

Challenges in aggregating invasive species occurrence data

Rebekah D. Wallace and Charles T. Barger



**Center for Invasive Species
and Ecosystem Health**
UNIVERSITY OF GEORGIA

Bugwood

- Started 1994 - Outreach/Extension Focused
- Create freely available tools to help groups work together

- EDDMapS Data Coordinator 2011-present
 - Work with agencies, programs, groups to enter existing data into the EDDMapS Database
 - Work with verifiers in the record review process
 - Process specific data requests



Museums/Herbaria have standards!



United States Department of the Interior

NATIONAL PARK SERVICE
WASHINGTON, D.C. 20240

IN REPLY REFER TO:

1 March, 1983

GUIDELINES FOR THE USE OF NPS NATURAL HISTORY MUSEUM LABELS

By : Jonathan Bayless
NPS, WASO

The following Museum specimen labels for NPS natural history collections are available from the Division of Museum Services, Harpers Ferry Center.

<u>NPS FORM #</u>	<u>Label type</u>	<u>Height by width * (inches)</u>
10-500	Vetebrate Wet Specimen Label	3 X 5
10-501	Vertebrate Specimen Label	3/4 X 3 1/4

<http://npshistory.com/publications/curation/nh-museum-labels.pdf>

CIDOC Standards, guidelines

What is a standard?

The British Standards Institution (BSI), the world's oldest standards setting organisation (1901), says:

Put at its simplest, a standard is an agreed, repeatable way of doing something. It is a published document that contains a technical specification or other precise criteria designed to be **used consistently as a rule, guideline, or definition**. Standards help **to make life simpler** and to increase the reliability and the effectiveness of many goods and services we use. Standards are created by **bringing together the experience and expertise of all interested parties** such as the producers, sellers, buyers, users and regulators of a particular material, product, process or service." (our **Bolding**).

To the advantages given above can be added: **delivering interoperability**.

- [LIDO - Lightweight Information Describing Objects, 2010](#)
- [The CIDOC Conceptual Reference Model \(latest official release\)](#)
- [Object ID](#)
- [Introduction to Metadata, third edition, edited by Murtha Baca](#)
- [Statement of principles of museum documentation, 2012](#)
- [Statement on Linked Data identifiers for museum objects, 2012](#)
- [International Guidelines for Museum Object Information: The CIDOC Information Categories, 1995](#)
- [CIDOC Fact sheet No 3 - Recommendations for identity photographs, 2010](#)
- [CIDOC Fact Sheet No 2 - Labelling and marking objects, 1994](#)
- [CIDOC Fact Sheet No 1 - Registration step by step, 1993](#)

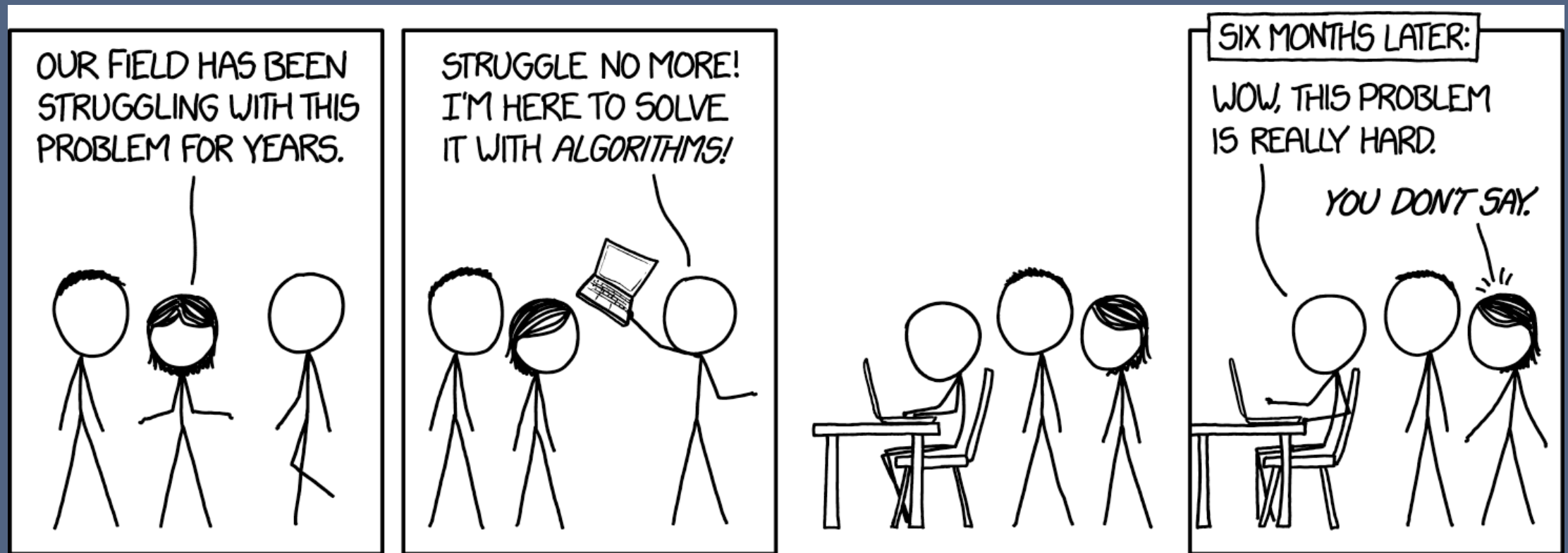
<http://network.icom.museum/cidoc/standards/cidoc-standards-guidelines/>

