

Integrating Data Quality Feedback

A Data Provider's Perspective

Mare Nazaire
Rancho Santa Ana Botanic Garden
SPNHC + TDWG, Dunedin 2018

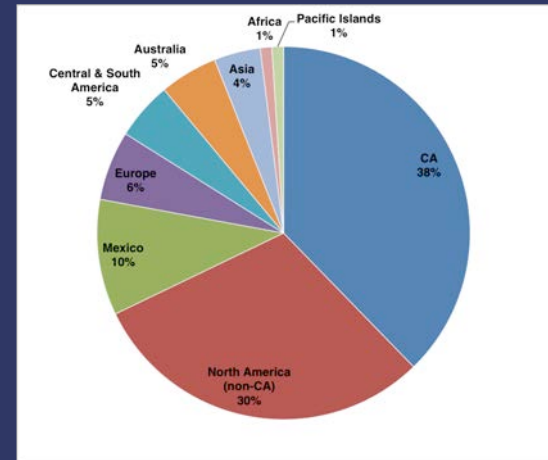


Natural history collections have always contained a wealth of data... a preserved organism and its label are a scientific specimen that has **great intrinsic value**. ... Before the advent of computers, natural history collections were physical databases from which geographic or ecological analyses and reports could be extracted by human visitation and transcription, usually a laborious and time-consuming task. ... **Computerization of label data** makes such reports on distribution and ecology of species more **readily available** to potential users; they **add value to the data**. **Interconnecting the databases brings robustness** to the information that natural history collections can provide to policy-making bodies; **appreciation of robust data will lead in turn to appreciation of the collections** from which those data were taken. Interconnectivity [facilitates] a common goal: the discovery and description of the world's biota.

- Meredith A. Lane
Roles of Natural History Collections
Ann. Missouri Bot. Gard. 83: 536-545. 1996.

RSA Herbarium

- 1.23 million specimens of vascular plants
- Combined holdings: RSA & POM College
- Curatorial staff: 5.5 FTE



RSA Herbarium

Digitizing efforts

- ~10,000 / yr manual entry
- ~12,000 / yr bulk imports
- ~12,000 / yr images

- 43% databased:
 - California
 - Selected taxa / regions
- 25% georeferenced
- 4% imaged



A Brief History of RSA's Databasing

- 1987: databasing type specimens, DBase II – limited fields
- 1993: experimentation with FMP - label making
- 1995: databasing southern CA specimens
flat files converted to FMP
- 1999: RSA & UCR merge FMP systems
- 2013: FMP crash

The screenshot displays the RSA Herbarium Inventory (Sequoia) web interface. The main record is for *Agoseris retrorsa*, collected by Mary DeDecker on May 11, 1954. The interface is divided into several sections:

- Header:** Shows the record ID (RSA-614784), family (Asteraceae), and collector (Mary DeDecker_1954-May-11_A).
- Localities:** Compares the "Old System" and "New System" localities. Both specify Inyo County, California, USA, with a primary physiographic area of Sierra Nevada. The old system locality text is "Sierra Nevada: Division Creek, Mt. Whitney Quadrangle. Growing throughout Oak grove.. Elev. 6000 Feet." The new system locality text is "Sierra Nevada: Division Creek. Growing throughout Oak grove."
- Coordinates and Elevation:** Fields for Latitude, Longitude, and Elevation (6000 FT) are visible.
- Specimen Info:** Includes collector (Mary DeDecker), collection number (204), date (May 1954), and other fields like "primary key" (18304), "CNPS listing", "State Status", "Federal Status", and "EO #".
- Actions:** A "New Inventory Record" button is present. At the bottom, there are buttons for "Georeference this locality", "Sula Export", "Naomi Export for Online DB", "Naomi's UC/JEPS Extract", "Naomi's Export for M. Wall", "Export CCH", "Export Baja", and "Specify Export".

A Brief History of RSA's Databasing

- 2014 – 2017: data migration from FMP to Specify

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
Barcode	Herbarium	Accession	Ac Living	Living Collec	Family	Genus1	Species1	Species Author1	ssp/ var author	ssp/ var author	ssp/ var author	ssp/ var author	Gen SpDs	Det	DetM	DetRe	Type	Stat	Country	State	County	County	Locality	
83	RSA	602805			Caryophyllaceae	Minuartia	nuttallii	(Pursh) Briq.		ssp. gracilis							Oct 2008	non-type	US	CA	Los Angeles	Los Angeles	San Gabriel Mountains: West Baldy	
84	RSA	604600			Caryophyllaceae	Polycarpon	depressum	Nutt.									Aug 2004	non-type	US	CA	Los Angeles	Los Angeles	Foothills of San Gabriel Mountains: Betwe	
85	RSA	606997			Caryophyllaceae	Silene	lemmonii	S. Wats.									Aug 2004	non-type	US	CA	Los Angeles	Los Angeles	San Gabriel Mountains: Northwest slope	
86	RSA	600454			Caryophyllaceae	Caryophyllaceae												non-type	US	CA	Fresno	Fresno	Southern Sierra Nevada: Rensselaer Mine	
87	RSA	613287			Caryophyllaceae	Dianthus	armeria	L.				D. L. Banks					1998	non-type	US	District of Columbia	District of Columbia	Washington D.C.	Washington D.C.	
88	RSA	602077			Caryophyllaceae	Spargula	arvensis	Linnaeus										non-type	US	CA	Los Angeles	Los Angeles	San Gabriel: Garibaldi Avenue and San Ga	
89	RSA	613505			Caryophyllaceae	Cerastium	glomeratum	Thull.				L. Moore					1998	non-type	US	CA	Monterey	Monterey	Point Lobos: Village Site	
90	RSA	603702			Caryophyllaceae	Sagina	occidentalis	S. Wats.										non-type	US	CA	Monterey	Monterey	Point Lobos: End of road just S of Hidden	
91	RSA	604234			Caryophyllaceae	Silene	stellata	(L.) Ait. f.										non-type	US	Missouri	Boone	Boone	Rollins Spring, just S of Columbia.	
92	RSA	613484			Caryophyllaceae	Cerastium	glomeratum	Thull.				D. L. Banks					1998	non-type	US	District of Columbia	District of Columbia	Washington	American University Campus.	
93	RSA	605253			Caryophyllaceae	Stellaria	pubera											non-type	US	Maryland	Montgomery	Montgomery	Minnehaha Creek: near Glenn Echo.	
94	RSA	605255			Caryophyllaceae	Agrostemma	githago	L.										non-type	US	VA	Fairfax	Fairfax	1 mile ESE of Langley: Side of Leesburg Ro	
95	RSA	613479			Caryophyllaceae	Arenaria	serpyllifolia	L.				D. L. Banks					1998	non-type	US	District of Columbia	District of Columbia	Washington [D.C.]: American University Co		
96	RSA	607106			Caryophyllaceae	Cerastium	fontanum	Baumg.		ssp. vulgare							Jul 2004	non-type	US	CA	Los Angeles	Los Angeles	San Gabriel Mountains: Rincon Ranger Sta	
97	RSA	608504			Caryophyllaceae	Silene	gallica	L.										non-type	US	CA	Los Angeles	Los Angeles	Santa Monica Mountains: Mandeville Can	
98	RSA	608647			Caryophyllaceae	Stellaria	media	(L.) Villars										non-type	US	CA	Los Angeles	Los Angeles	La Verne: ...	
99	RSA	608899			Caryophyllaceae	Minuartia	douglasii	(Fenzl ex Torr. & A. Gray) Mattf.										non-type	US	CA	Los Angeles	Los Angeles	San Gabriel Mountains: San Dimas Canyon	
100	RSA	609019			Caryophyllaceae	Silene	laciniata	Cav.		ssp. major		C.L. Hitchc. & Maguire					Jul 2004	non-type	US	CA	Los Angeles	Los Angeles	San Gabriel Mountains: Burbank Canyon.	
101	unaccessioned	-5734			Caryophyllaceae	Silene	gallica	L.										non-type	US	CA	Los Angeles	Los Angeles	San Dimas: Puddingstone.	
102	RSA	608701			Caryophyllaceae	Cerastium	fontanum	Baumg.		ssp. vulgare								1998	non-type	US	CA	Riverside	Riverside	San Jacinto Mountains: Strawberry Creek.
103	RSA	608473			Caryophyllaceae	Sagina	saginoides	(L.) Karst.		var. hesperia								1998	non-type	US	CA	Riverside	Riverside	West fork of Tahquitz Creek.
104	RSA	609974			Caryophyllaceae	Spergularia	salina	J. & K. Presl		var. salina								non-type	US	CA	Los Angeles	Los Angeles	San Pedro: South of 200 E. Signal Street.	
105	RSA	611075			Caryophyllaceae	Spergularia	bocconei											1998	non-type	US	CA	Los Angeles	Los Angeles	San Pedro: Harbor Blvd. at 4th Street
106	RSA	593905			Caryophyllaceae	Spergularia	marina	(L.) Griseb.										non-type	US	CA	Riverside	Riverside	San Jacinto Valley, ca 2 air miles NE of Win	
107	RSA	593904			Caryophyllaceae	Spergularia	marina	(L.) Griseb.										non-type	US	CA	Riverside	Riverside	San Jacinto Valley, ca 2 air miles NE of Win	
108	RSA	593902			Caryophyllaceae	Spergularia	platensis	(Camb.) Fenzl										non-type	US	CA	Riverside	Riverside	San Jacinto Valley, ca 2 air miles NE of Win	
109	RSA	593878			Caryophyllaceae	Polycarpon	tetraphyllum	(L.) L.										non-type	US	CA	Riverside	Riverside	San Jacinto Valley, ca 2 air miles NE of Win	
110	RSA	593877			Caryophyllaceae	Loeflingia	squarrosa	Nutt.										non-type	US	CA	Riverside	Riverside	San Jacinto Valley, ca 2 air miles NE of Win	
111	RSA	593699			Caryophyllaceae	Spergularia	marina	(L.) Griseb.										non-type	US	CA	Riverside	Riverside	San Jacinto Valley, San Jacinto Wildlife Ref	
112	RSA	593691			Caryophyllaceae	Spergularia	rubra	J. Presl. & C. Presl.										non-type	US	CA	Riverside	Riverside	San Jacinto Valley, San Jacinto Wildlife Ref	
113	RSA	593690			Caryophyllaceae	Loeflingia	squarrosa	Nutt.										non-type	US	CA	Riverside	Riverside	San Jacinto Valley, San Jacinto Wildlife Ref	
114	RSA	593800			Caryophyllaceae	Minuartia	pusilla	(S. Wats.) Mattf.										non-type	US	CA	Riverside	Riverside	San Jacinto Mountains, Garner Valley.	
115	RSA	593793			Caryophyllaceae	Silene	antirrhina	L.										non-type	US	CA	San Diego	San Diego	Cuyamaca Mountains, burned area at wes	
116	unaccessioned	-5890			Caryophyllaceae	Silene	rubra											non-type	US	WA	King	King	Seattle: Fort Lawton.	
117	RSA	597604			Caryophyllaceae	Arenaria	macradenia	S. Watson		var. arcuifolia								non-type	US	CA	San Bernardino	San Bernardino	San Gabriel Mountains, Baldy Mesa Ridge.	
118	unaccessioned	-6041			Caryophyllaceae	Silene		L.										non-type	US	CA	San Bernardino	San Bernardino	San Gabriel Mountains, Telegraph Peak. P	
119	RSA	611609			Caryophyllaceae	Silene	antirrhina	L.										non-type	US	CA	Los Angeles	Los Angeles	San Gabriel Mountains: Spring Hill, San Ar	
120	RSA	611634			Caryophyllaceae	Loeflingia	squarrosa	Nutt.		var. squarrosa								non-type	US	CA	Los Angeles	Los Angeles	Liebre Mountains region: Plum Canyon so	
121	RSA	611638			Caryophyllaceae	Minuartia	pusilla	(S. Wats.) Mattf.										non-type	US	CA	Los Angeles	Los Angeles	Liebre Mountains region: North base of L	
122	RSA	611641			Caryophyllaceae	Minuartia	pusilla	(S. Wats.) Mattf.										non-type	US	CA	Los Angeles	Los Angeles	Liebre Mountains region: Portal Ridge, no	
123	unaccessioned	-6190			Caryophyllaceae	Silene	lemmonii	S. Wats.										1998	non-type	US	CA	Los Angeles	Los Angeles	Devil's Canyon, Camground vicinity.

