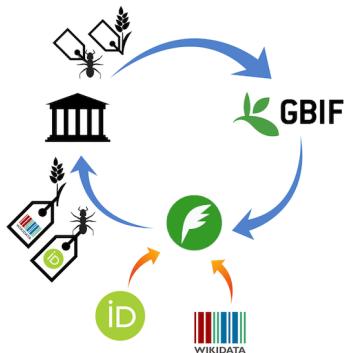


Workshop: People Identifiers

Roundtrip to Collection Management Systems



Wednesday, December 6, 2023

08:00 – 12:00 EST /
14:00 – 18:00 CET /
13:00 – 17:00 UTC

[Time Zone Converter](#)

Purpose

Put the disambiguation of people's names into practice with a particular focus on collections management systems. The purpose is for you to share how you or your colleagues do this (or intend to do this) in the software you use at your respective institutions through incorporation of logic, scripts, and/or persistent and unique identifiers. We will discuss and demonstrate how the Frictionless Data Packages available on Bionomia might facilitate this process. We will set the stage with a broader discussion about the purpose of person identifiers, where they can be found, and our motivations for applying, using, and sharing them.

Logistics & Audience

This will be a remote, 4hr workshop with breaks over Zoom. Sessions will be recorded and later posted for others who were unable to attend. Anticipated attendees will be natural history collections data managers, informaticians, and enthusiasts (some new to this work) who wish to learn practical solutions to organize and link people names in their collections management systems to shared concepts of identity. This workshop will be widely announced and we should expect 100+ attendees.

Code of Conduct

(Adapted from the [Entomological Collections Network](#) 2023 annual meeting Code of Conduct)

Bionomia, its maintainer(s), supporter(s) and enthusiast(s) promote a welcoming environment that is safe, collaborative, supportive, and productive. We value the diversity of

views and backgrounds reflected among all attendees at this and other events. We are committed to providing a positive environment for all, regardless of gender, sexual orientation, ability, religion, socioeconomic status, career status or ethnicity. All organizers, speakers, and attendees are expected to adhere to the Code of Conduct.

Our respectful dialogue policy asks that participants make every effort to maintain constructive discourse with other attendees at all times. We wish to foster an environment that welcomes everyone into conversation. This also includes speakers who respect and honour designated time limits and attendees who likewise respect and honour speaking and listening times.

Expected Behavior

- Treat everyone with respect and consideration
- If you see inappropriate or disrespectful behavior or language, please speak up, either to the offender or to the organizer(s).

Unacceptable Behavior

- Harassment and intimidation, including any verbal or written conduct designed to denigrate, threaten, intimidate, or coerce another attendee
- Discrimination based on gender or gender identity, sexual orientation, age, disability, physical appearance, body size, race, religion, national origin, or culture
- Verbal abuse of any attendee or speaker

Consequences

- Anyone requested to stop unacceptable behavior is expected to comply immediately
- Immediate removal from the meeting without warning
- Prohibit attendance at any future meeting

Reporting Unacceptable Behavior

If you are the subject of unacceptable behavior or have witnessed any such behavior, please immediately notify an organizer.

Agenda

Time (UTC)	Title	Presenter
13:00 – 13:05	Welcome, Agenda, Code of Conduct	David Shorthouse https://orcid.org/0000-0001-7618-5230
13:05 – 13:15	Bionomia: Mission, Integrations, and Updates	David Shorthouse https://orcid.org/0000-0001-7618-5230
13:15 – 13:25	Collecting ORCIDs from the Research Landscape	Shivendra Naidoo https://orcid.org/0009-0005-6221-4682
13:25 – 13:35	Wikidata: The Rosetta Stone of the Linked Data Web	Lydia Pintscher https://orcid.org/0000-0002-3939-2115
13:35 – 13:45	Questions & Answers	
13:45 – 13:55	The Impact of Who: in the database, in the wild	Deborah Paul https://orcid.org/0000-0003-2639-7520
13:55 – 14:15	Disambiguation Exhibition	
	The Tower of Babel in international collections	Krisztina Lohonya https://orcid.org/0000-0002-5313-8674
	Betty Johnson: A disambiguation example (Approx 5 - 7 min presentation)	Siobhan Leachman https://orcid.org/0000-0002-5398-7721
	The Gandalf Effect	Sabine von Mering https://orcid.org/0000-0003-2982-7792
14:15 – 14:30	Break	
14:30 – 14:40	TaxonWorks, GBIF, and Bionomia: how can we move from a one-way to a roundtrip?	Tommy McElrath https://orcid.org/0000-0003-0390-4227
14:40 – 14:50	A workflow for working with three CMS's and an intern at RBGE.	Robert Cubey https://orcid.org/0000-0001-7902-3843
14:50 – 15:00	Disambiguating and roundtripping with EMu and Wikidata at Te Papa	Gareth Watkins
15:00 – 15:10	EarthCape CMS demo	Evgeniy Meyke https://orcid.org/0000-0003-1986-2631
15:10 – 15:20	Roundtripping through Wikidata and Darwin Core	Mathias Dillen https://orcid.org/0000-0002-3973-1252
15:20 – 15:35	Specify agents data model and identifiers Semi-automatically adding identifiers from frictionless data to SPECIFY CMS	Grant Fitzsimmons Joaquim Santos https://orcid.org/0000-0002-2160-4968

15:35 – 15:45	Roundtripping at the University of Oslo	Rukaya Johaadien https://orcid.org/0000-0002-2857-2276
15:45 – 16:00	Questions & Answers	
16:00 – 16:15	Break	
16:15 – 16:45	Breakout Groups	
16:45 – 16:55	Breakout Group Reports	
16:55 – 17:00	Wrap-Up & Next steps	David Shorthouse https://orcid.org/0000-0001-7618-5230

Breakout Groups

16:15 – 16:45: Candidate ideas dependent on interest, enrollment, and volunteer leads

A: Hands-On Collections Management Systems

- Agents, people, and users
- People data attributes, identifiers
- Populating dwc:recordedByID and dwc:identifiedByID

B: Tools of the Trade & Disambiguation Best Practices

- Parsing people names
- OpenRefine and decision-making processes
- Wikidata and ORCID application programming interface

C: Standards & Exchange of Links or Annotations

- Frictionless Data from Bionomia
- Alternate structures such as W3C Web Annotations, Nanopublications, OpenDS

Background Reading and Additional Resources

Roundtripping In Practice

Dillen M, Plank A (2023) Milestone MS32 The design and prototype of a workflow integrating Wikidata into validation and linking. ARPHA Preprints.