Occurrence data collection... we try!

- Everyone has their own data they need to collect
- Methods of collection widely vary
- Reluctant to collect more than they need
- Want others to collect to fit their standards
- How to accommodate all, or even just most?



Just need some standardization!



R Monroe. xkcd – Here to help <https://xkcd.com/1831/>



Center for Invasive Species
and Ecosystem Health
UNIVERSITY OF GEORGIA

EDD MapS
find • map • track

What is occurrence data? “X” was “Here”

X

- Can be an organism at any level of taxonomy, documented in any possible way, or not at all!
 - Scientific Name
 - Common Name
 - Short-hand/abbreviated name
 - Database code

Here

- Can be a location documented at any level of precision
 - GPS
 - County, State
 - State
 - Country
 - Etc.



Bad Practices – Coded Data (without metadata)

G	H	I	J	
	Icon	Name	Description	No
109° 23.914206' W	Flag, Blu	BH11000001	02-JUN-14 9:29:21AM	12-
109° 31.590155' W	Flag, Blu	BH11000002	02-JUN-14 12:18:10PM	12-
110° 8.312879' W	Flag, Blu	BH11000005	07-OCT-14 9:41:43AM	07-
109° 14.853095' W	Flag, Blu	BH11000031	08-SEP-14 11:05:27AM	18-
109° 16.627060' W	Flag, Blu	BH11000032	08-SEP-14 11:33:36AM	18-
109° 39.534308' W	Flag, Blu	BOAT DOCK	05-SEP-06 10:55:17AM	15-
109° 45.688474' W	Flag, Blu	BOTTOM	05-SEP-06 10:32:26AM	15-
109° 16.871672' W	Flag, Blu	CT11000003	02-JUL-14 9:28:14AMSPRAYED	02-
109° 27.199634' W	Flag, Blu	CT11000004	03-SEP-14 10:30:43AM	03-
109° 42.852355' W	Flag, Blu	DEAD H P R	05-SEP-06 11:51:50AM	25-
109° 22.304378' W	Flag, Blu	DK12000001	08-JUN-14 9:30:23AM	18-
109° 20.447508' W	Flag, Blu	GALLMIDGE	07-OCT-14 12:29:52PM	17-
109° 43.176363' W	Flag, Blu	HC11000004	01-JUL-14 12:08:45PM	01-
109° 26.652136' W	Flag, Blu	HC14000001	09-MAY-14 10:52:22AM	09-
109° 29.897834' W	Flag, Blu	HT11000001	02-JUN-14 12:53:11PM	12-
109° 24.333521' W	Flag, Blu	HT11000003	02-JUN-14 1:44:36PM	12-