A Brief History of RSA's Databasing

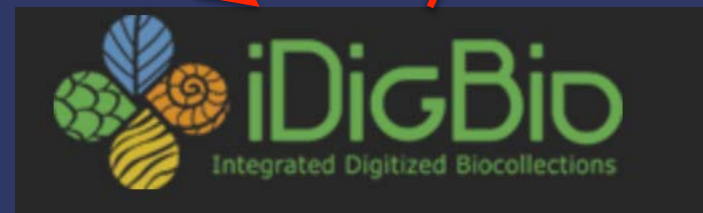
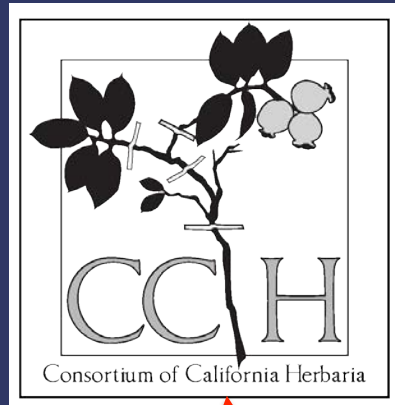
- Data migration: ~450,000 records

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
Barcode	Herbarium	Accession	Ac Living	Living Collec	Family	Genus1	Species1	Species Author1	ssp/ var author	ssp/ var subtaxon	ssp/ var author	Gen Sp/Pr. S.	Det	DetM	DetPr	Type	Statu	Country	State	County	County	County	County	County	Locality	
83	RSA	602805			Caryophyllaceae	Minuartia	nuttallii	(Pursh) Briq.		ssp. gracilis							Oct 2008	non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	San Gabriel Mountains: West Baldy	
84	RSA	604600			Caryophyllaceae	Polycarpon	depressum	Nutt.									Aug 2004	non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	Foothills of San Gabriel Mountains: Betwe	
85	RSA	606997			Caryophyllaceae	Silene	lemmonii	S. Wats.									Aug 2004	non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	San Gabriel Mountains: Northwest slope	
86	RSA	600454			Caryophyllaceae	Caryophyllaceae												non-type	US	CA	Fresno	Fresno	Fresno	Fresno	Southern Sierra Nevada: Rensselaer Mine.	
87	RSA	613287			Caryophyllaceae	Dianthus	armeria	L.									1998	non-type	US	CA	District of Columbia	District of Columbia	District of Columbia	District of Columbia	Washington D.C.	
88	RSA	602077			Caryophyllaceae	Spargula	arvensis	Linnaeus										non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	San Gabriel: Garibaldi Avenue and San Ga	
89	RSA	613505			Caryophyllaceae	Cerastium	glomeratum	Thull.									1998	non-type	US	CA	Monterey	Monterey	Monterey	Monterey	Point Lobos: Village Ste	
90	RSA	603702			Caryophyllaceae	Sagina	occidentalis	S. Wats.										non-type	US	CA	Monterey	Monterey	Monterey	Monterey	Point Lobos: End of had just S of Hidden	
91	RSA	604234			Caryophyllaceae	Silene	stellata	(L.) Ait. f.										non-type	US	Missouri	Boone	Boone	Boone	Boone	Rollins Spring, just S of Columbia.	
92	RSA	613484			Caryophyllaceae	Cerastium	glomeratum	Thull.									1998	non-type	US	District of Columbia	Washington	Washington	Washington	Washington	American University campus.	
93	RSA	605253			Caryophyllaceae	Stellaria	pubera											non-type	US	Maryland	Montgomery	Montgomery	Montgomery	Montgomery	Minnehaha Creek: near Glenn Echo.	
94	RSA	605255			Caryophyllaceae	Agrostemma	githago	L.										non-type	US	VA	Fairfax	Fairfax	Fairfax	Fairfax	1 mile ESE of Langley Side of Leesburg Ro	
95	RSA	613479			Caryophyllaceae	Arenaria	serpyllifolia	L.									1998	non-type	US	CA	District of Columbia	District of Columbia	District of Columbia	District of Columbia	Washington [D.C.]: American University Co	
96	RSA	607606			Caryophyllaceae	Cerastium	fontanum	Baumg.		ssp. vulgare							Jul 2004	non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	San Gabriel Mountains: Rincon Ranger Sta	
97	RSA	608504			Caryophyllaceae	Silene	gallica	(L.) Villars										non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	San Gabriel Mountains: Mandeville Can	
98	RSA	608647			Caryophyllaceae	Stellaria	media	(L.) Villars										non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	La Verne: ...	
99	RSA	608899			Caryophyllaceae	Minuartia	douglasii	(Fenzl ex Torr. & A. Gray) Mattf.										non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	San Dimas: San Dimas Canyon	
100	RSA	608925			Caryophyllaceae	Silene	laciniata	Cav.		ssp. major							Jul 2004	non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	San Gabriel Mountains: Burbank Canyon.	
101	RSA	608925			Caryophyllaceae	Silene	gallica	L.										non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	San Dimas: Puddingstone.	
102	RSA	608925			Caryophyllaceae	Cerastium	fontanum	Baumg.		ssp. vulgare							1998	non-type	US	CA	Riverside	Riverside	Riverside	Riverside	San Jacinto Mountains: Strawberry Creek.	
103	RSA	608473			Caryophyllaceae	Sagina	saginoides	(L.) Karst.		var. hesperia							1998	non-type	US	CA	Riverside	Riverside	Riverside	Riverside	West fork of Tahouitz Creek.	
104	RSA	609974			Caryophyllaceae	Spargularia	salina	J. & K. Presl		var. salina								non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	San Pedro: South of 200 E. Signal Street.	
105	RSA	611075			Caryophyllaceae	Spargularia	bocconei										1998	non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	San Pedro: Harbor Blvd. at 4th Street	
106	RSA	593905			Caryophyllaceae	Spargularia	marina	(L.) Griseb.										non-type	US	CA	Riverside	Riverside	Riverside	Riverside	San Jacinto Valley, ca 2 air miles NE of Win	
107	RSA	593904			Caryophyllaceae	Spargularia	marina	(L.) Griseb.										non-type	US	CA	Riverside	Riverside	Riverside	Riverside	San Jacinto Valley, ca 2 air miles NE of Win	
108	RSA	593902			Caryophyllaceae	Spargularia	platensis	(Camb.) Fenzl										non-type	US	CA	Riverside	Riverside	Riverside	Riverside	San Jacinto Valley, ca 2 air miles NE of Win	
109	RSA	593878			Caryophyllaceae	Polycarpon	tetraphyllum	(L.) L.										non-type	US	CA	Riverside	Riverside	Riverside	Riverside	San Jacinto Valley, ca 2 air miles NE of Win	
110	RSA	593877			Caryophyllaceae	Loeflingia	squarrosa	Nutt.										non-type	US	CA	Riverside	Riverside	Riverside	Riverside	San Jacinto Valley, ca 2 air miles NE of Win	
111	RSA	593699			Caryophyllaceae	Spargularia	marina	(L.) Griseb.										non-type	US	CA	Riverside	Riverside	Riverside	Riverside	San Jacinto Valley, San Jacinto Wildlife Ref	
112	RSA	593691			Caryophyllaceae	Spargularia	rubra	J. Presl. & C. Presl.										non-type	US	CA	Riverside	Riverside	Riverside	Riverside	San Jacinto Valley, San Jacinto Wildlife Ref	
113	RSA	593690			Caryophyllaceae	Loeflingia	squarrosa	Nutt.										non-type	US	CA	Riverside	Riverside	Riverside	Riverside	San Jacinto Valley, San Jacinto Wildlife Ref	
114	RSA	593800			Caryophyllaceae	Minuartia	pusilla	(S. Wats.) Mattf.										non-type	US	CA	Riverside	Riverside	Riverside	Riverside	San Jacinto Mountains, Garner Valley.	
115	RSA	593793			Caryophyllaceae	Silene	antirrhina	L.										non-type	US	CA	San Diego	San Diego	San Diego	San Diego	Cuyamaca Mountains, burned area at wes	
116	unaccessioned	5890			Caryophyllaceae	Tissa	rubra											non-type	US	WA	King	King	King	King	Seattle: Fort Lawton.	
117	RSA	597604			Caryophyllaceae	Arenaria	macradenia	S. Watson		var. arcuifolia							Jul 2004	non-type	US	CA	San Bernardino	San Bernardino	San Bernardino	San Bernardino	San Gabriel Mountains, Baldy Mesa Ridge.	
118	unaccessioned	6041			Caryophyllaceae	Silene		L.										non-type	US	CA	San Bernardino	San Bernardino	San Bernardino	San Bernardino	San Gabriel Mountains, Telegraph Peak. P	
119	RSA	611609			Caryophyllaceae	Silene	antirrhina	L.									Aug 2004	non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	San Gabriel Mountains: Spring Hill, San Ar	
120	RSA	611634			Caryophyllaceae	Loeflingia	squarrosa	Nutt.		var. squarrosa								non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	Liebre Mountains region: Plum Canyon so	
121	RSA	611638			Caryophyllaceae	Minuartia	pusilla	(S. Wats.) Mattf.										non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	Liebre Mountains region: North base of Li	
122	RSA	611641			Caryophyllaceae	Minuartia	pusilla	(S. Wats.) Mattf.										non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	Liebre Mountains region: Portal Ridge, no	
123	unaccessioned	6190			Caryophyllaceae	Silene	lemmonii	S. Wats.										1998	non-type	US	CA	Los Angeles	Los Angeles	Los Angeles	Los Angeles	Devil's Canyon, Camground vicinity.