<https://doi.org/10.3897/aphapreprints.e114920>

Disambiguation of People's Names

Groom Q, Bräuchler C, Cubey RWN, Dillen M, Huybrechts P, Kearney N, Klazenga N, Leachman S, Paul DL, Rogers H, Santos J, Shorthouse DP, Vaughan A, von Mering S, Haston EM (2022) The disambiguation of people names in biological collections. Biodiversity Data Journal 10: e86089. <https://doi.org/10.3897/BDJ.10.e86089>

Educational Resources

Phillips, M. A., Aina, O., Flemming, A., Zeringue-Krosnick, S. E., Kovacs, J., Leachman, S., Mabry, M. E. (2023). Uncovering Hidden Figures of Natural History Collections Using Digital Data Sleuthing & Storytelling. BCEENET- Biological Collections in Ecology & Evolution Network, QUBES Educational Resources. <https://doi.org/10.25334/FP7P-V513>

Privacy Considerations

Copas, K. 2020. GBIF meets GDPR: Best practices for data publishing institutions
<https://assets.ctfassets.net/u017ejk9rkwj/4QS15IK32vFFdE5BdwvYot/056dbb471b76c94f363d20c8dfac13ae/200625-03-GDPR-Copas.pdf>

Workshops and Presentations

Supporting inclusive and sustainable research infrastructure for systematics (SISRIS) by connecting scientists and their specimens, [Presentations and Materials](#)

See also <https://bionomia.net/workshops>

Exchange Standards

Biodiversity Information Standards (TDWG), Attribution Interest Group,
<https://www.tdwg.org/community/attribution/>

Darwin Core Quick Reference Guide, <https://dwc.tdwg.org/terms/>

Questions and Comments from the Zoom Chat

(thanks to Deb Paul, Joaquim Santos, and Robert Cubey)

David Shorthouse presents: Bionomia: Mission, Integrations, and Updates

- participant (TK) asks for links to papers cited in your talk David.
 - “Manifesto” paper I shared earlier: <https://doi.org/10.36850/mr8>

Shivendra Naidoo presents: Collecting ORCIDs from the Research Landscape

NN  to ORCID for having an explicit statement re data that they will NOT hold: "ORCID does NOT collect or store other identifying or sensitive information such as date of birth, gender, credit card numbers or other financial information, medical information, physical address, phone numbers, social security numbers, passport, or national identity numbers. There are no fields in the ORCID data structure to hold any of this information."

<https://info.orcid.org/orcid-trust>

- **Wouter Addink:** can a trusted organization in orcid also add custom data to an orcid record like taxonomic specialization? (Shivendra answered live in the recording).

Lydia Pintscher Presents: Wikidata: The Rosetta Stone of the Linked Data Web

<https://lydiapintscher.de/talks/WikidataTheRosettaStoneoftheLinkedDataWeb.pdf>

- **Nicky Nicolson:** Wikidata does seem to encourage the assignation of gender onto person records (with things like property missing reports, entity schema etc), will this be reviewed given that gender is now recognised as a property that may violate privacy?
 - Lydia Pintscher 13:48 Yes there are ongoing discussions. What the outill be we will see because it is a community/editor decision at the end.
 - Nicky Nicolson (Kew) 13:50 Can you supply a link to these ongoing discussions please?
 - Lydia Pintscher 13:51 The latest one was a session as part of the data modelling days last week:
https://www.wikidata.org/wiki/Wikidata:Events/Data_Modelling_Days_2023
 - NN Thanks. I also wonder how wikidata might scale to be able to deal with issues around holding data on living people (as the population of academic authors via the scholia project): there perhaps needs to be a mechanism to quickly deal with information that should not have been added and to revert & remove from history, if necessary.
 - Lydia: There is this policy for it:
https://www.wikidata.org/wiki/Wikidata:Living_people

Deborah Paul Presents: The Impact of Who: in the database, in the wild

Deb asks: How many aliases are you known by? (see distribution in chat)

- Mathias Dillen: Alternative approach to this aliases question: What is the maximum number of ids for you in a single resource?
 - MD: in TaxonWorks, there's no limit to ids we can have and we store the aliases too and the role connected to these aliases.
 - Teresa MM: Arctos - unlimited aliases

Krisztina Lohonya Presents: The Tower of Babel in international collections

- **James Macklin 14:08** This highlights the critical need to record the verbatim text of names as part of data capture and not just an assertion made by the transcriber! **Images of the label for reference** can help if available and linkable from a record no matter where it lives.
 - DP: and the transcriptions of those labels mean that downstream others of us don't have to figure out handwriting 😊
 - Rachel Walcott Labels (stuck and loose) are a key link between specimen and person
 - MD: IIIF can help here in subsetting an image in a standardized way, only depicting the label in question.

- **Holly Little 14:08** @Krisztina Lohonya can you share the link for the “Personal names around the world” from one of your slides?
 - Krisztina Lohonya 14:09
<https://www.w3.org/International/questions/qa-personal-names> :)
 - Nicky Nicolson (Kew) 14:10 This is a little old but also a good list:
<https://www.kalzumeus.com/2010/06/17/falsehoods-programmers-believe-about-names/>
- Margaret: No reason not to use the native name label eg Hangul for Korean names (transliterations vary: McCune Reischauer, Revised romanization and whatever google does...)
- NN This is a little old but also a good list:
<https://www.kalzumeus.com/2010/06/17/falsehoods-programmers-believe-about-names/>
- **jshieh:** Issues arise also the rules from where the community that creates the entity. Mixture of order of personal names entered in Wikidata raised reconciliation difficulty especially for CJK names.
- Personal names around the world:
<https://www.w3.org/International/questions/qa-personal-names>

Siobhan Leachman Presents: [Betty Johnson: A disambiguation example](#)

- **jshieh:** @Siobhan Leachman she/her Experience shared very much with the Mrs spousal project I am plowing through!
- **James Macklin:** @Siobhan Leachman she/her Did you somehow report the duplicate record back to Harvard? Or do "we" expect Harvard to discover your hard work via Wikidata or Bionomia? i.e., what is the roundtrip expectation...?
 - SL: I expect the Harvard Index to query Wikidata and get ALL these duplicates. Emailing on each institution that has duplicates would take too much of my time.
 - SL: There are many institutions and databases that have the same duplication issues, and Wikidata is also in this group!
 - MD: IIRC we tried to suggest corrections to the Harvard index a few years ago, but they were very slow to respond and act. Things may have changed since then, but I suspect there is not much resources available at their end to maintain and update the database.
 - SL: Yeah, lack of resources is a common theme.
 - TM: I would assume that the Harvard data came from MCZBase and they can merge agents if they know they need to, but my guess is that they don't know and they are not actively looking at Wikidata to find these things. Lack of resources at the data source is definitely an issue.
 - JM: @Teresa Mayfield-Meyer Actually not the case. The Harvard Index is a stand-alone database that used to be connected to an old CMS (HUH did/does not use MCZBase). At least it was the case when I was there. I am not sure who is responsible for curating the Index now?
 - SL: I've been in discussions where I've been told someone from IPNI was improving it.
 - JM: Interesting, maybe Nicky can speak to that? There has always been a relationship between IPNI and the Index along with the Australian Index based on nomenclature...
 - MD: When I did the analysis in Wikidata some years ago, the overlap with IPNI author IDs was very big. There is still a (horrible) graph here:
<https://github.com/matdillen/STSM-wikidata-people/blob/master/plots/larg>

- [est%20links.png](#) Not sure if this is maintained and whether there are direct links to the Harvard Index in IPNI
- SL: I'm aware there is currently a lot of work going on in Wikidata to link IPNI ids to people items. Harvard index it appears mainly me working on that dataset (at least in the Mix'n'match tool)
 - JS: I reported na error recently to IPNI and got an answer from **Gandhi, Kanchi Natarajan** <gandhi@oeb.harvard.edu> who corrected things and confirmed all changes would take effect on Harvard Index and IPNI
 - SvM: We have worked closely with the IPNI team in the project about plant genera named for women. They quickly updated errors we pointed out and it was great discussing nomenclatural questions with them.
 - JM: Yes, Kanchi is an invaluable expert for sorting out complicated nomenclature including the relationships to people :-)

Rachel Walcott asks: Could someone please post a link to the Frictionless Data packages mentioned

- MD: Example:
<https://bionomia.net/dataset/b740eaa0-0679-41dc-acb7-990d562dfa37/datapackage.json>
- SL: I can give you an example.
<https://bionomia.net/dataset/e2980e63-d152-4219-8c1e-0ffdef3ea6aa?page=1> Look to the top right hand side. They are generated for each dataset
- DP: <https://bionomia.net/collection-data-managers> see the "Incorporating Enhancements" section that explains Frictionless Data Packages

Sabine von Mering Presents: The Gandalf Effect

- Chris Kreussling: General Question: How do all these systems work with citizen/community science data, observers, and identifiers? It seems like all these systems only care about collected specimens. Other records - such as photographic records on iNaturalist - seem to be ignored/disregarded. I know GBIF and GloBI, at least, import my iNaturalist observations. But those don't seem to make it to others, such as Bionomia.
 - ECU: On Inaturalist you have the option to link your orcid ID
- Tommy McE: Bionomia ... uncovering immortals since 2018 ...
- SL: I LOVE those warnings [from Bionomia]. As well as indicating issues with data they can also really help when disambiguating.
- Rob Cubey: @Sabine von Mering I'm now going to have a look at the RBGE specimens for Tonduz records to check "mismatch with lifespan". 😊

Tommy McElrath Presents: TaxonWorks, GBIF, and Bionomia: how can we move from a one-way to a roundtrip?

Work we need to empower

Roundtripping Specimens

- WITH "Buffered Collecting Event"
- W/O assigned "Person" in TaxonWorks
 - 770,584 collection objects
 - 526,697 already associated in Bionomia
 - 176,648 collection objects MINIMUM that would be associated with a person
 - likely more

- A Dating App (to evaluate and select matching people) to get data from Bionomia back into TaxonWorks. TW plans to develop this.
-

Robert Cubey Presents: A workflow for working with three CMS's and an intern at RBGE.

- **Nicky Nicolson (Kew) 14:46** Sounds like Steve Baskauf's code (vanderbot) was the source? Yes, that's confirmed in the repo.
- Github repo shared by Rob <https://github.com/DanielPain/RBGE-vanderbot-wikidata>
- **annika.hendriksen 14:55** @Robert Cubey - Royal Botanic Garden Edinburgh How did connect the three CMS's for the publication of the data on f.e. your website : by identifying the entities in each system with a Wikidata identifier? And if there wasn't: did you create Wikidata items? And are you able to register Aliasses (not to be confused with name variations due to typo's or misinterpretations) Answer: I am afraid we did not connect the three CMS's (each has unique functionality) - we only looked for "know" figures i.e. Well know or famous plant collectors - and looked for records for these individuals in each of the three CMS's.

Then de-duplicated in the individual's systems (if appropriate) and then looked to see if there was a wikidata record for the individual, if not we created the wiki-data record and added the Q number to the three CMS's.

So, no central agents / person table across the CMS's just a link to a unifying "Wikidata" record. Aliases were added to both wikidata items and to the three CMS's (very inefficient workflow, but very satisfying tidying up all the "loose ends".)

As we push our data out to GBIF we push out the wikidata Q number for our collectors so allowing their automatic pick up by Bionomia.

Gareth Watkins Presents (recording): Disambiguating and roundtripping with EMu and Wikidata at Te Papa

- <https://blog.tepapa.govt.nz/2023/12/05/agreeing-to-agree-roundtripping-people-identifiers-with-wikidata>

Evgeniy Meyke Presents: EarthCape CMS demo

peeking inside EarthCape (a live! demo)

This screenshot shows a list of herbarium specimens. The columns include Catalogue Number, Last Modified On, Last Modified By, Family, Genus, Species, Subspecific Taxon, Taxonomic Name, Type, Accepted Name, Storage Name, and Storage Kew Region. The data shows various plant species from different families like Fabaceae, Rosaceae, and Asteraceae.

This screenshot shows a detailed view of a herbarium object (K001490816). It includes tabs for Collections, Taxonomic names, Spreadsheets, SEED, Visibility Test, Occurrences, LIVING, and associated research. The 'Occurrences' tab is active, showing occurrence details like Occurrence Info, Types, Derived Objects, Condition, Cultivation, Restrictions, Images, and Associated Research.

This screenshot shows a list of seed viability test results. The columns include Created On, Created By, Id, Status, Pass Fail, Type, Viability Assess., Test Type, Start Date, End Date, Catalogue Number, and Days. The data shows multiple entries for different dates and users, with various status outcomes like 'Pass', 'Fail', and 'In pro.'

people data in EarthCape
keeping verbatim / legacy data

The screenshot shows the Earthscape Windows Client AU application window. The top menu bar includes 'File', 'Recording...', 'Home', 'View', 'Tools', 'People - Earthscape Windows Client AU', and a timestamp '01:52:06'. The left sidebar contains a 'Navigation' pane with various organizational and taxonomic categories. The main area features a grid view with columns for People, Collection Objects, HERBARIUM, FUNGARIUM, SPIRIT, Organizations, Collections, Taxonomic names, Spreadsheets, and SEED. The grid displays data rows corresponding to the categories listed in the sidebar.

People	Collection Objects	HERBARIUM	FUNGARIUM	SPIRIT	Organizations	Collections	Taxonomic names	Spreadsheets	SEED
Int. b.	First N. b.	Last Name	Middle Name	User Account	Organization	Department	Job Title	Email	Country
sub ...	J.	Wylie		—	—	—	—	—	—
L.E.		Wylie sub Wood							
K.J.		Wyles Poort							
F.L.		Wyman							
L.F.		Wynd							
	F.	Wynd							
		Wynne-Samuel							
	B.	Wynn-Jones							
	F.	Wynn-Jones							
		Wyrdrack							
	K.	WYRWA			POLISH ACADEMY OF SCIENCES		INSTITUTE OF PLANT GENETICS (B168)		kwyr...
	P.	WYSE-JACKSON			NATIONAL BOTANIC GARDENS, DUBLIN		(9427)		peter...
	P.	WYSE-JACKSON			BOTANIC GARDENS CONSERVATION INT.		(A172)		
	P.	Wyses-Jackson			University College, Dublin.		University College, Dublin. (Dept.)		
	X.I.	X.M. van der Burgt							
	P.	Xaba							
		Xander van der Burgt							
	M.	Xanthos							
	M.D.	Xanthos							
	M.	Xanthos			Royal Botanic Gardens Kew, UK.		Royal Botanic Gardens Kew, UK. (Dept.)		
		Martin							
		Xanthos							

- merge tool to be built so that it can be used on other record types as well
 - integrate with Bionomia and Wikidata - would need consultation with others working on this to build this functionality
 - EarthCape pinging Bionomia API

	Diagnostic	score	orfid	fullname	given	family	orcid	wikidata	uri	fullname_reverse	label	is_public	has_occurrences	count	other_names
✓	Recording...	56.24935	Charles Robt. Charles	Charles	Robt.	Charles	Darwin	Q1015	http://	Darwin, Charles Robert Charles	Charles, Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.90037	Sue Darvin -	Sue	Darwin A.			Q10518299	http://	Darwin, Abbott, Sue	Sue Darvin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.90037	Denis Darvin -	Denis			Darwin	0000-0	http://	Darwin, Denis	Denis, Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	Denis Horatio Darwin	Horatio	Denis	Morales	Darwin	Q420447	http://	Darwin, Horatio, Denis	Horatio, Denis, Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	Hannah Wedderburn Darwin	Wedderburn	Hannah	D	Darwin	Q172181	http://	Darwin, Hannah D.	Hannah D., Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	Henry Darwin	Henry	D	Rogers	Darwin	Q113607470	http://	Darwin, Henry Rogers	Rogers, Henry, Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	Anne Belle O. Anne Be. Owens D.	Owens D.			Darwin	Q113607470	http://	Darwin, Anne Belle Owens	Owens, Anne B., Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	Darwin Thos.	Darwin		Hickman	Darwin	0000-0	http://	Darwin, Thomas Hickman	Hickman, Thomas Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	D & Darwin -	D	5	Edison	Darwin	0000-0	http://	Darwin, D & Darwin Edison	Edison, D & Darwin Paul	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	Leonard Darwin	Leonard	Darwin		Darwin	Q954932	http://	Darwin, Leonard	Leonard, Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	George Darwin	George	Darwin		Darwin	Q250216	http://	Darwin, George	George, Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	Erasmus Darwin	Erasmus	Darwin		Darwin	Q10749	http://	Darwin, Erasmus	Erasmus, Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	Robert Darwin	Robert	Darwin		Darwin	Q009112	http://	Darwin, Robert	Robert, Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	Robert Darwin	Robert	Darwin		Darwin	Q04193	http://	Darwin, Robert	Robert, Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		55.27579	Daniel Andrew	Daniel	Andrew		Darwin	Q1205217	http://	Darwin, Daniel Andrew	Daniel, Darwin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		13.94354	Robert Els.	Robert	Els.		Easton	Q1555142	http://	Easton, Robert Els.	Easton, Robert Els.	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		8.969431	Francis Victor M.	Francis	Victor M.			Q133435	http://	Galdiano, Victor M. Francis	Galdiano, Victor M. Francis	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		7.65819	Charles Chsc.	Charles	Chester			Q9243528	http://	Chester, Charles	Charles, Chester	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	
✓		7.65819	Emmer Bell	Emmer	Bell			Q5066516	http://	Bell, Emmer	Emmer, Bell	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0	

- **annika.hendriksen** 15:12 @Evgeniy Meyke (earthcape) Do You register the Botanist Author Abbreviation also ? a great property to use while reconciling with Wikidata ;)
 - Evgeniy Meyke (earthcape) 15:15 Thanks for a tip! Notat the moment but I will definitely look into it.
 - Would love to chat more @Evgeniy Meyke (earthcape) - very cool live Demo. We're [at TaxonWorks] thinking about the "dating app" (user interface and functions) in the context of your demo. The "how discern, evaluate, and grab the matches returned" into the database.

..... Existing models for Agents

Frictionless data packages (e.g. Bionomia)

- "Extended DWC-A"
- Machine-readable metadata but
- No standardized schemas
- Custom terms

```
"name": "users",
"format": "csv",
"mediatype": "text/csv",
"encoding": "utf-8",
"profile": "tabular-data-resource",
"schema": {
  "fields": [
    {
      "name": "id",
      "type": "integer"
    },
    {
      "name": "name",
      "type": "string",
      "skos:exactMatch": "http://schema.org/name"
    },
    {
      "name": "familyName",
      "type": "string",
      "skos:exactMatch": "http://schema.org/familyName"
    },
    {
      "name": "particle",
      "type": "string"
    },
    {
      "name": "givenName",
      "type": "string",
      "skos:exactMatch": "http://schema.org/givenName"
    },
    {
      "name": "alternateName",
      "type": "array",
      "skos:exactMatch": "http://schema.org/alternateName"
    },
    {
      "name": "sameAs",
      "type": "string",
      "format": "url"
    }
  ]
}
```

<https://github.com/AgentschapPlantentuinMeise/roundtrip>

Converts Bionomia frictionless data to DwC agent attribution extension



..... Roundtripping now: Darwin Core

term	definition	vocabulary
occurrenceID	ID of the occurrence the agent acted upon.	
agentType	The nature of the agent.	yes
agentIdentifierType	The type of identifier for the agent.	yes
identifier	A string conforming to an identification system.	
name	The name of the agent.	
alternateName	An alias for the item. Other full name agent may have been known under such as maiden name.	
verbatimName	As written on occurrence, such as the collection or determination label.	
action	The name of the single action written as a verb in past tense.	yes
role	The name of the role the agent played in the context of executing the action.	yes
displayOrder	The display order for the agent that executed the action when more than one agent was a participant.	integer
identificationID	An identifier for the identification, i.e. the body of information associated with the assignment of a scientific name.	
startedAtTime	Start is when an action is deemed to have been started by an agent.	ISO date
endedAtTime	End is when an action is deemed to have been ended by an agent.	ISO date

<https://github.com/tdwg/attribution>

More info in DiSSCo Prepare D5.4:

<https://dx.doi.org/10.34960/ajxs-zr25> (down atm)
[Alternate URL](#)



- role term not used (was dropped) but would be useful.

..... Two open action points!

- Can we get a version of the DwC Attribution extension approved and/or implemented in the short term?
 - <https://github.com/tdwg/attribution>
- Can we (currently!) find a common way to cite specimens in a resolvable manner?
 - https://www.wikidata.org/wiki/Wikidata:Property_proposal/GBIF_occurrence_ID



- please go to GitHub to contribute to the discussion
- Can we get a version of the DwC Attribution extension approved and/or implemented in the short term?
- See <https://github.com/tdwg/attribution>
- Can we (currently!) find a common way to cite specimens in a resolvable manner?
- - https://www.wikidata.org/wiki/Wikidata:Property_proposal/GBIF_occurrence_ID

Grant Fitzsimmons Presents: Specify agents data model and identifiers

Agent Duplication

In a normalized model, duplicate agents can arise from messy legacy data, poor data entry practices, ambiguous initials, and other factors.



- A.C.Bentley
- A. Bentley
- Andy Bentley
- Andrew Bentley
- A.C.B.

Joaquim Santos

Presents: Semi-automatically adding identifiers from frictionless data to SPECIFY CMS

Accessing Identifiers

Agents have a 1:M **Identifiers** subform that captures any unique identifiers associated with an agent.

You can also capture an identifier in a regular text field.

The screenshot shows the 'Query Results: Agent' interface. At the top, there are fields for Agent Type (Person), Title (Mr.), Job Title (Collections Manager), Last Name (Bentley), First Name (Andrew), Middle Initial (C), Abbreviation, Email (abentley@ku.edu), Org Type, and Remarks. Below these is a 'Identifiers' subform. A red arrow points to the 'Identifiers' section. Inside, there is a table with one row showing an ORCID identifier: 0000-0002-3093-1258, Identifier Type: ORCID. Below the table are buttons for ORCID, Wikidata, and Bionomia.

This can be customized by the collections admin

Accessing Identifiers

You can add weblink buttons to connect Bionomia, ORCID, Wikidata, or any other site to your identifier field!

The screenshot shows the 'Query Results: Agent' interface. It has the same fields and subforms as the previous screenshot. However, the 'Identifiers' subform now contains three buttons at the bottom: ORCID, Wikidata, and Bionomia. Three red arrows point upwards from the bottom of the page towards these buttons.

This allows you to use the same ID with Bionomia as with ORCID

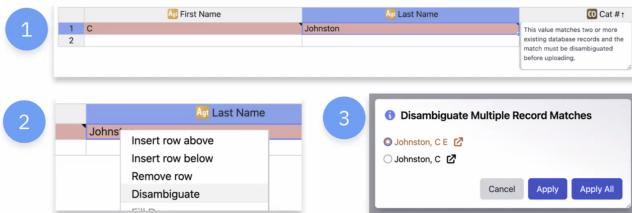
Bulk Adding Identifiers

You can add add identifiers to existing agents in bulk by using the **WorkBench**, Specify's bulk importing tool.

The screenshot shows the 'Data Mapper' interface in WorkBench. At the top, there are tabs: Data Set, Import Identifiers, Meta Data, Tools, Data Mapper, Validate, Results, Upload, Revert, and Save. The main area displays a table of agent data. The columns are: Firstname, MiddleInitial, Lastname, Identifier, and IdentifierType. The IdentifierType column contains ORCID values such as 0000-0002-3093-1258, 0000-0002-8775-2739, 0000-0002-2875-0474, etc. The table has 19 rows. At the bottom, there are buttons for Search, Replace, and various reporting and saving options.

Agent Disambiguation

The **WorkBench** can also disambiguate agents when uploading data.



- tool that resolves duplicates

Resolving Duplicates

Specify has a **record merging** tool that allows users to clean up duplicates and automatically store the variant names.

You can see **date ranges** associated with activities done by each agent and **preview associated records**.

This screenshot shows the 'Merge Records' dialog in the Specify software. It displays two records for 'Bentley, Andy C' and 'Bentley, Andrew C'. The left pane lists fields with dropdown menus for previewing values from both records. The right pane shows detailed history and audit information for each record, including timestamps, user actions, and merge details. At the bottom, there are buttons for 'Preview', 'Merge', and 'Cancel'.

Notes from Chat -

Grant Fitzsimmons: "Can we get a version of the DwC Attribution extension approved and/or implemented in the short term?

- <https://github.com/tdwg/attribution>

Can we (currently!) find a common way to cite specimens in a resolvable manner?

- https://www.wikidata.org/wiki/Wikidata:Property_proposal/GBIF_occurrence_ID

David Shorthouse "Please do include full URIs for people identifiers in recordedByID and identifiedByID."

Teresa Mayfield-Meyer "Occurrence IDs describe occurrences NOT specimens...." Answer:

Mathias Dillen: "The wikidata property is called occurrence id, but is actually about the id for the GBIF record (gbifID), not dwc:occurrenceID"

Wouter Addink: "GBIF occurrence ID is an id for an occurrence not a specimen and also it is not persistent so it is not usable to refer to a specimen. To solve this we are working on minting digital specimen DOIs" Answer: Mathias Dillen; "GBIF has implemented measures to keep their IDs as persistent as possible. These IDs are resolvable, recognizable and unambiguous in the GBIF data model. Bionomia is an example of a resource which uses the GBIF ID as its primary link back to the specimen data. That has been problematic over the years, but this year has seen dramatic improvement due to the changes implemented by GBIF to prevent gbifID drift."

- getting Users information from Bionomia Frictionless Data Pkg, back into Specify

- developed methods to help automate
 - get list of agents from the frictionless data package that do not yet have an identifier inside the local Specify database

Agent

Agent Type:	Person	Title:	First Name:	Jack Rodney	Middle Initial:	<input style="width: 20px; height: 20px;" type="button" value="..."/>
Last Name:	Laundon	Email:				
Abbreviation:						
Remarks:	<input style="width: 100%; height: 40px;" type="text"/>					
Date Of Birth:			Date Of Death:			
Agent Identifier:						
Addressess						
<input type="checkbox"/> Is Current <input type="checkbox"/> Is Primary						
Address:						
Address2:						
City:	State:	Postal Code:	Country:		Fax:	<input style="width: 40px; height: 20px;" type="button" value="Grid"/>
Room/Building:	Phone 1:	Phone 2:				
{<} {<} {>} {>}<						
Variants						
Var Type:	Name:					
{<} {<} {>} {>}<						
Agent Specialties						
<input style="width: 40px; height: 20px;" type="button" value="Grid"/>						

Agent	Agent Type: Person	Title:	First Name: Jack Rodney	Middle Initial:																		
Last Name:	Laundon	Email:																				
Abbreviation:																						
Remarks:																						
Date of Birth:	1934	Date Of Death:		2016-12-31																		
Agent Identifier: http://www.wisidata.org/entity/O13426189																						
▼ Addresses <table border="1"> <tr> <td><input type="checkbox"/> Is Current</td> <td><input type="checkbox"/> Is Primary</td> </tr> <tr> <td>Address:</td> <td colspan="3"></td> </tr> <tr> <td>Address2:</td> <td colspan="3"></td> </tr> <tr> <td>City:</td> <td>State:</td> <td>Postal Code:</td> <td>Country:</td> </tr> <tr> <td>Room/Building:</td> <td>Phone1:</td> <td>Phone2:</td> <td>Fax:</td> </tr> </table>					<input type="checkbox"/> Is Current	<input type="checkbox"/> Is Primary	Address:				Address2:				City:	State:	Postal Code:	Country:	Room/Building:	Phone1:	Phone2:	Fax:
<input type="checkbox"/> Is Current	<input type="checkbox"/> Is Primary																					
Address:																						
Address2:																						
City:	State:	Postal Code:	Country:																			
Room/Building:	Phone1:	Phone2:	Fax:																			
 < > > 																						
<input style="border: 1px solid black; padding: 2px; margin-right: 10px;" type="button" value="Grid"/> <input style="border: 1px solid black; padding: 2px;" type="button" value="Print"/>																						
▼ Variants <table border="1"> <tr> <td>Var Type: Variant</td> <td>Name: Laundon, Jack</td> </tr> </table>					Var Type: Variant	Name: Laundon, Jack																
Var Type: Variant	Name: Laundon, Jack																					
 < > > 																						
 < > > 																						
<input style="border: 1px solid black; padding: 2px; margin-right: 10px;" type="button" value="Grid"/> <input style="border: 1px solid black; padding: 2px;" type="button" value="Print"/>																						
▼ Agent Specialties <table border="1"> <tr> <td>Specialty:</td> <td>Description:</td> </tr> </table>					Specialty:	Description:																
Specialty:	Description:																					

- sometimes multiple matches, chooses one he thinks is correct

Bionomia Agents

Bionomia Agents																																							
Import agents table from Bionomia Update existing agent data from Bionomia table																																							
<table border="1"> <tr><td>1</td><td>2</td><td>-</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td><td>23</td><td>x</td></tr> </table>														1	2	-	14	15	16	17	18	19	20	21	22	23	x												
1	2	-	14	15	16	17	18	19	20	21	22	23	x																										
<table border="1"> <thead> <tr> <th>ID</th><th>Name</th><th>Family Name</th><th>Particle</th><th>Given Name</th><th>Alternate Name</th><th>Same As</th><th>ORCID</th><th>Wikidata</th><th>Birth Date</th><th>Birth Date Precision</th><th>Death Date</th><th>Death Date Precision</th></tr> </thead> <tbody> <tr> <td>46655</td><td>Elias, hermano</td><td>Elias</td><td></td><td>hermano</td><td></td><td>["Elias","Bro Elias","Elias, frère","Fr. Elias","Hno. Elias","hermano Elias"]</td><td></td><td>Q21511810</td><td>1870-01-01</td><td>year</td><td>1937-01-01</td><td>year</td></tr> </tbody> </table>														ID	Name	Family Name	Particle	Given Name	Alternate Name	Same As	ORCID	Wikidata	Birth Date	Birth Date Precision	Death Date	Death Date Precision	46655	Elias, hermano	Elias		hermano		["Elias","Bro Elias","Elias, frère","Fr. Elias","Hno. Elias","hermano Elias"]		Q21511810	1870-01-01	year	1937-01-01	year
ID	Name	Family Name	Particle	Given Name	Alternate Name	Same As	ORCID	Wikidata	Birth Date	Birth Date Precision	Death Date	Death Date Precision																											
46655	Elias, hermano	Elias		hermano		["Elias","Bro Elias","Elias, frère","Fr. Elias","Hno. Elias","hermano Elias"]		Q21511810	1870-01-01	year	1937-01-01	year																											
Save Data Save LastName/FirstName AgentID_Converted Text1_Converted LastName MiddleInitial FirstName Abbreviation DateOfBirth DateOfDeath AgentVariant AgentType Expeditions Group Remarks Collector Determiner																																							
<input type="button" value="Save"/>																																							

Specify 6.8.03

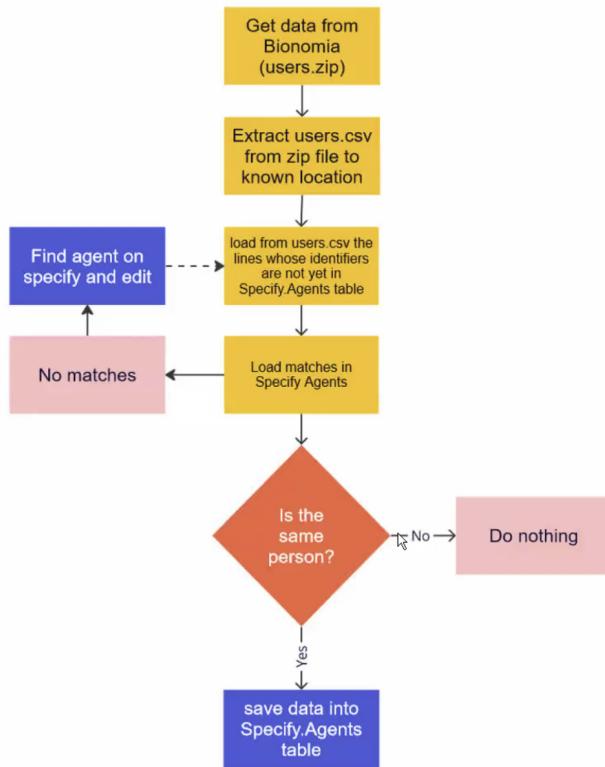
File Edit Data System Tabs Help

Welcome Data Trees Reports Interactions Statistics Query Workbench BatchEdit Attachments

Record Sets Search Results - 2

Last Name	First Name	Middle Initial	Address	Address2	Agent Identifier
Elias					
Elias [Frère]					

Workflow



miro

- SL: Q for Joaquim: Bionomia lists multiple aliases for a deceased collector sourced from Wikidata. Do you import ALL those aliases in your system?
 - JS: Yes. All the aliases in the Bionomia file are being imported
 - Bjorn: For the 6 MUSIT in Oslo collections I used about 10 hours to look through the missing attributions and implement the new ID's and did implement some new synonyms. Last time I did this was about one year ago. I sent a few wishes

to the MUSIT developer team.