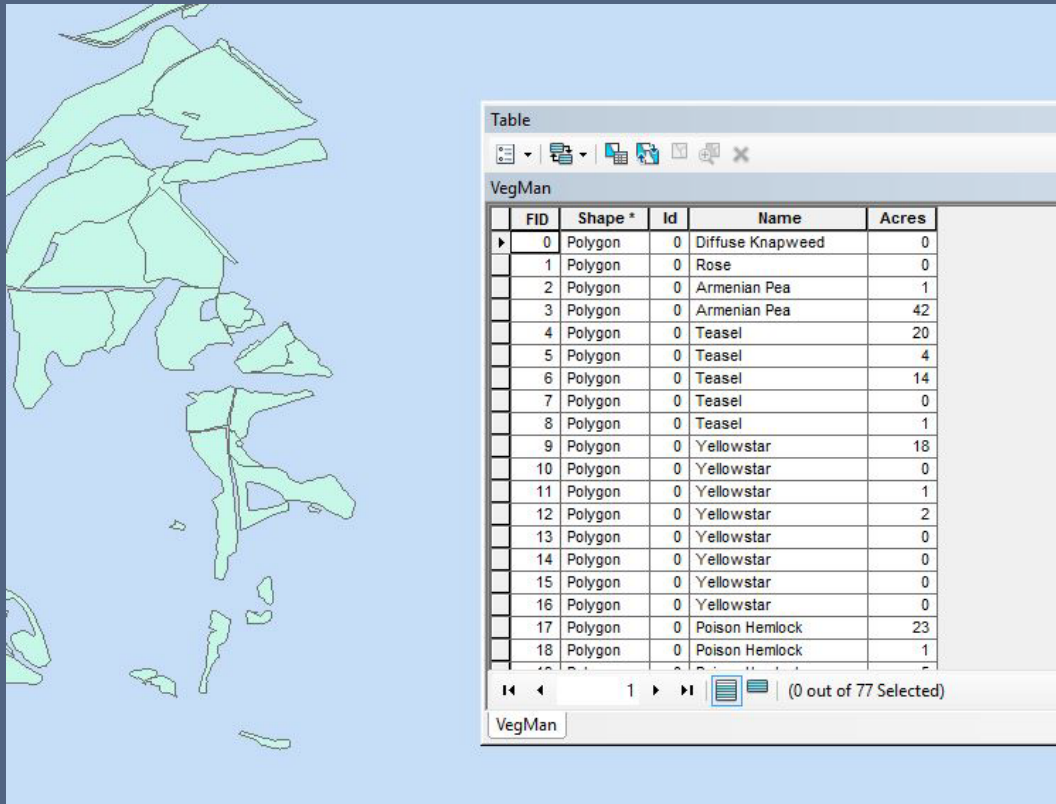
What species is being reported? CT – Canada Thistle or Camel Thorn? What do the numbers after the name mean?

	A	B	C
1	Species	ParkCodex	
2	Artemisia vulgaris	CUVA	}
3	Ammophila arenaria	LEWI	
4	Ammophila breviligulata	LEWI	
5	Cakile maritima	LEWI	
6	Capsella bursa-pastoris	LEWI	
7	Carex kobomugi	LEWI	
8	Daphne laureola	LEWI	
9	Daucus carota	LEWI	
10	Geranium dissectum	LEWI	
11	Geranium robertianum	LEWI	
12	Hypericum androsaemum	LEWI	
13	Juncus bufonius	LEWI	
14	Lotus pedunculatus	LEWI	
15	Mycelis muralis	LEWI	
16	Plantago coronopus	LEWI	
17	Soliva sessilis	LEWI	
18	Strix varia	LEWI	
19	Trifolium arvense	LEWI	
20	Trifolium subterraneum	LEWI	
21	Acer platanoides	PISC	
22	Alliaria petiolata	PISC	
23	Fallopia japonica	PISC	
24	Glechoma hederacea	PISC	
25	Halyomorpha halys	PISC	
26	Hedera helix	PISC	

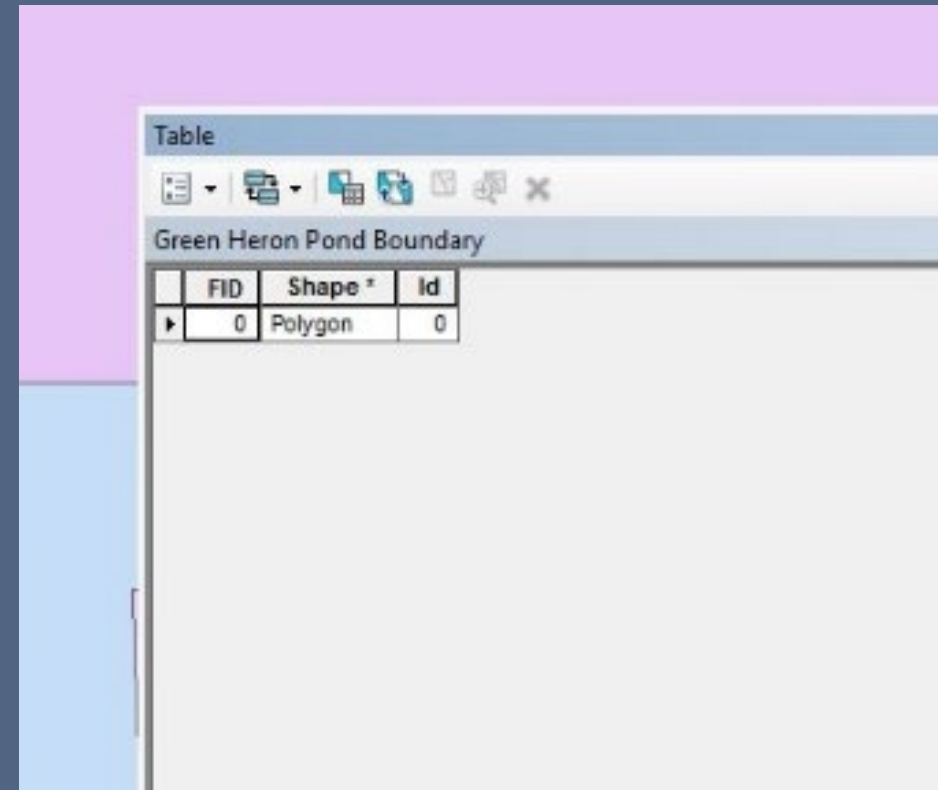
What parks are these reported in? National, State, City, County, Municipality, Regional?

Bad Practices - Missing Data

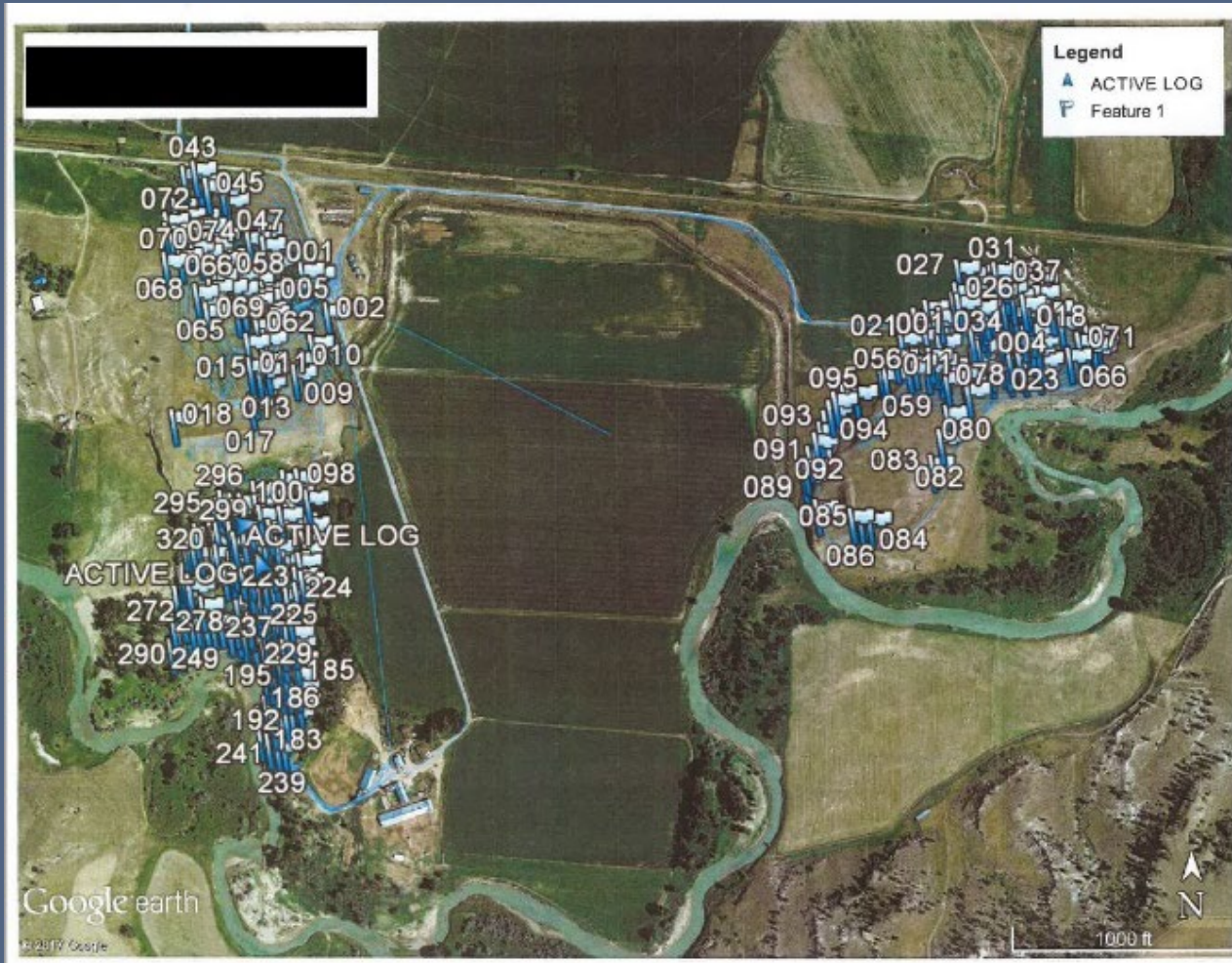
No Dates



No Species, No Dates



Bad Practices – Missing Data



“Paper Map”

- Map not provided with the KML/KMZ used to create the pdf that was submitted and so is missing
 - Location/Coordinates
 - Species Names
 - Date
 - Optional/Extra information
 - Area
 - Treatment
 - Phenology
 - Etc.