Cleaning, formatting, standardization

- Combining / separating fields
- Removing negative accessions, hidden spaces, odd characters
- Standardizing geography, taxonomy

Publishing Data



Publishing Data

The screenshot shows the Symbiota Connectors interface. The top navigation bar includes: Welcome, Data, Trees, Reports, Interactions, Statistics, Query, Symbiota, Workbench, and BatchEdit. The left sidebar shows 'Actions' with 'New Connector' and 'Symbiota Connectors' (containing 'RSA'). The main panel displays details for the 'Symbiota Connector' named 'RSA':

- Connector Name: RSA
- Status: Data to Send to Symbiota
- Description:
- Key: 78b924c4-8b4b-4134-86fc-5f266c9dfd4c
- Export Mapping Name: Symbiota

Current Status

- Sent to Symbiota: 04/04/2018
- DB Cache Status: OK
- DB Cache Created: 06/20/2018
- Changes Since Last Send: 524252
- New or Edited Records: 524252
- Deleted Records: 0

A secondary window titled 'Sending to Symbiota' is visible in the background, showing a progress bar for 'Updating cache'.

The screenshot shows the 'Data Exporter' window with a table of mappings and their update status. The table has three columns: Mapping, Last Update, and Status.

Mapping	Last Update	Status
RSA web portal	2018-08-06 13:23:31.0	2045 out of date records
CCH	2018-08-07 09:53:51.0	Up to date
CCH Baja California	2018-07-03 12:15:18.0	4952 out of date records
Marcus E. Jones	2018-06-19 13:59:46.0	2 out of date records
Symbiota	Never	Needs to be built

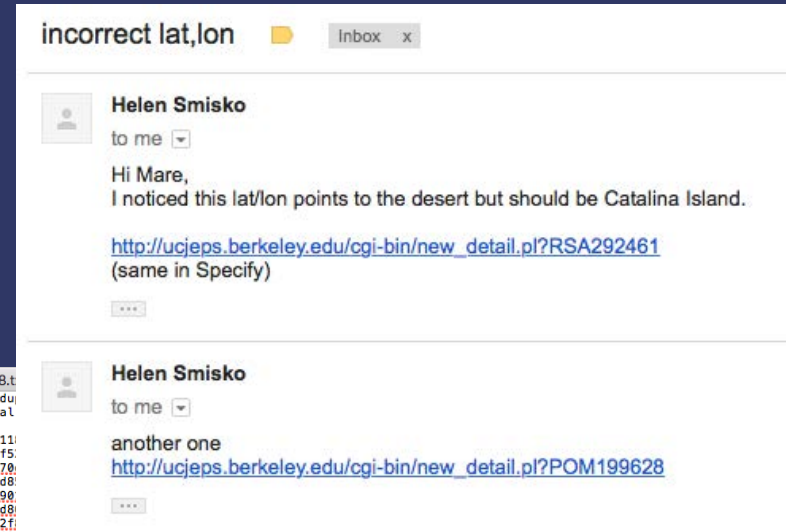
Below the table, there is a progress bar labeled 'Updating cache'. At the bottom of the window, there are several buttons: Help, Show SQL, Export, Build / Update, Export for Web Portal, and Quit.

Data Quality Feedback

Sources of feedback

- Individual user (1 – few records)
- Aggregators (large batches)
 - CCH
 - iDigBio

```
log_forMare_March2018.t
These are the records that have been skipped because they are du
the database. All of these are the 2nd instance of the original
These do not have barcode numbers
skipping: ACC: Duplicate accession number found, POM10056==>0311
skipping: ACC: Duplicate accession number found, POM10120==>28f5
skipping: ACC: Duplicate accession number found, POM102930==>070
skipping: ACC: Duplicate accession number found, POM11059==>b2e8
skipping: ACC: Duplicate accession number found, POM11959==>0890
skipping: ACC: Duplicate accession number found, POM126404==>088
skipping: ACC: Duplicate accession number found, POM127859==>a2f1
skipping: ACC: Duplicate accession number found, POM128426==>469
skipping: ACC: Duplicate accession number found, POM13688==>25ae1c7d-bbdb-4746-b2dc-cbe6656d248a
skipping: ACC: Duplicate accession number found, POM14551==>011cd666-8503-421b-bafd-48af45b66a02
skipping: ACC: Duplicate accession number found, POM145644==>1752bada-b2df-4f18-a2f8-6e2842627ff2
skipping: ACC: Duplicate accession number found, POM14635==>0b6c8b30-2d9e-4f84-b94c-e9c37b1f46d1
skipping: ACC: Duplicate accession number found, POM14669==>2e10110c-a63a-4c76-bf08-100b1f99c9f1
skipping: ACC: Duplicate accession number found, POM149409==>2d12ded6-cce0-430a-e047-d2babbfd188ac
skipping: ACC: Duplicate accession number found, POM161774==>5e754264-d028-4782-91f9-d8b23b4ff060
skipping: ACC: Duplicate accession number found, POM162077==>d0ce59c8-0d5d-4126-9565-6f20f854ead6
skipping: ACC: Duplicate accession number found, POM170269==>da620bf2-0909-438e-914d-b0c70c53b3b8
skipping: ACC: Duplicate accession number found, POM175473==>c9ada67f-4805-4c3a-abd5-d9c518dd7559
skipping: ACC: Duplicate accession number found, POM179210==>7be1564c-5489-4506-91cb-2b1a0b255a65
skipping: ACC: Duplicate accession number found, POM179767==>0978e274-9292-46e0-bc95-c09c52cb173
skipping: ACC: Duplicate accession number found, POM181034==>ea0b0e0a-7f50-495c-a4f4-1967aee1139b
skipping: ACC: Duplicate accession number found, POM181568==>7ef21d35-9a13-4750-a8fd-6031e6baf42
skipping: ACC: Duplicate accession number found, POM183493==>d216a4d4-a544-42aa-881f-ed0d0da3892c
skipping: ACC: Duplicate accession number found, POM183854==>45a2cceb-42e4-46ff-bc9f-3753b344473f
skipping: ACC: Duplicate accession number found, POM18620==>9008c7f4-c46c-42fa-8b2d-064f98f1c734
skipping: ACC: Duplicate accession number found, POM192628==>beb7873f-8615-42c5-b431-4be78182e33b
skipping: ACC: Duplicate accession number found, POM195168==>83917099-ee5f-4b05-9521-bb2aba09f021
skipping: ACC: Duplicate accession number found, POM195438==>94d0199a-fc85-4c66-9af6-c069ad0017a
skipping: ACC: Duplicate accession number found, POM203237==>e10281d9-6574-489e-812c-0914c36a9e66
skipping: ACC: Duplicate accession number found, POM213001==>1d8f35a5-b54a-4b0d-88fd-a0e0b9cfd5b1
skipping: ACC: Duplicate accession number found, POM214936==>55b35a1e-8f44-44a9-a54d-ff1868398eb8
skipping: ACC: Duplicate accession number found, POM221102==>adc2e623-6003-447c-a26f-1df6e5318f4f
skipping: ACC: Duplicate accession number found, POM24056==>5dcfd212-aead-41a7-9e1f-d399437553ba
skipping: ACC: Duplicate accession number found, POM25257==>c9dcac8c-ca89-4012-b39c-c42ac34b6293
skipping: ACC: Duplicate accession number found, POM260184==>f0a0c672-8004-445d-90f6-c2e2ca97adfe
skipping: ACC: Duplicate accession number found, POM260877==>ae7f80b3e-f67-4273-a911-f5f96e306d66
skipping: ACC: Duplicate accession number found, POM263158==>45d70c1d-94fa-4306-903a-7918b28ac71
skipping: ACC: Duplicate accession number found, POM269458==>77ace89e-48af-4b31-b5f1-91cb7463069
skipping: ACC: Duplicate accession number found, POM272868==>e81dc0d2-e6fd-415d-9912-d4ead0c7b5b0
skipping: ACC: Duplicate accession number found, POM273143==>60850782-c068-4ac1-bfb0-399f8c386611
skipping: ACC: Duplicate accession number found, POM273215==>885f2363-c17f-4315-b77e-70de77e3ce10
skipping: ACC: Duplicate accession number found, POM273861==>4ac8456a-a71c-45c7-b80c-121826ae6ae3
skipping: ACC: Duplicate accession number found, POM274585==>30b7194e-6670-421a-b12c-61c147091885
skipping: ACC: Duplicate accession number found, POM276248==>80596b6d-842f-4421-bb04-038c4085ce50
skipping: ACC: Duplicate accession number found, POM276249==>06b97800-47a0-4b7c-8fa2-5ded12fe8666
skipping: ACC: Duplicate accession number found, POM276623==>a5fce88a-71da-4315-a8e6-412aec297373
skipping: ACC: Duplicate accession number found, POM277124==>bb322977-8286-4eed-aaef-3120ee66548d
skipping: ACC: Duplicate accession number found, POM277912==>98b976cb-7a72-436a-8d99-859e92f7a99f
```



Data Quality Feedback

- Georeferencing errors
 - some can be easy fixes
 - requires re-georeferencing

RSA133434 georeference

Inbox x



Katie Gallagher

to me

Hi Mare,
I hope this email finds you well!

Quick note: I think RSA133434 might be georeferenced incorrectly. Or maybe not at all. Thought you might want to know. :)

http://ucjeps.berkeley.edu/cgi-bin/new_detail.pl?accn_num=RSA133434&YF=1



Consortium of California Herbaria

Participants News Search About Help Donate



RANCHO SANTA ANA BOTANIC GARDEN

Accession
Detail
Results

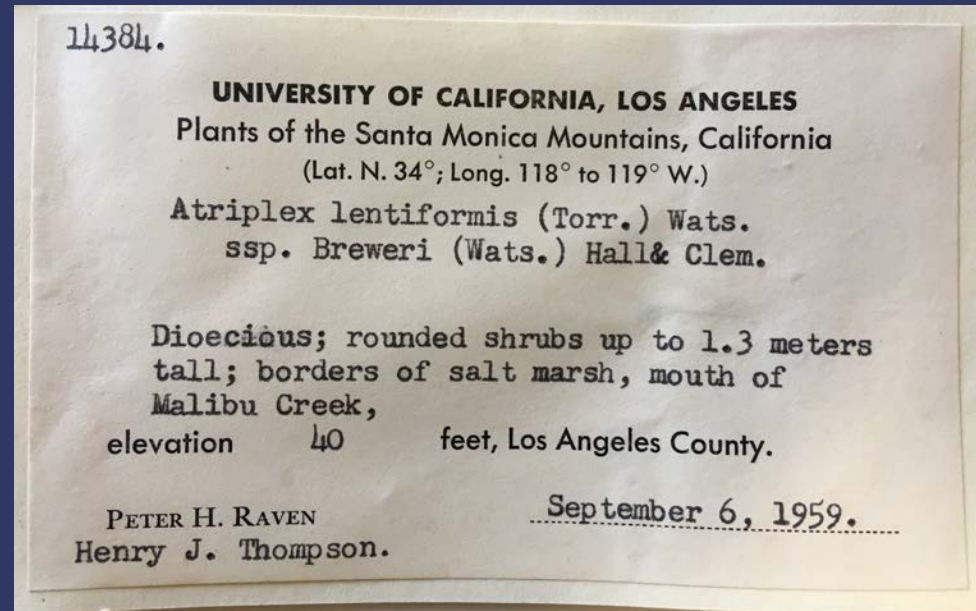
RSA is the home institution for this record.

Please cite data retrieved from this page: Data provided by the participants of the Consortium of California Herbaria (ucjeps.berkeley.edu/consortium/; Tue Jul 17 12:31:44 2018). Records are made available under the [CCH Data Use Terms](#).

Specimen number	RSA133434
Determination	<i>Atriplex lentiformis subsp. breweri</i> <small>More information: Jepson Online Interchange</small>
Collector, number, date	P. H. Raven, H. J. Thompson, 14384, 1959-9-6
Verbatim date	1959-09-06
County	Los Angeles
Locality	At mouth of Malibu Creek
Elevation	12m
Habitat	Borders of salt marsh
Coordinates	34.00000 -118.00000 <small>BarkeleyMapper [or without layers, here]</small>
Datum	not recorded
Coordinate source	Not recorded
Voucher information	<i>other label numbers</i> 50c5ecb2-cd15-4592-be5c-17cb20778fb8 <i>verbatim elevation</i> 40 ft
Annotations and/or curatorial actions	current determination (uncorrected): <i>Atriplex lentiformis subsp. breweri</i>
Notes	Dioecious; rounded shrubs up to 1.3 meters tall.;
Comment	

Data Quality Feedback

- Georeferencing errors
 - some can be easy fixes
 - requires re-georeferencing



▼ Search Results - 12						
Country	State	County	Locality	Locality continued	Latitude1	Longitude1
United States	California	Los Angeles County	Salt marsh at mouth of Malibu Creek.			
United States	California	Los Angeles County	Salt marsh at mouth of Malibu Creek.			
United States	California	Los Angeles County	Salt marsh at mouth of Malibu Creek.			
United States	California	Los Angeles County	At mouth of Malibu Creek			
United States	California	Los Angeles County	Borders of salt marsh at mouth of Malibu Creek.			
United States	California	Los Angeles County	At mouth of Malibu Creek		34.0000000000	-118.0000000000
United States	California	Los Angeles County	At mouth of Malibu Creek		34.0000000000	-118.0000000000
United States	California	Los Angeles County	At mouth of Malibu Creek,			
United States	California	Los Angeles County	At mouth of Malibu Creek.		34.0000000000	-118.1190000000
United States	California	Los Angeles County	At mouth of Malibu Creek.		34.0000000000	-118.0000000000
United States	California	Los Angeles County	At mouth of Malibu Creek.		34.0000000000	-118.0000000000
United States	California	Los Angeles County	At mouth of Malibu Creek.		34.0000000000	-118.0000000000
United States	California	Los Angeles County	At mouth of Malibu Creek.		34.0000000000	-118.0000000000

Data Quality Feedback

- Georeferencing errors
 - requires locating physical specimen

specimen record with error

Inbox x



Joy England <jengland@rsabg.org>

Mar 28



to me

Hi Mare,

I found this record for a CNPS listed species that was mapped wrong:

http://ucjeps.berkeley.edu/cgi-bin/new_detail.pl?RSA377334

Better coordinates for the locality are 34.1112, -118.3733 (error radius 1500m). Also it appears that the databaser believed the locality was on the Angeles National Forest which appears in brackets with the person's initials and date. That note should be deleted from the locality field because it is an inaccurate interpretation of the original locality description, which is not on National Forest land.

Thanks!

