

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. DO NOT EXCEED FIVE PAGES.

NAME: Poelen, Jorrit

eRA COMMONS USER NAME (credential, e.g., agency login):

POSITION TITLE: Research Scholar

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE (if applicable)	END DATE MM/YYYY	FIELD OF STUDY
University of Groningen, Groningen, the Netherlands	MS	2000	Experimental Physics

A. Personal Statement

The overall goal of the proposed work is to harness the collective power of bioinformatics, artificial intelligence (AI) and ecology to transform our understanding of taxon-specific zoonotic risk. I have the expertise, communication skills and motivation necessary to contribute to the proposed work.

After spending about a decade in industry, I was re-introduced to academia in 2011 by attending the Ecological Integration Symposium at Texas A&M College Station. As an invited visual artist, I was asked to take a minute or two to introduce my video installation “Evidence of Use” to the conference attendees. I noticed the audience freeze when I suggested that all scientific data should be liberated for the unrestricted enjoyment of scientists and artists alike. Fortunately, after the talk, I was approached by an enthusiastic audience member who introduced me to his big dream and his excel spreadsheet with fish stomach diet data. This marked my unlikely entry into the field of ecological informatics and the start of my collaboration with Jim Simons, then marine biologist at Texas A&M Corpus Christi. Our complementary skills and shared enthusiasm led to the materialization of the Gulf of Mexico Species Interactions (GoMexSI) and Global Biotic Interactions (GloBI) projects. Both projects re-purpose existing species datasets in an attempt to realize Charles Elton’s vision to “[...] provide conceptions which can link up into some complete scheme the colossal store of facts about natural history [...]” (Elton CS. 1927. *Animal Ecology*. doi:10.5962/bhl.title.7435). GoMexSI’s focus is on synthesis of marine food webs in the Gulf of Mexico, while GloBI aims to make it easier to share and find any openly available species interaction (e.g., virus-host, parasite-host, plant-pollinator, predator-prey) claim.

As a small, independent, open source and open data developer, I have no choice but to make use of openly available datasets, leverage cheap (or open) compute infrastructures, and establish a wide range of mutually beneficial research collaborations with institutional scientists. I have noticed that my frugal approach to software design, automation and collaborations have resulting in a sustained, economically viable, way of life with beneficial side effects such as:

1. Providing a de-facto continuous peer-review service of existing open datasets and -services
2. Developing pragmatic approaches to link biological data at scale
3. Engaging in, and enabling of, “risky” scientific studies that re-use existing datasets to answer (global) questions
4. Catalyzing discussions around data integration, frugal tools, data preservation, data standards and ontologies

Over the years, I have come to realize that many of my contributions to the academic community cannot be measured in, or are often detrimental to, traditional metrics like citation indexes, grant moneys, institutional affiliation, and academic tenure. With this, I would like to think that, as a free-living parasite/symbiont, I help my hosts evolve by transferring infectious new ideas across disciplinary and institutional

boundaries.

In summary, my background is uniquely suited to successfully address the extra-ordinary cross-disciplinary data integration challenges outlined in Dr. Upham's proposal "Intelligently predicting viral spillover risks from bats and other wild mammals."

B. Positions and Honors

Positions and Employment

1999 - 2001	Research Associate, University of California, Berkeley, Berkeley, CA
2001 - 2002	Software Engineer, Tekadence Inc., San Francisco, CA
2002 - 2006	Software Engineer, GE Healthcare, Zeist
2006 - 2008	Software Engineer, GE Healthcare, Barrington, IL
2008 -	Freelance Software Engineer, Self, Oakland, CA
2010 - 2011	Software Engineer, Xoom Inc., San Francisco, CA
2011 - 2012	Software Engineer, Pandora Media Inc., Oakland, CA
2012 - 2013	Assistant Director, University of California, San Francisco, Helen Diller Family Comprehensive Cancer Center Translational Informatics Program, San Francisco, CA
2020 -	Research Scholar, Ronin Institute, Montclair, NJ

Other Experience and Professional Memberships

Honors

C. Contribution to Science

1. Providing a de-facto continuous peer-review service of existing open datasets and -services.

As funding agencies and institutions are embracing open science principles, many projects, institutions and infrastructures claim to provide open-access and FAIR (findable, accessible, interoperable, re-usable) data. However, traditional human review methods used by scientific journals and data publication platforms lack the ability to verify and monitor the openness and FAIR-ness of published data at scale. I have contributed to designing, implementing and promoting automated processes to help implement method to continuously review the availability and interoperability of openly published datasets. For instance, a tool developed in collaboration with Pensoft Publishers helps to monitor the availability of species interaction data in Pensoft publications like ZooKeys and the Biodiversity Data Journal (Dimitrova et al. 2020). Other examples include automated and continuous review processes for Global Biotic Interactions (Poelen et al. 2014) and automated detection of ambiguous taxonomic names at scale (Thessen et al. 2018).

- a. Dimitrova M, Poelen J, Zhelezov G, Georgiev T, Agosti D, Penev L. Semantic Publishing Enables Text Mining of Biotic Interactions. Biodiversity Information Science and Standards. 2020 September 28; 4:-.
- b. Thessen A, Poelen J, Collins M, Hammock J. 20 GB in 10 minutes: a case for linking major biodiversity databases using an open socio-technical infrastructure and a pragmatic, cross-institutional collaboration. PeerJ Computer Science. 2018 September 17; 4:e164-.
- c. Poelen J, Simons J, Mungall C. Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. Ecological Informatics. 2014 November; 24:148-159.

2. Developing pragmatic approaches to link biological data at scale. Over the last decade, hundreds of millions of dollars of public and private funding has dramatically increased the availability of biodiversity and ecological data through digital communication networks. While some initiatives have establishes ways to links between datasets (e.g., NCBI Linkout program), creating cross-

disciplinary research data products remains a challenge. As the volume of datasets that I work with grows, I contribute to developing novel approaches to help link biological data at scale. For instance, I've shown how to re-use existing infrastructures used by open source software and open science projects to integrate decentralized and heterogeneous species interaction datasets (Poelen et al. 2014, Poelen 2017). In addition, I've contributed to developing methods to use scalable nanopublications to disseminate scientific data (Kuhn et al. 2018) and used open compute platforms to link quickly link major biodiversity databases (Thessen et al. 2018). Also, I helped develop a method to version and reliably reference (i.e., link) entire biodiversity data networks using decentralized and location agnostic data tracking techniques (Elliott et al. 2020).

- a. Elliott M, Poelen J, Fortes J. Toward reliable biodiversity dataset references. *Ecological Informatics*. 2020 September; 59:101132-.
- b. Kuhn T, Banda J, Willighagen E, Ehrhart F, Evelo C, