classification.

2. Intended Functionality

- Ecologists, like most users of organism names, care about how to label organisms, and how to interpret labels others have placed on organisms.
- We wish for ease in combining datasets.
- We abhor name changes and ambiguity

Core elements of VegBank



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 lists are available 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

Flora North America

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.

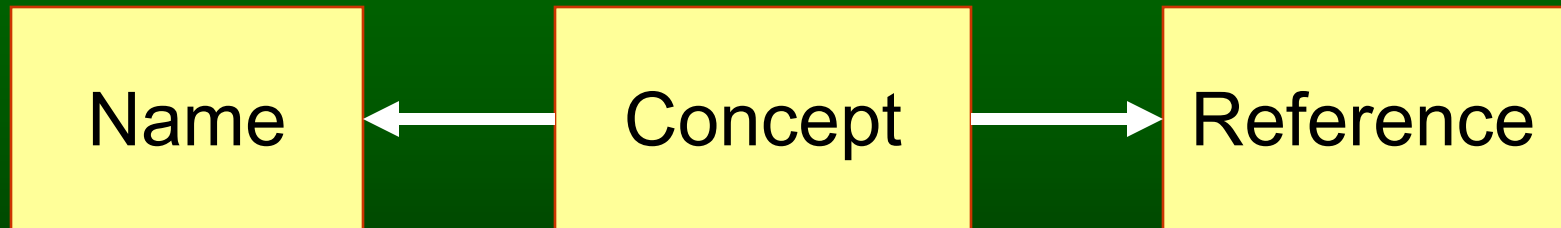
Intended functionality

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

3. Taxonomic theory

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

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



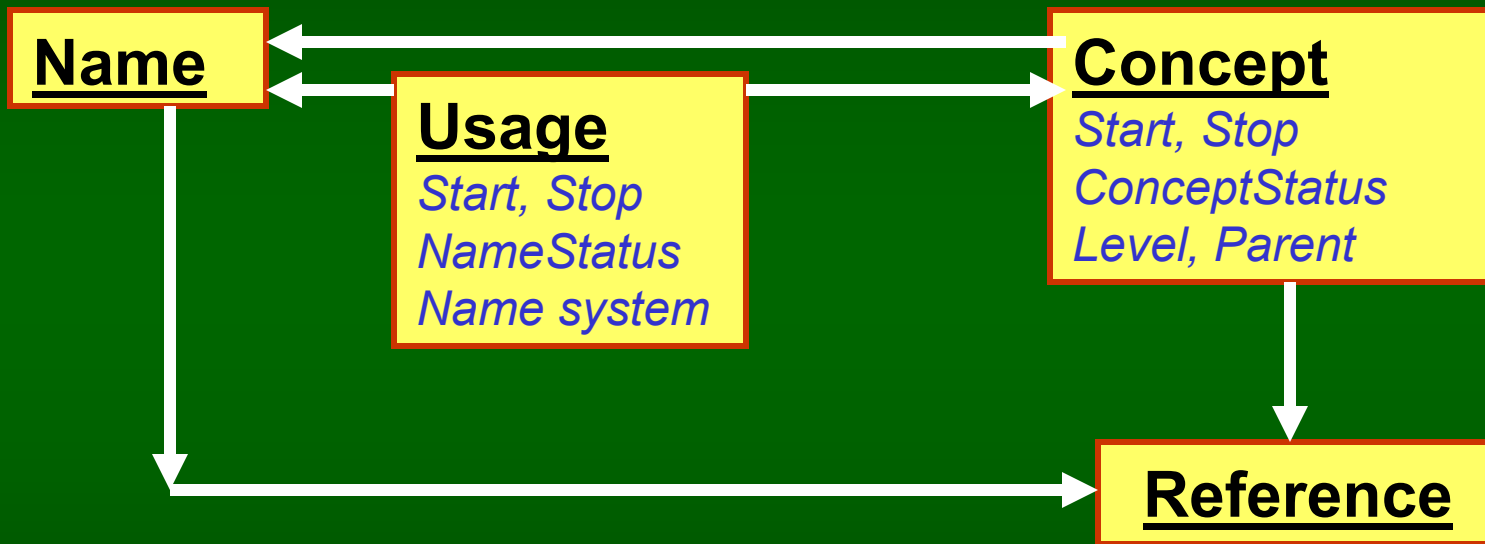
A usage represents a unique association of a concept with a name.



- 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

Data relationships

VegBank taxonomic data model



Single party, dynamic perspective

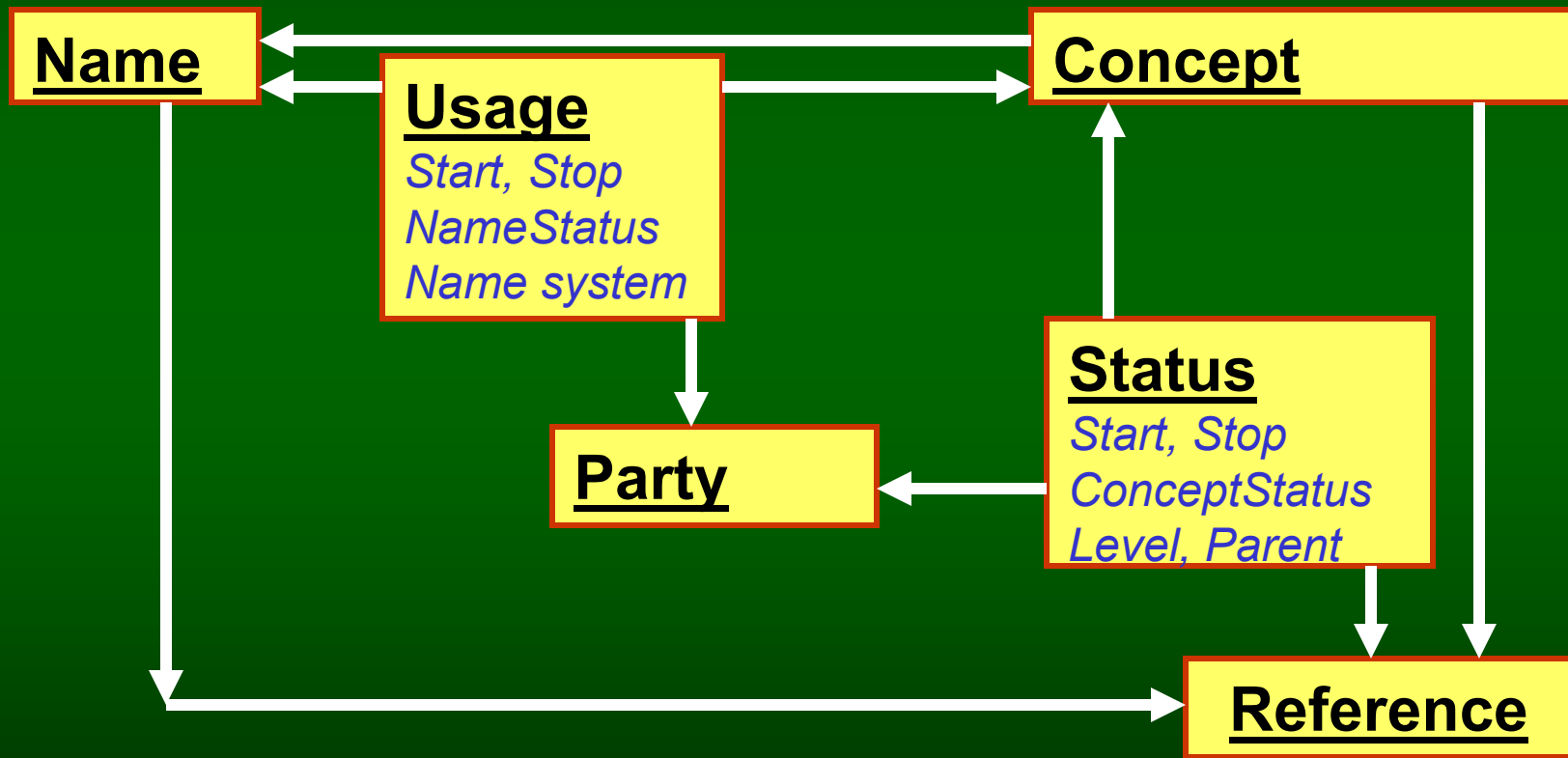
Party Perspective

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.

Data relationships

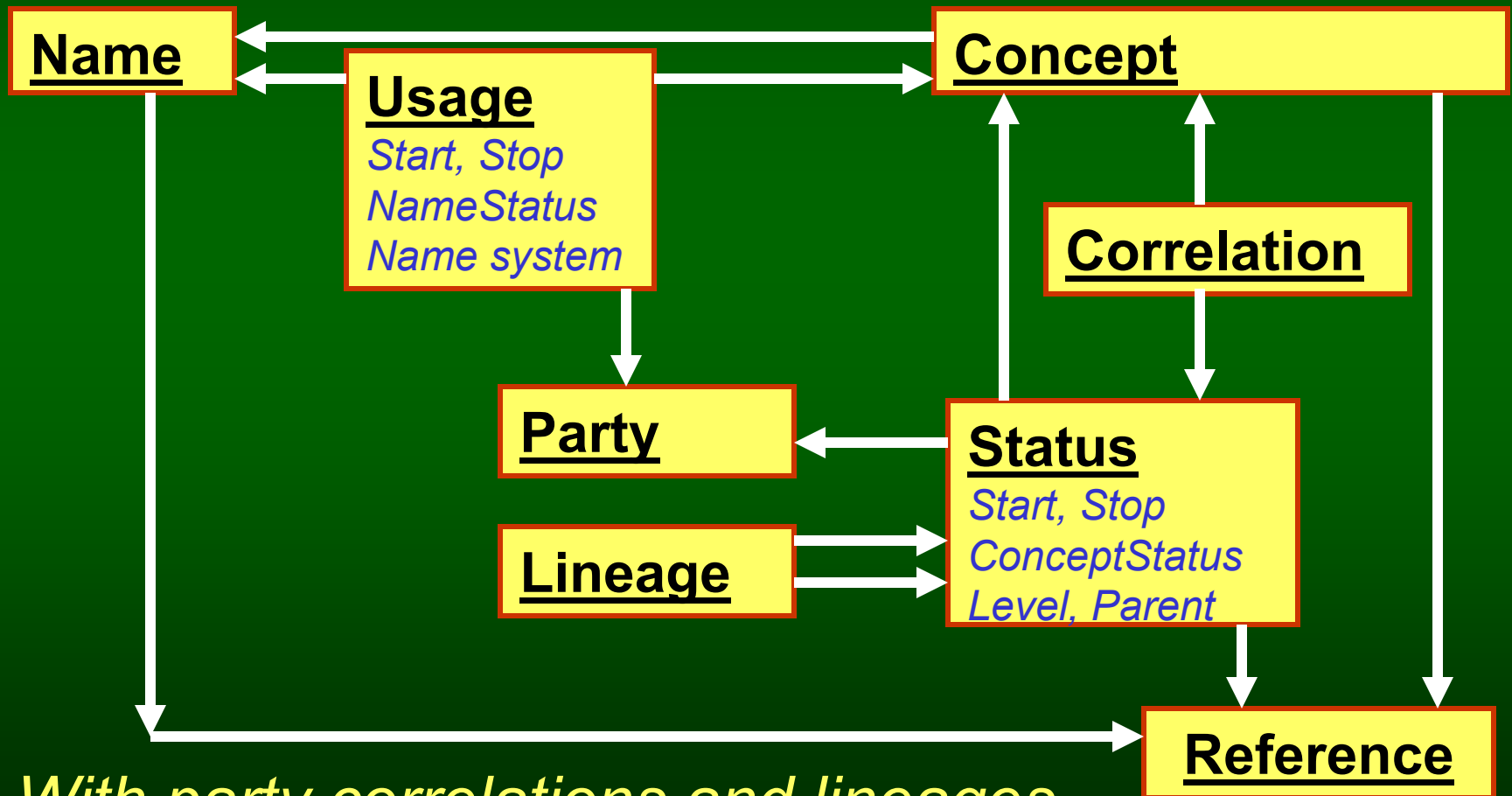
VegBank taxonomic data model



Multiple parties, dynamic perspectives

Data relationships

VegBank taxonomic data model



Primary differences between the VegBank and IOPI models

The IOPI model is optimized for

- describing taxonomic decisions represented in literature.

The VB model is optimized for

- stability in accepted concepts (super concepts),
- support of multiple dynamic party perspectives,
- support of multiple name systems.

4. State of Development

1. VegBank

2. Collaborators

- NatureServe Biotics4
- USDA PLANTS & ITIS

Status of VegBank

- Working Prototype; open for deposit Nov 1
- Production version July 2004
- Access version on VegBranch today
- Efficient data exchange by July 2004
- Functionality mandated in draft FGDC standards
- IAVS working group established for data exchange standards established 2003

VegBank 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 demonstration groups in depth

Concept workbench

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

NatureServe Biotics 4

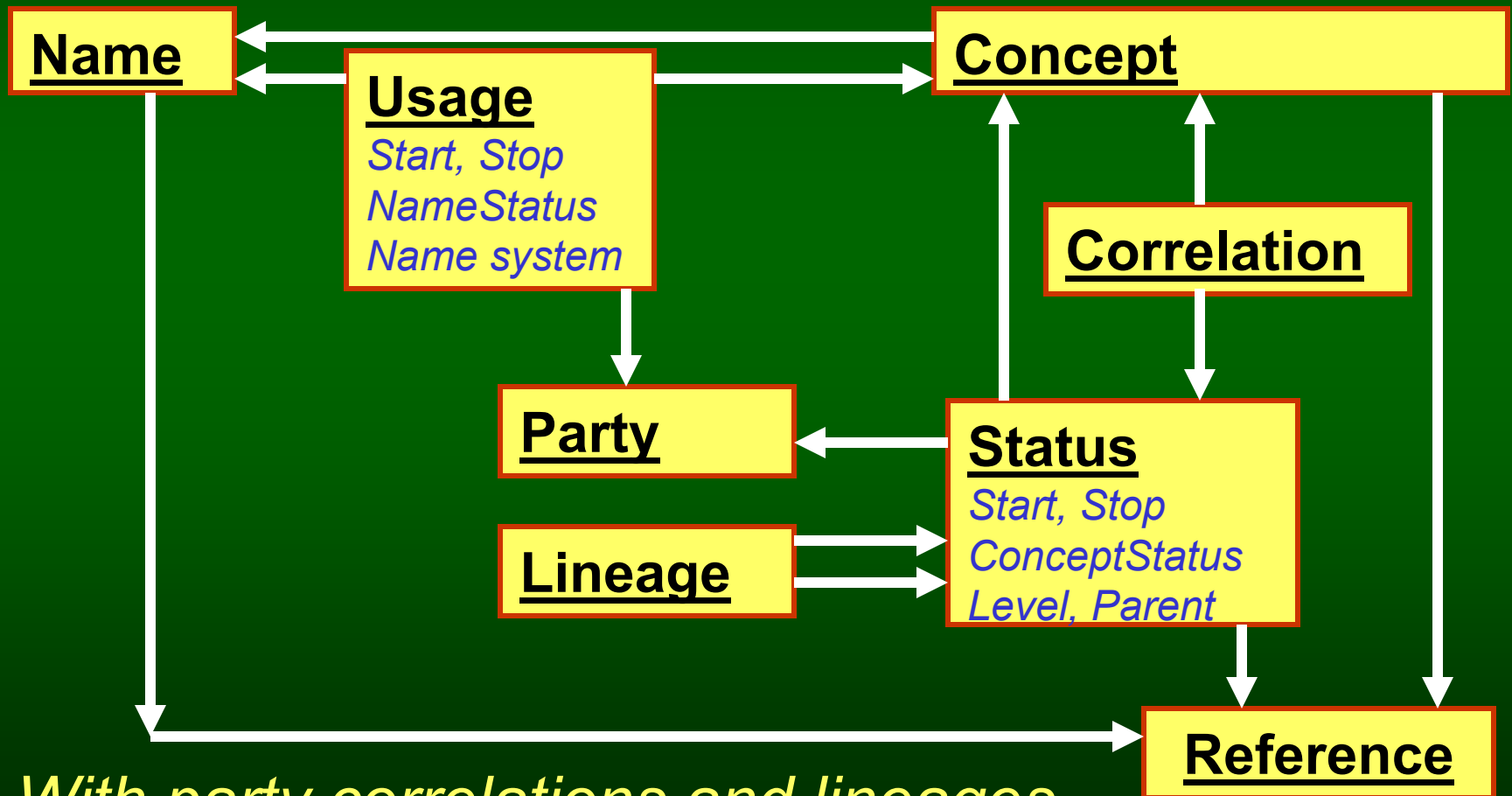
- In production
- Loading the same initial concepts
- Already using concepts where strong differences between state programs
- Routine data exchange with Vegbank under development.

PLANTS & ITIS

- Frequent communication during redesign.
- Design documents for conversion to concept-based system complete
- Collaboration to assure that the concepts we develop will be employed in the new system
- Uncertainties about funding

Data relationships

VegBank taxonomic data model



Core elements of the IOPI model