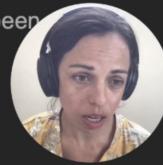
- NN: @David Shorthouse have you opinion on the wikidata "bionomia id" property (or fed back to them on that)? Since your design intention is that bionomia IDs are only ever orcid or wikidata q ids
 - DS: answered outloud. Simplifies Bionomia work, queries for matching IDs (in both places) is simplified.
 - Bionomia is a “project” ... key to

Rukaya Johaadien Presents: [Roundtripping at the University of Oslo](#)

Possible future problem: duplicated effort

Duplicated effort: Next time BP will have to look through & discard all the same incorrect attributions as the previous time

- Possible solutions:
 - BP corrects data on Bionomia
 - BP corrects data on a new attribution user interface which records that which he considers to be misattributions
 - Each time we only look at **new** attributions which have been attributed/helped on or something in that case)



Scaling?

Problems: Scale and engagement

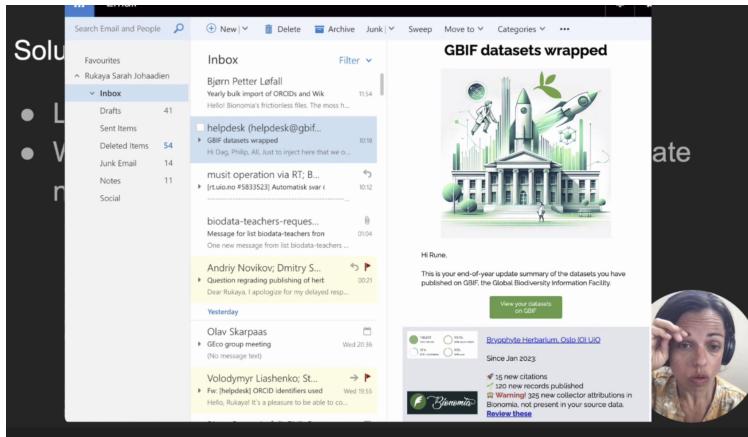
Scale and engagement: Currently we only have BP doing this - but we need to expand to all the curators of all collections in Norway published on GBIF

Having a network for the Norwegian GBIF Node is useful here, however we don't have much engagement from our curators - they are happy their data gets published but not much beyond that



Solutions

- imagining what to do about the need to see “progress” / Status



Solutions

- Leverage the GBIF Norway network further
- We plan to build a twice-yearly email status update newsletter “GBIF datasets wrapped”
- Future plan to move the MUSIT NHM UiO people name synonyms to other museum collection’s MUSIT deployments (starting with Tromsø)
- seeking more automated round-tripping



Q: Nicky Nicolson: “@David Shorthouse have you opinion on the wikidata “bionomia id” property (or fed back to them on that)? Since your design intention is that bionomia IDs are only ever orcid or wikidata q ids”

Ans: in person - paraphrased “use wikidata id’s rather than Bionomia - as Bionomia is a project”

Ans: NN: Yeah I thought of it because you said don’t hold a field like “bionomia ID” in your CMS: maybe wikidata is muddying the waters there

Ans: Mathias Dillen: “Bionomia ID does have significant added value compared to just listing the Wikidata or ORCID. It indicates that the person record with that ID was processed in Bionomia, which is not true for most ORCIDs and Wikidata Qids.”

Question from Deb: What else can we do with personal identifiers?

Ans: Rob Cubey - Expeditions (see-

https://www.wikidata.org/wiki/Wikidata:WikiProject_Research_expeditions

From Chat:

Shelley James: “Improved/more reliable duplicate discovery”

Sabine von Mering: “Transdisciplinary provenance research”

Rachel Walcott: “Connect Labels and handwriting and Panama bank accounts ;-)”

Nicky Nicolson: “where we are depending on few experts for expert id”

Annika.hendriksen: “named entity recognition on HTR label texts”

Lauren Hughes: "Humanities research and looking at their Authoritative lists with Library of Congress is NHM, London research being explored just now"

Q: Tommy McElrath9INHS): e.g. the new DWC term: verbatimLabel

Ans: rukaya "<https://dwc.tdwg.org/terms/#dwc.verbatimLabel>"

Ans: Lauren Hughes: Verbatim - can be verbatim label or verbatim museum register book or verbatim index card depending on where the data was first sourced for digital

Ans: Elspeth Haston" Would be so good to have verbatim data captured, but it's also a HUGE amount of work for the people transcribing too."

Ans: rukaya "Not really with handwriting OCR as good as it is (e.g. google cloud vision)"

Ans: annika.hendrik "transcribing handwriting is often an interpretation ;S"

"

Open Discussion Time (just before and into break)

Q: James Macklin: "If we have too many IDs that are interrelated is there a risks of circular logic and associated errors?

- AWeeks: Anyone here connecting to the Humanities world with this people work?
 - SL: Lisa Delesio reached out for info on a particular person in Bionomia
 - TM: German researcher example
 - NN: We have some plant humanities researchers here are Kew: e.g Caroline Cornish <https://www.kew.org/science/our-science/people/caroline-cornish>
 - LHughes: We are building useful 'DWC' fields with in collaboration with humanities/historians to expand the potential of our people/expeditions records
 - KL: labels and the story handwriting tells, also person who collected with Alfred Russell Wallace, ...
 - I present you, Eaton. He had a particularly difficult handwriting.
https://en.wikipedia.org/wiki/Alfred_Edwin_Eaton
 - https://en.wikipedia.org/wiki/Elliott_Coues The spiritualist attending seances with Alfred Russel Wallace 😊
 - Jcsagebiel: Our information on paleontologist Jack Wilson resulted in a book, a couple of museum exhibits, and a live show on stage at the historic Paramount theatre in Austin, Texas.
 - KL: We actually have a collector from whom we have a huge collection and we know he was arrested in Algeria for collecting illegally on his lunchtime
 - DE: I have occasionally found collection curators (museum staff) crediting their own very young children as co-collectors :-)
 - TK: We have been linking digitized archives of TDA Cockerell to related specimen records. So much cool info
- JM: often don't know why, correspondence clarifies, gives insights, gives CONTEXT
 - Collection Managers might have lots of this information.