GEOLocate Web Application

Workbench 2 possible locations found

Georeference Options | Draw polygon Place marker Measure

Locality String: laurel canyon

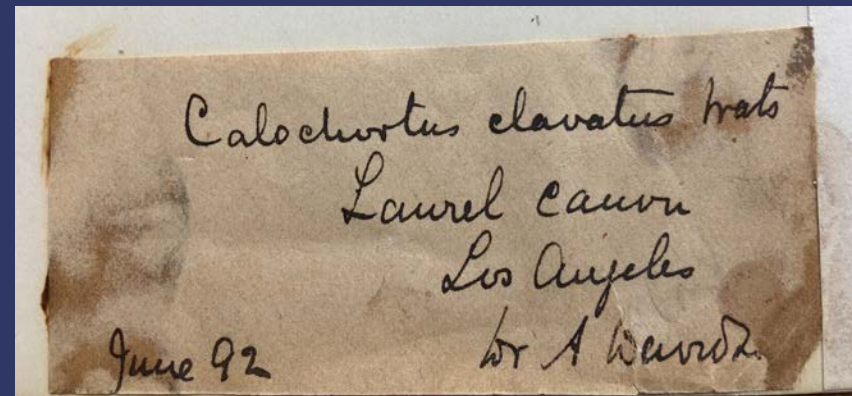
Country: UNITED STATES OF AMERICA

State: California

County: Los Angeles

latitude: 34.101675 longitude: -118.3648 uncertainty: 1506 m error polygon

34.101675 -118.3648 1506 Unavailable



Data Quality Feedback

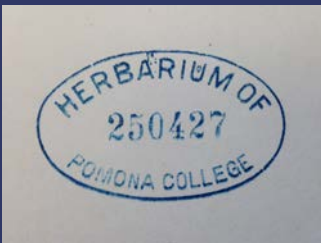
- Duplicate accessions



```
log_forMare_March2018.txt
These are the records that have been skipped because they are duplicates that have already been found in
the database. All of these are the 2nd instance of the original RSA POM accession
These do not have barcode numbers
skipping: ACC: Duplicate accession number found, POM10056==>03118c00-30a5-4336-807f-c383b46144ab
skipping: ACC: Duplicate accession number found, POM10120==>28f52e93-2c01-4c94-80f6-40874fa31c52
skipping: ACC: Duplicate accession number found, POM102930==>070eb4de-6470-4347-ab7b-71ec47a3b149
skipping: ACC: Duplicate accession number found, POM11059==>b2d85817-b529-4074-9071-a9a58c876f58
skipping: ACC: Duplicate accession number found, POM11059==>d89019e7-4716-4568-bb54-8ec6edcafa0a
skipping: ACC: Duplicate accession number found, POM126404==>0d86d602-66a6-44fc-a832-56d5144ef697
skipping: ACC: Duplicate accession number found, POM127859==>a2f8f7f6-b34d-4ff6-b40c-6e63d43c02d9
skipping: ACC: Duplicate accession number found, POM128426==>4693c987-6e8c-464c-b4f5-bf24bdfaeae4
skipping: ACC: Duplicate accession number found, POM13688==>25ae1c7d-bbdb-4746-b2dc-cbe6656d248a
skipping: ACC: Duplicate accession number found, POM14551==>911cd6e6-8503-421b-bafd-48af45b66a92
skipping: ACC: Duplicate accession number found, POM145644==>1752bada-b2df-4f18-a2f8-6e2842627ff2
skipping: ACC: Duplicate accession number found, POM14635==>0b6c8b10-2dde-4f84-b84c-e9c37b1f4dd1
skipping: ACC: Duplicate accession number found, POM14669==>2ef0110c-e63a-4d76-8fd8-100bff99bc9f
skipping: ACC: Duplicate accession number found, POM149409==>2d12ded6-cce0-430a-a047-d2babfd188ac
skipping: ACC: Duplicate accession number found, POM161774==>5e754264-d028-4782-91f9-d8b23b4ff6d0
skipping: ACC: Duplicate accession number found, POM162077==>d0ce59c8-0d5d-4126-9565-6f20f854ead6
skipping: ACC: Duplicate accession number found, POM170269==>da620bf2-0909-438e-914d-b0c70c53b3b8
skipping: ACC: Duplicate accession number found, POM175473==>c9ada67f-4805-4c3a-abd5-d9c518dd7559
skipping: ACC: Duplicate accession number found, POM179210==>7be1564c-5489-4506-91cb-2b1a0b255a65
skipping: ACC: Duplicate accession number found, POM179767==>0979e274-9292-46a0-bc85-cc09c52cbf73
skipping: ACC: Duplicate accession number found, POM181034==>eab0e60a-7f5e-450c-a4f4-19671eeff139b
skipping: ACC: Duplicate accession number found, POM181568==>7ef21d35-9a13-4750-a8fd-6031e6bafa42
skipping: ACC: Duplicate accession number found, POM183493==>d216a4d4-a544-42aa-881f-ed0d0da3892c
skipping: ACC: Duplicate accession number found, POM183854==>45a2cceb-42e4-46ff-bc9f-3753b344473f
skipping: ACC: Duplicate accession number found, POM18620==>9008c7f4-c46c-42fa-8b2d-064f98f1c734
skipping: ACC: Duplicate accession number found, POM192628==>beb7873f-8615-42c5-b431-4be78182e33b
skipping: ACC: Duplicate accession number found, POM195168==>83817099-ee5f-4b05-9521-bb2aba09f821
skipping: ACC: Duplicate accession number found, POM195438==>94d81e9a-fc85-4c66-9bf6-cd69add0077a
skipping: ACC: Duplicate accession number found, POM203237==>ef0281d9-6574-489e-812e-a914c36a59e6
skipping: ACC: Duplicate accession number found, POM213001==>1d8f35a5-b54a-4b0d-88fd-a0e0b9cfd5b1
skipping: ACC: Duplicate accession number found, POM214936==>55b35a1e-8f44-44a9-a54d-fff868398eb8
skipping: ACC: Duplicate accession number found, POM221102==>adc2e623-6003-447c-a26f-1df6e5318f4f
skipping: ACC: Duplicate accession number found, POM24056==>5dcdfd212-ead-41a7-9e1f-d399437553ba
skipping: ACC: Duplicate accession number found, POM255257==>c9dcaec8-ca89-4012-b39d-c42ac34b6293
skipping: ACC: Duplicate accession number found, POM260184==>f0a0ce72-8804-445d-90f6-e2ede9a97adfe
skipping: ACC: Duplicate accession number found, POM260877==>ad760ab3-ef67-4273-af91-5f596e30dd65
skipping: ACC: Duplicate accession number found, POM263158==>45d70c1d-94fa-43d6-903a-7918b828ac71
skipping: ACC: Duplicate accession number found, POM269458==>77ace89e-48af-4b31-b5f1-91c0b7463069
skipping: ACC: Duplicate accession number found, POM272868==>e81dc0d2-e6fd-415d-9912-d4ead0c7b5b0
skipping: ACC: Duplicate accession number found, POM273143==>60850782-c068-4ac1-bfb0-399f8c386661
skipping: ACC: Duplicate accession number found, POM273215==>885f2363-c17f-4315-b77e-70de77e3cef0
skipping: ACC: Duplicate accession number found, POM273861==>4ac8456a-a7fc-45c7-b802-121826ae6ae3
skipping: ACC: Duplicate accession number found, POM274585==>3bb7194d-c670-4721-b12c-61c47a9a0c85
skipping: ACC: Duplicate accession number found, POM276248==>80506b6d-842f-44e1-b8b4-938c4885ce50
skipping: ACC: Duplicate accession number found, POM276249==>06b97800-47a0-4b7c-8fa2-5ded12f5e866
skipping: ACC: Duplicate accession number found, POM276623==>a5fce88a-71da-4315-a8e6-412aec297373
skipping: ACC: Duplicate accession number found, POM277124==>bb322977-8286-4eed-aaef-3120ee66548d
skipping: ACC: Duplicate accession number found, POM277912==>98b976cb-7a72-436a-a0d9-859e9217a09f
```