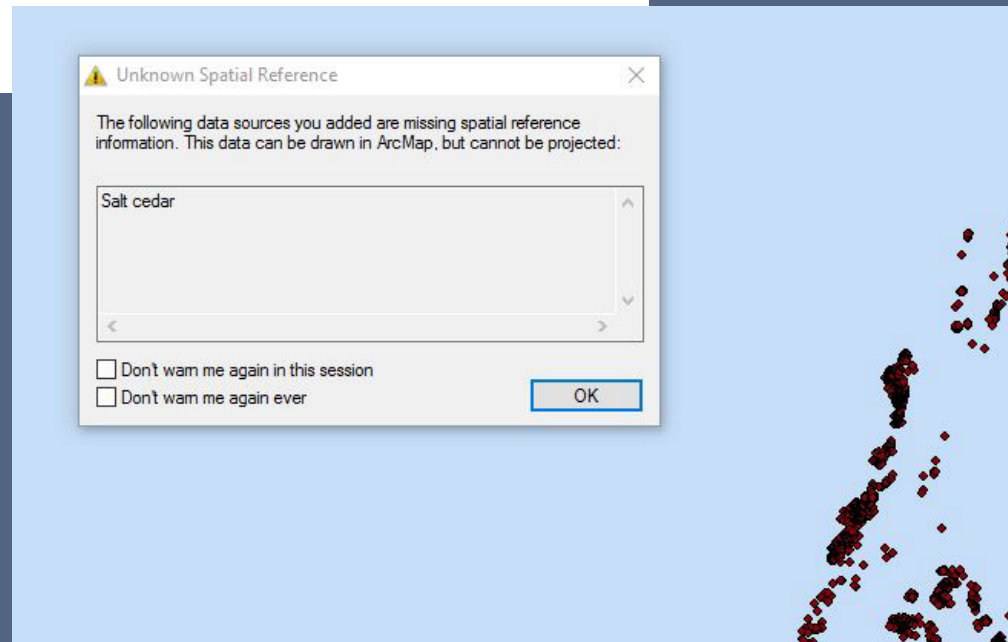
Bad Practices – Incomplete File Packets

File ID	Filename
1046	BlackHenbane 2016_Points.shp

File ID	Filename
1405	MPCD_2017Grant_Pictures.shp
1404	MPCD_2017Grant_Points.shp
1403	MPCD_2017Grant_Polygons.shp

GilpinPoints.dbf	2/3/2018 3:16 PM	DBF File
GilpinPoints.prj	2/3/2018 3:16 PM	PRJ File
GilpinPoints.sbn	2/3/2018 3:16 PM	SBN File
GilpinPoints.sbx	2/3/2018 3:16 PM	SBX File
GilpinPoints.shp	2/3/2018 3:16 PM	SHP File
GilpinPoints.shx	2/3/2018 3:16 PM	SHX File

Anything with the same Name, send it together



Bad Practices - Duplicate Data

Table

USFS Affected Area 2017_Polygons

FID	Shape	title	descriptio	type	start	end	area	areaunits	project	creator
0	Polygon	Affected Area		Polygon	9/5/2017 1:52:16 AM	9/5/2017 1:52:16 AM	75.9	Acres	USFS Affected Area 2017	slvwma-admin
1	Polygon	Affected Area		Polygon	9/11/2017 7:22:34 AM	9/11/2017 7:22:34 AM	46.18	Acres	USFS Affected Area 2017	slvwma-admin
2	Polygon			Polygon	9/5/2017 1:30:29 AM	9/5/2017 1:30:29 AM	52.39	Acres	USFS Affected Area 2017	slvwma-admin

Table

USFS2017

FID	Shape *	OID_	Name	FolderPath	SymbolID	AltMode	Base	Clamped	Extruded	Snippet
0	Polygon ZM	0	Polygon	MapItFast Project Export/Polygons	0	0	0	-1	0	<table style="border:1px solid black; t
1	Polygon ZM	0	Polygon	MapItFast Project Export/Polygons	0	0	0	-1	0	<table style="border:1px solid black; t
2	Polygon ZM	0	Polygon	MapItFast Project Export/Polygons	1	0	0	-1	0	<table style="border:1px solid black; t



Bad Practices – Incomplete/Unclear Data

Missing Zone/Datum

H	I	J	K
GPS_Height	Northing	Easting	Date
292.8	4873719.669	652937.134	5/2/2008
291.5	4873736.606	652942.147	5/2/2008
292.5	4873794.424	653008.627	5/2/2008
292.0	4873799.951	653019.027	5/2/2008
292.6	4873853.726	653161.170	5/2/2008
284.5	4873841.988	653201.100	5/2/2008
285.9	4873833.996	653218.567	5/2/2008
282.7	4873884.728	653327.649	5/2/2008
284.1	4873911.796	653369.224	5/2/2008
283.9	4873916.395	653393.760	5/2/2008
288.0	4873945.927	653427.867	5/2/2008
275.5	4874020.914	653521.550	5/2/2008
275.0	4873949.857	653474.427	5/2/2008

Missing Units

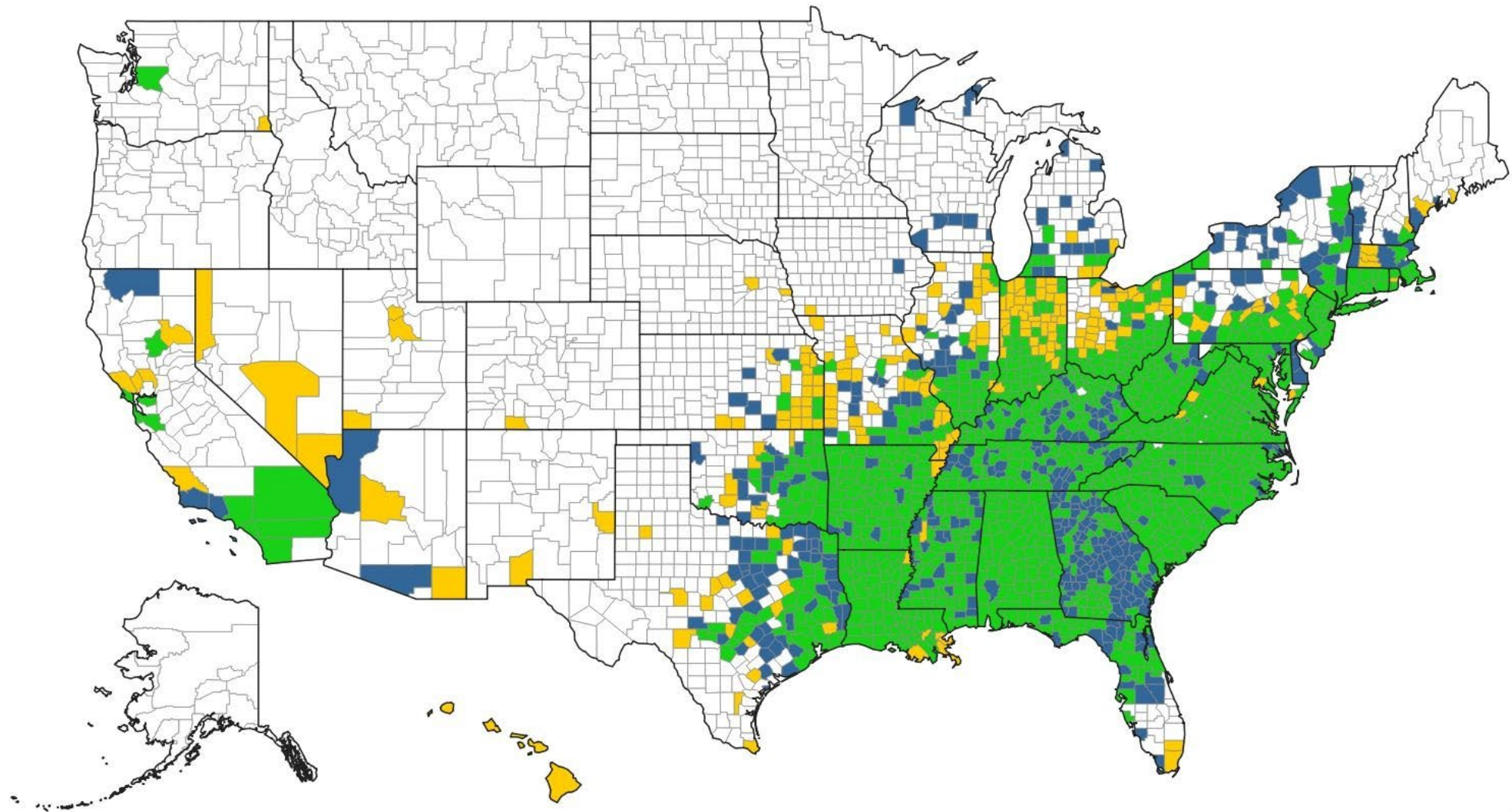
A	B	C	D
Id	Area_of_Ph		
0	965		
	64		
0	36		
0	30		
0	18		
0	5		
0	0		

Meaning of "0" Area/Count?

FID	Shape*	Id	Name	Acres
0	Polygon	0	Diffuse Knapweed	0
1	Polygon	0	Rose	0
2	Polygon	0	Armenian Pea	1
3	Polygon	0	Armenian Pea	42
4	Polygon	0	Teasel	20
5	Polygon	0	Teasel	4
6	Polygon	0	Teasel	14
7	Polygon	0	Teasel	0
8	Polygon	0	Teasel	1
9	Polygon	0	Yellowstar	18
10	Polygon	0	Yellowstar	0
11	Polygon	0	Yellowstar	1
12	Polygon			
13	Polygon		Yellowstar	0
			Yellowstar	0
15	Polygon		Yellowstar	0
16	Polygon		Yellowstar	0
17	Polygon	0	Poison Hemlock	23
18	Polygon	0	Poison Hemlock	1

Does "0" mean "not found" or "field non-nullable, defaults to 0"

Japanese honeysuckle (*Lonicera japonica*)



Legend

- No Data
- Literature only
- Observation only
- Both

Why not make an invasive species occurrence a specimen?

➤ Impractical!

- Time – Collecting specimen, taking data, taking to facility
- Money – Materials for collecting, staff time to collect/preserve, facility
- Space – Space to prepare specimen, Facility space
- Prioritization of all of the above resources spent elsewhere – Prevention/Control/Eradication, Education
- Occurrences happen at the same place through time to document changes, take a specimen every time?



Standardization of Occurrence Data

Documentation of the occurrence provides more evidence
without the need to create a specimen

➤ Original Minimums

- Who
- What
- Where
- When

➤ New Minimums?

- Original +
- Record Basis
- Record Level Identifiers
- Location Minimums



Plus more for Invasive Species!

“Minimums”

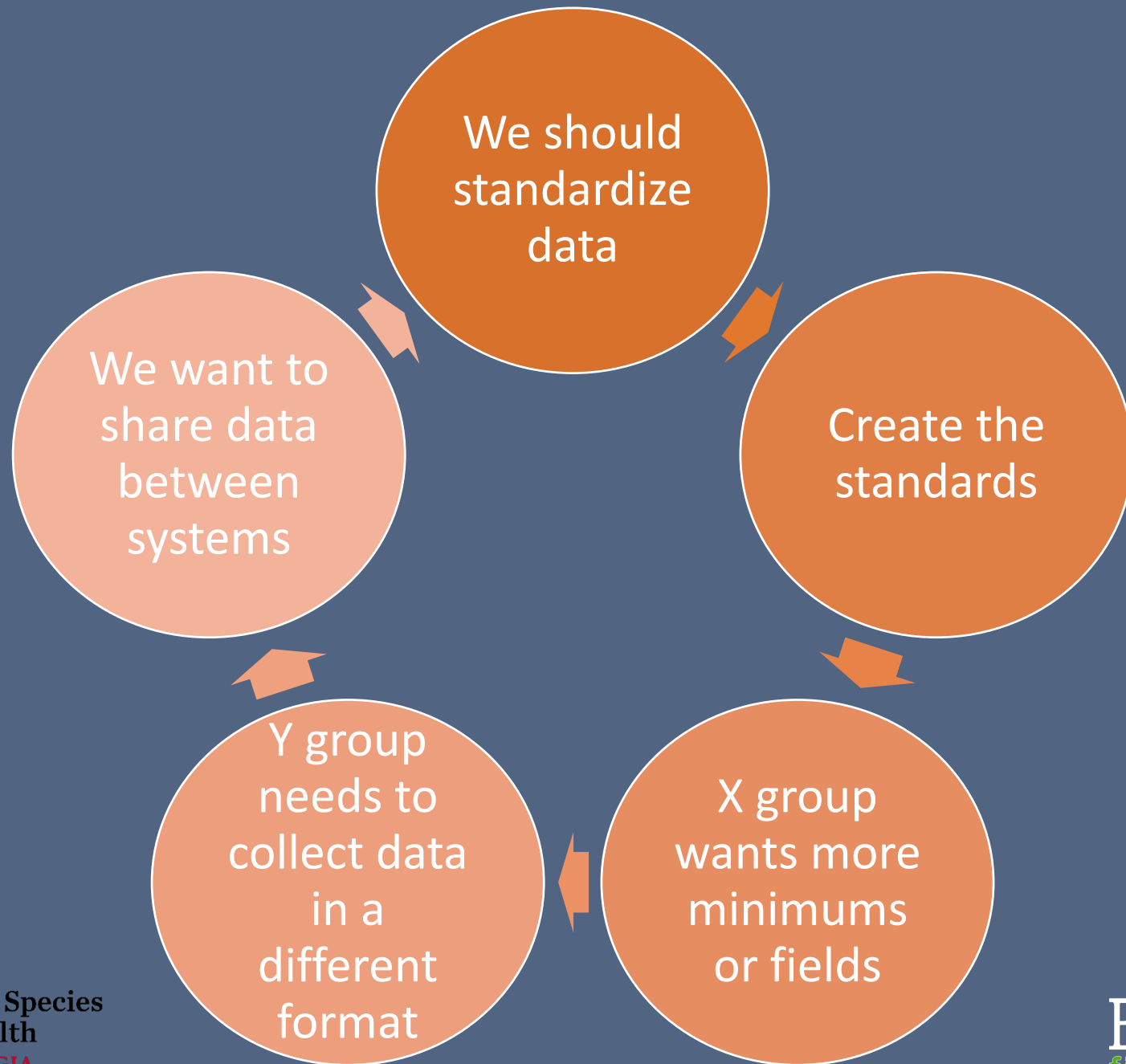
- Management Status
- Quantifications
- Host (if exists)

“Recommended”

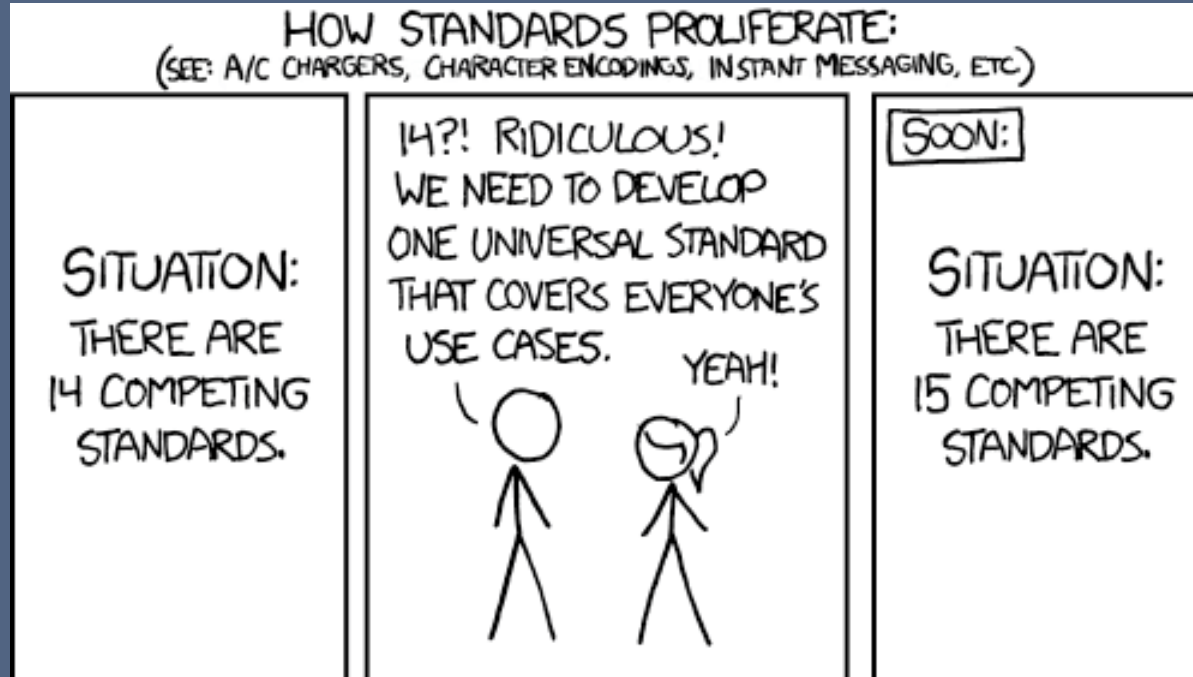
- Treatment methods/info
- Population Status
- Verification Info

Easy enough, right?





There's an xkcd for that...



R Monroe. xkcd - standards <https://xkcd.com/927/>



Center for Invasive Species
and Ecosystem Health
UNIVERSITY OF GEORGIA

EDD MapS
find • map • track



NAISMA Mapping Standards

Mapping Standards For Program Managers

**Approved by:
North American Invasive Species Management Association**

October 17, 2018



Center for Invasive Species
and Ecosystem Health
UNIVERSITY OF GEORGIA

EDD MapS
find • map • track 

So, what can we do?

- You cannot “force” a group to conform, but you can publicize the benefits of standardization and aggregation and collaboration
- Make it as easy as possible to aggregate data
- Involve them in as much of the data process as possible to have them invested in its success



So, what can we do?

- Promote the concept of data preservation and long-term use in ways that they would benefit
 - “The greater science good” is fantastic, but “funding agencies want to know if your long-term plan is working” is more relevant to the professional’s needs
- Create tools to help them add value to their data
 - Graphs
 - Maps
 - Statistics
 - Reports



Rebekah D. Wallace
EDDMapS Data Coordinator
bekahwal@uga.edu

Questions?