Data Quality Feedback

- Duplicate accessions
 - replace barcode as identifier



```
These are the records that the parser has switched to using the barcode instead of the original RSA/POM accession
logging: ACC: Barcode field not NULL, using barcode for ID 817804==>RSA0036212<
logging: ACC: Barcode field not NULL, using barcode for ID POM1000298==>RSA0076363<
logging: ACC: Barcode field not NULL, using barcode for ID POM100168==>RSA0003896<
logging: ACC: Barcode field not NULL, using barcode for ID POM100169==>RSA0003976<
logging: ACC: Barcode field not NULL, using barcode for ID POM100315==>RSA0000492<
logging: ACC: Barcode field not NULL, using barcode for ID POM100316==>RSA0000491<
logging: ACC: Barcode field not NULL, using barcode for ID POM10034==>RSA0081600<
logging: ACC: Barcode field not NULL, using barcode for ID POM10047==>RSA0066543<
logging: ACC: Barcode field not NULL, using barcode for ID POM1005==>RSA0112486<
logging: ACC: Barcode field not NULL, using barcode for ID POM100674==>RSA0029864<
logging: ACC: Barcode field not NULL, using barcode for ID POM100675==>RSA0029866<
logging: ACC: Barcode field not NULL, using barcode for ID POM10067==>RSA0081597<
logging: ACC: Barcode field not NULL, using barcode for ID POM100691==>RSA0008946<
logging: ACC: Barcode field not NULL, using barcode for ID POM10070==>RSA0114414<
logging: ACC: Barcode field not NULL, using barcode for ID POM100719==>RSA0102439<
logging: ACC: Barcode field not NULL, using barcode for ID POM10074==>RSA0016428<
logging: ACC: Barcode field not NULL, using barcode for ID POM100816==>RSA0013865<
logging: ACC: Barcode field not NULL, using barcode for ID POM100818==>RSA0009430<
logging: ACC: Barcode field not NULL, using barcode for ID POM100819==>RSA0009431<
logging: ACC: Barcode field not NULL, using barcode for ID POM100820==>RSA0009447<
logging: ACC: Barcode field not NULL, using barcode for ID POM100821==>RSA0009448<
logging: ACC: Barcode field not NULL, using barcode for ID POM10082==>RSA0016946<
logging: ACC: Barcode field not NULL, using barcode for ID POM10084==>RSA0016907<
logging: ACC: Barcode field not NULL, using barcode for ID POM10090==>RSA0076245<
logging: ACC: Barcode field not NULL, using barcode for ID POM10091A==>RSA0076246<
logging: ACC: Barcode field not NULL, using barcode for ID POM10093==>RSA0008280<
logging: ACC: Barcode field not NULL, using barcode for ID POM1009==>RSA0112487<
logging: ACC: Barcode field not NULL, using barcode for ID POM100==>RSA0008293<
logging: ACC: Barcode field not NULL, using barcode for ID POM10115==>RSA0091276<
logging: ACC: Barcode field not NULL, using barcode for ID POM101185==>RSA0097662<
logging: ACC: Barcode field not NULL, using barcode for ID POM101244==>RSA0041047<
logging: ACC: Barcode field not NULL, using barcode for ID POM10131==>RSA0078861<
logging: ACC: Barcode field not NULL, using barcode for ID POM101426==>RSA0045086<
logging: ACC: Barcode field not NULL, using barcode for ID POM101427==>RSA0057703<
logging: ACC: Barcode field not NULL, using barcode for ID POM101430==>RSA0057701<
logging: ACC: Barcode field not NULL, using barcode for ID POM101444==>RSA0045087<
logging: ACC: Barcode field not NULL, using barcode for ID POM101448==>RSA0057639<
```


Data Quality Feedback

- Discrepancies in elevation
 - requires locating the physical specimen

```
logging: ELEV: Sierra 9901 ft. (3018 m.) greater than Sierra maximum (8850 ft.): discrepancy=
1051 RSA78542
logging: ELEV: Sierra discrepancy= 1051 greater than 500 ft, elevation changed to NULL RSA78542

logging: ELEV: Siskiyou 14927 ft. (4550 m.) greater than Siskiyou maximum (14170 ft.):
discrepancy= 757 POM44448
logging: ELEV: Siskiyou 17716 ft. (5400 m.) greater than Siskiyou maximum (14170 ft.):
discrepancy= 3546 RSA790239
logging: ELEV: Siskiyou 21850 ft. (6660 m.) greater than Siskiyou maximum (14170 ft.):
discrepancy= 7680 RSA789814
logging: ELEV: Siskiyou 21850 ft. (6660 m.) greater than Siskiyou maximum (14170 ft.):
discrepancy= 7680 RSA789826
logging: ELEV: Siskiyou discrepancy= 3546 greater than 500 ft, elevation changed to NULL
RSA790239
logging: ELEV: Siskiyou discrepancy= 757 greater than 500 ft, elevation changed to NULL
POM44448
logging: ELEV: Siskiyou discrepancy= 7680 greater than 500 ft, elevation changed to NULL
RSA789814
logging: ELEV: Siskiyou discrepancy= 7680 greater than 500 ft, elevation changed to NULL
RSA789826
```

Data Quality Feedback

iDigBio

Rancho Santa Ana Botanic Garden Herbarium

Specimen Records: 517,391

Media Records: 0

iDigBio Last Ingested Date: 2018-04-23



The combined Herbarium of Rancho Santa Ana Botanic Garden and Pomona College (RSA-POM) is a museum-quality collection of vascular plant specimens. With current holdings totaling over 1,200,000 specimens, the Herbarium is the third largest in California. The Herbarium is recognized throughout the world for its strength in documenting the diversity, distribution, variation, and ecology of more than 6500 species of flowering plants, conifers, and ferns in California, which constitutes nearly 50% of the total collection. The holdings from Southern California exceed 250,000 and are unsurpassed by any other herbarium.

Approximately 95% of the collection is composed of mounted sheets and filed according to a standardized system of classification. Ancillary collections that augment the collection include a cone & fruit collection, wood collection, fluid preserved collection, and pollen and anatomy slide collection.

Contacts

Name	<i>none</i>	Name	<i>none</i>
Role	<i>none</i>	Role	<i>none</i>
Email	seinetadmin@asu.edu	Email	seinetadmin@asu.edu

[Data Corrected](#) [Data Use](#) [Raw](#)


























This table shows any data corrections that were performed on this recordset to improve the capabilities of iDigBio Search. The first column represents the correction performed. The last two columns represent the number and percentage of records that were corrected. A complete list of the data quality flags and their descriptions can be found [here](#). Clicking on a data flag name will take you to a search for all records with this flag in this recordset.

Flag	Records With This Flag	(%) Percent With This Flag
idigbio_isocountrycode_added ⓘ	516413	99.811
dwc_continent_added ⓘ	516217	99.773
dwc_datasetid_added ⓘ	368812	71.283
dwc_parentnameusageid_added ⓘ	368812	71.283
dwc_taxonomicstatus_added ⓘ	368812	71.283
gbif_canonicalname_added ⓘ	368812	71.283
gbif_genericname_added ⓘ	368812	71.283
gbif_taxon_corrected ⓘ	368812	71.283
dwc_taxonid_replaced ⓘ	368445	71.212
dwc_phylum_replaced ⓘ	366679	70.871
dwc_class_added ⓘ	358690	69.327
gbif_vernacularname_added ⓘ	345255	66.73
dwc_taxonrank_added ⓘ	321940	62.224
dwc_multimedia_added ⓘ	284638	55.014
dwc_scientificnameauthorship_replaced ⓘ	273352	52.833
gbif_reference_added ⓘ	268652	51.924
taxon_match_failed ⓘ	152639	29.502
geopoint_datum_missing ⓘ	102701	19.85
dwc_originalnameusageid_added ⓘ	76138	14.716
dwc_taxonrank_replaced ⓘ	58371	11.282
dwc_specificepithet_replaced ⓘ	35976	6.953
dwc_taxonremarks_added ⓘ	26985	5.216
dwc_genus_replaced ⓘ	19394	3.748
geopoint_low_precision ⓘ	13078	2.528
dwc_infraspecificepithet_added ⓘ	11028	2.131

Data Quality Feedback

iDigBio

- formatting

Flag	Records With This Flag	(%) Percent With This Flag
idigbio_isocountrycode_added 	516413	99.811
dwc_continent_added 	516217	99.773
dwc_datasetid_added 	368812	71.283
dwc_parentnameusageid_added 	368812	71.283
dwc_taxonomicstatus_added 	368812	71.283
gbif_canonicalname_added 	368812	71.283
gbif_genericname_added 	368812	71.283
gbif_taxon_corrected 	368812	71.283
dwc_taxonid_replaced 	368445	71.212
dwc_phylum_replaced 	366679	70.871
dwc_class_added 	358690	69.327
gbif_vernacularname_added 	345255	66.73
dwc_taxonrank_added 	321940	62.224
dwc_multimedia_added 	284638	55.014
dwc_scientificnameauthorship_replaced 	273352	52.833
gbif_reference_added 	268652	51.924
taxon_match_failed 	152639	29.502
geopoint_datum_missing 	102701	19.85
dwc_originalnameusageid_added 	76138	14.716
dwc_taxonrank_replaced 	58371	11.282
dwc_specificepithet_replaced 	35976	6.953
dwc_taxonremarks_added 	26985	5.216
dwc_genus_replaced 	19394	3.748
geopoint_low_precision 	13078	2.528
dwc_infraspecificepithet_added 	11028	2.131

Data Quality Feedback

iDigBio

- formatting
- may require Specify tech

Flag	Records With This Flag	(%) Percent With This Flag
idigbio_isocountrycode_added	516413	99.811
dwc_continent_added	516217	99.773
dwc_datasetid_added	368812	71.283
dwc_parentnameusageid_added	368812	71.283
dwc_taxonomicstatus_added	368812	71.283
gbif_canonicalname_added	368812	71.283
gbif_genericname_added	368812	71.283
gbif_taxon_corrected	368812	71.283
dwc_taxonid_replaced	368445	71.212
dwc_phylum_replaced	366679	70.871
dwc_class_added	358690	69.327
gbif_vernacularname_added	345255	66.73
dwc_taxonrank_added	321940	62.224
dwc_multimedia_added	284638	55.014
dwc_scientificnameauthorship_replaced	273352	52.833
gbif_reference_added	268652	51.924
taxon_match_failed	152639	29.502
geopoint_datum_missing	102701	19.85
dwc_originalnameusageid_added	76138	14.716
dwc_taxonrank_replaced	58371	11.282
dwc_specificepithet_replaced	35976	6.953
dwc_taxonremarks_added	26985	5.216
dwc_genus_replaced	19394	3.748
geopoint_low_precision	13078	2.528
dwc_infraspecificepithet_added	11028	2.131

Data Quality Feedback

iDigBio

- formatting
- may require Specify tech
- still need explanation

Flag	Records With This Flag	(%) Percent With This Flag
idigbio_isocountrycode_added	516413	99.811
dwc_continent_added	516217	99.773
dwc_datasetid_added	368812	71.283
dwc_parentnameusageid_added	368812	71.283
dwc_taxonomicstatus_added	368812	71.283
gbif_canonicalname_added	368812	71.283
gbif_genericname_added	368812	71.283
gbif_taxon_corrected	368812	71.283
dwc_taxonid_replaced	368445	71.212
dwc_phylum_replaced	366679	70.871
dwc_class_added	358690	69.327
gbif_vernacularname_added	345255	66.73
dwc_taxonrank_added	321940	62.224
dwc_multimedia_added	284638	55.014
dwc_scientificnameauthorship_replaced	273352	52.833
gbif_reference_added	268652	51.924
taxon_match_failed	152639	29.502
geopoint_datum_missing	102701	19.85
dwc_originalnameusageid_added	76138	14.716
dwc_taxonrank_replaced	58371	11.282
dwc_specificepithet_replaced	35976	6.953
dwc_taxonremarks_added	26985	5.216
dwc_genus_replaced	19394	3.748
geopoint_low_precision	13078	2.528
dwc_infraspecificepithet_added	11028	2.131

Data Quality Feedback

iDigBio

- formatting
- may require Specify tech
- still need explanation
- worth correcting?

Flag	Records With This Flag	(%) Percent With This Flag
idigbio_isocountrycode_added	516413	99.811
dwc_continent_added	516217	99.773
dwc_datasetid_added	368812	71.283
dwc_parentnameusageid_added	368812	71.283
dwc_taxonomicstatus_added	368812	71.283
gbif_canonicalname_added	368812	71.283
gbif_genericname_added	368812	71.283
gbif_taxon_corrected	368812	71.283
dwc_taxonid_replaced	368445	71.212
dwc_phylum_replaced	366679	70.871
dwc_class_added	358690	69.327
gbif_vernacularname_added	345255	66.73
dwc_taxonrank_added	321940	62.224
dwc_multimedia_added	284638	55.014
dwc_scientificnameauthorship_replaced	273352	52.833
gbif_reference_added	268652	51.924
taxon_match_failed	152639	29.502
geopoint_datum_missing	102701	19.85
dwc_originalnameusageid_added	76138	14.716
dwc_taxonrank_replaced	58371	11.282
dwc_specificepithet_replaced	35976	6.953
dwc_taxonremarks_added	26985	5.216
dwc_genus_replaced	19394	3.748
geopoint_low_precision	13078	2.528
dwc_infraspecificepithet_added	11028	2.131

Data Quality Feedback

When the data quality feedback you're receiving is wrong:

- Example – rev_geocode_flip_both_sign



Data Quality Feedback


When the data quality feedback you're receiving is wrong:

- Example – rev_geocode_flip_both_sign

Specimen Record
Plantae > Magnoliophyta > Lamiales > Acanthaceae

Barleria lancifolia T.Anderson
From Rancho Santa Ana Botanic Garden Herbarium

Continent	Africa	Institution Code	Rsa
Country	South Africa	Collection Code	Rsa-pom
Locality	The Transvaal. Soutpansberg District, North Of The Soutpansberg, On The Farm Davenham. Valley Running Past House.	Catalog Number	Rsa0044463
Latitude	-29	Collected By	K. Balkwill, M-j Balkwill
Longitude	22	Date Collected	1991-05-08



click or hover to wake

Leaflet | Map data © OpenStreetMap contributors

Lat: -29, Long: 22



Lat: -22, Long: 29

South Africa. The Transvaal. Soutpansberg District, North of the Soutpansberg.

Data Quality Feedback

When the data quality feedback you're receiving is wrong:

- Example – rev_geocode_flip_both_sign

Family	Scientific Name	Date Collected	Country	Institution Code	Basis of Record	Columns
Acanthaceae	Barleria lancifolia	1991-05-08	South Africa	RSA	PreservedSpecimen	view
Convolvulaceae	Dichondra occidentalis	1990-03-10	Mexico	RSA	PreservedSpecimen	view
Acanthaceae	Barleria heterotricha	1991-05-08	South Africa	RSA	PreservedSpecimen	view
Acanthaceae	Barleria galpinii	1995-04-29	South Africa	RSA	PreservedSpecimen	view
Acanthaceae	Barleria bremekampii	1995-04-29	South Africa	RSA	PreservedSpecimen	view

- Only 1 of 5 records flagged indicated an actual error

Other Issues with Data Quality

CCH

- georeferencing and re-ingesting coordinates back into home institution database



Consortium of California Herbaria

[Participants](#) [News](#) [Search](#) [About](#) [Help](#)

Use this page to provide coordinates for records that lack them

THE BERKELEYMAPPER INTERFACE AND FUNCTIONALITY HAVE CHANGED. More modifications are expected. The circle drawing tool and the line drawing tool are now in the top center of the displayed map.
This is what needs to be copied from the circle tool, for example: "Coordinate: 36.94680 / -120.53314 (500 meters radius)" You can **triple-click** to select the line.
Select the map type at the upper right of the map.
Display the individual record markers by deselecting the Marker clusterer.
Control the zoom with the slider or the box-drawing magnifier below the slider.
Comments/questions to dbaxter at berkeley.edu

NOTA BENE: Try to choose records that have error radius estimates for cloning.

<p>Scientific Name</p> <input type="text"/>	<p>Source (1 or more; default is all sources)</p> <div style="border: 1px solid gray; padding: 2px;"><p>All sources</p><p>NY (New York Botanical Garden)</p><p>A, AMES, ECON, GH (Harvard University Herbaria)</p><p>UC (University Herbarium, UC Berkeley)</p><p>JEPS (Jepson Herbarium, UC Berkeley)</p><p>UCSB (UC Santa Barbara)</p><p>UCSC (UC Santa Cruz)</p><p>IRVC (UC Irvine)</p><p>UCR (UC Riverside)</p><p>DAV (UC Davis)</p><p>SBBC (Santa Barbara Botanic Garden)</p><p>CHSC (Chico State Herbarium, CSU Chico)</p><p>RSA (Rancho Santa Ana Botanic Garden Herbarium)</p><p>POM (Pomona Herbarium in RSA)</p></div>
<p>County (1 or more; default is all counties)</p> <div style="border: 1px solid gray; padding: 2px;"><p>All</p><p>Alameda</p><p>Alpine</p><p>Amador</p><p>Butte</p><p>Calaveras</p><p>Colusa</p><p>Contra Costa</p><p>Del Norte</p><p>El Dorado</p><p>Fresno</p><p>Glenn</p><p>Humboldt</p><p>Imperial</p></div>	<p>Collector (last name only)</p> <input type="text"/>
<p>Geographic Locality</p> <input type="text"/>	<p>Collection Number (numerical part only)</p> <input type="text"/>

When multiple words are searched for, only records that contain both words (in any order) will be returned. Words may be truncated. Wild card characters are not used in searches.

Only return records of *target taxa* that lack coordinates (may return a superabundance of records)

Return records of *all taxa* that lack coordinates

[Help with cloning coordinates](#)
[Help with adding coordinates](#)
Look up and map locations using [GeoLocate](#)
Get coordinates from Google Maps: [Touch maps](#)
Store and share geographic insights: [WikiMapia](#)

Please cite data retrieved from this page: *Data provided by the participants of the Consortium of California Herbaria.*

Problems, comments? Contact [dbaxter at berkeley.edu](mailto:dbaxter@berkeley.edu)
(Modified Aug 30 2012)

Data Quality Feedback

Steps involved to correct errors (often with individual records)

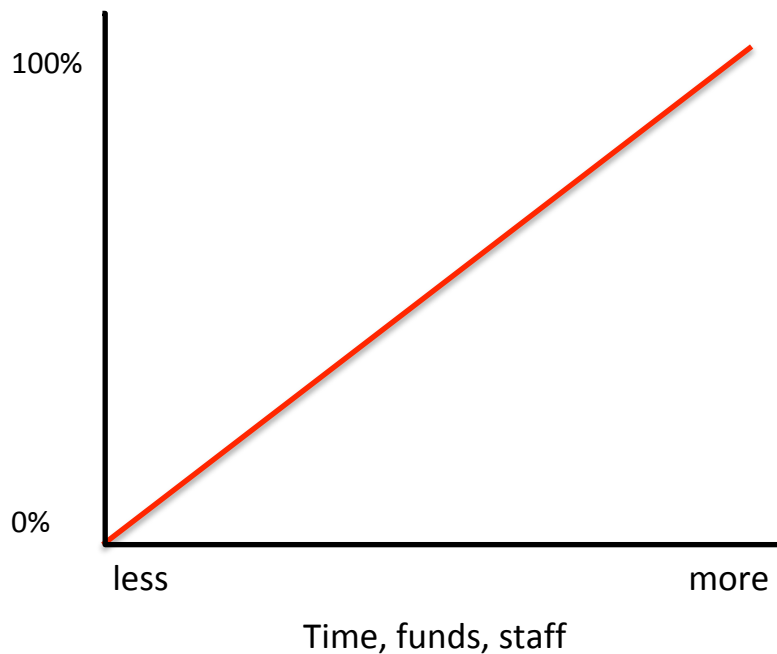
- Review specimen record in database
- Verify the source of the error: Is it a data entry error or an error on label?
 - Often entails pulling the specimen from the collection
- Check feedback against other sources
 - Georeferencing (GEOLocate, Google Earth)
 - Taxonomy (Tropicos, TNRS, IPNI)
- Make the correction **IF** the feedback is correct

Total time involved: 5 – 30 minutes / specimen

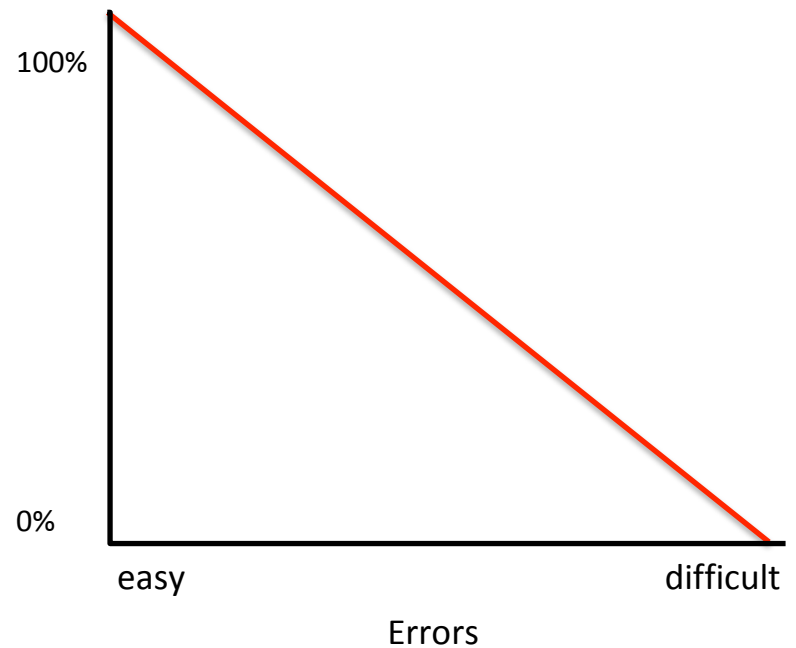
Data Quality Feedback

% chance of correcting feedback

Availability of resources



Level of difficulty



Challenges & Issues

- Lack of resources:
 - time
 - funds
 - staff
- Determinations: require expertise
- Specimen is missing!



Solutions to Improving Data Quality

Minimize errors:

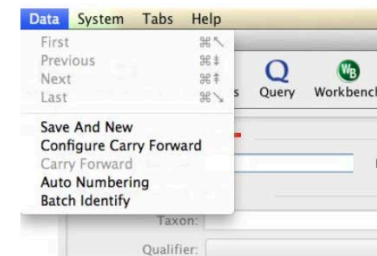
- Detailed data entry protocols
- Data entry training sessions
- QC checks for first few weeks
- Monthly QC checks on all data entry

DQ Feedback:

- Group errors into type
- Designated time to address errors

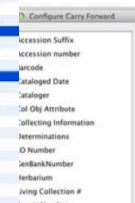
Entering New Records in Specify

1. On the task bar, click on the Data icon.
2. On the side bar, click **Collection Object** – this opens a form for entering collection data (often referred to as the “workspace”).
3. Before entering a record go to **Data Menu** (“Data” on the menu bar) and make sure the **Auto Numbering** option is unchecked. In the same **Data Menu** select the option of **Save and New** – this will allow you to go directly to a new empty form for adding a new record after saving the last one (in contrast to only saving and then having to click the + (add) button for creating a new collection object).



Barcode	Herbarium	Accession number	Accession Suffix	Status	Family	Genus	Species	Subspecies	Variety	Qualifier	Position	Current
RSAO121112	POM	293529		Current	Alismaceae	Alisma	gramineum					true
RSAO135469	RSA	86790		Current	Cartaceae	Opuntia	basilaris					true
RSAO1211094	POM	293544		Current	Fabaceae	Lupinus	bicolor					true
RSAO1210996	POM	293234		Current	Anacardiaceae	Rhus	integrifolia					true
RSAO140233	RSA	868481		Current	Fagaceae	Quercus	berberidifolia					true
RSAO140232	RSA	868480		Current	Fagaceae	Quercus	agrifolia		agrifolia			true
RSAO1211095	POM	293234		Current	Fagaceae	Quercus	agrifolia		agrifolia			true
RSAO140238	RSA	868486		Current	Rhamnaceae	Ziziphus	durata					true
RSAO140231	RSA	868479		Current	Fagaceae	Quercus	durata		gabrielensis			true
RSAO140230	RSA	868478		Current	Ranunculaceae	Ranunculus	hystriculus					true
RSAO1211098	POM	293234		Current	Utriculariaceae	Utricularia	vulgaris					true
RSAO1211098	POM	293234		Current	Lentibulariaceae	Utricularia	vulgaris					true
RSAO1211098	POM	293234		Current	Potamogetonaceae	Potamogeton	gramineus					true
RSAO121105	POM	293403		Current	Potamogetonaceae	Potamogeton	gramineus					true
RSAO121104	POM	293399		Current	Potamogetonaceae	Potamogeton	gramineus					true
RSAO121103	POM	293578		Current	Phrymaceae	Mimulus	guttatus					true
RSAO121102	POM	294075		Current	Onagraceae	Circaea	alpina		alpina			true
RSAO121100	POM	294785		Current	Fabaceae	Vicia	americana		americana			true
RSAO121098	POM	296330		Current	Poaceae	Deschampsia	cespitosa					true
RSAO126832	POM	294380		Current	Primulaceae	Lysimachia	oblongifolia					true
RSAO126830	POM	293267		Current	Hydrocharitaceae	Elodea	nuttallii					true
RSAO126827	POM	296026		Current	Polygonaceae	Persicaria	punctata					true
RSAO121113	POM	293531		Current	Alismaceae	Sagittaria	cuneata					true
RSAO121107	POM	296607		Current	Balsaminaceae	Impatiens	scalcarata					true
RSAO140235	RSA	868485		Current	Plantaginaceae	Micrantonon						true
RSAO126829	POM	293272		Current	Hydrocharitaceae	Elodea	canadensis					true
RSAO140239	RSA	868487		Current	Pinaceae	Picea						true
RSAO145409	POM	316971		Current	Cactaceae	Mammillaria						true
RSAO121116	POM	293956		Current	Asteraceae	Caillardiella	aristata					true
RSAO140236	RSA	868484		Current	Adoxaceae	Viburnum	edule					true
RSAO135651	RSA	867192		Current	Cactaceae	Coryphantha	vivipara		deserti			true
RSAO121093	POM	293502		Current	Fabaceae	Lupinus	bicolor					true
RSAO140237	RSA	868485		Current	Bignoniaceae	Tecoma	stans		angustatum			true
RSAO145406	POM	316968		Current	Cactaceae	Mammillaria						true
RSAO145407	POM	316970		Current	Cactaceae	Cereus						true
RSAO140234	RSA	868482		Current	Fagaceae	Quercus						true
RSAO135493	RSA	867191		Current	Cactaceae	Echinocereus						true
RSAO145411	POM	309552		Current	Cactaceae	Cereus						true
RSAO145410	POM	309552		Current	Cactaceae	Cereus						true
RSAO121097	POM	295540		Current	Fabaceae	Hosackia	oblongifolia		oblongifolia			true
RSAO145408	POM	316972		Current	Utriculariaceae	Utricularia	oblongifolia					true
RSAO121109	POM	293967		Current	Asteraceae	Senecio	trianthiflorus					true
RSAO121115	POM	293724		Current	Ericaceae	Phyllodoce	glandulifera					true
RSAO121114	POM	293375		Current	Saxifragaceae	Micranthes	lyalli					true
RSAO126828	POM	293556		Current	Ombrobrassicaceae	Castilleja	occidentalis					true
RSAO126826	POM	307793		Current	Lycopodiaceae	Diphasiastrum	alpinum					true
RSAO121111	POM	294627		Current	Liliaceae	Calochortus	apiculatus					true
RSAO121101	POM	293262		Current	Orchidaceae	Platanthera	stricta					true
RSAO126831	POM	293233		Current	Salicaceae	Salix	arctica					true
RSAO121099	POM	294636		Current	Amaryllidaceae	Allium	cernuum					true
RSAO126835	POM	293635		Current	Amaryllidaceae	Allium	cernuum					true
RSAO121095	POM	293522		Current	Fabaceae	Lupinus	polypophys		burkei			true

In Specify 6 you that might be checked:



Solutions to Improving Data Quality



Barcoding Bonanza with Cal Poly Pomona students

Solutions to Improving Data Quality



Acknowledgements

RSA curatorial staff: Erin Berkowitz,
Sierra Sutton, Lauren Tucker, Christina
Varnava, Grace Rice, LeRoy Gross,
Carrie Baldwin, Brandy Watts
Deb Paul, iDigBio
Joanna McCaffrey, iDigBio
Gil Nelson, iDigBio
Theresa Miller, Specify
Ed Gilbert, Symbiota

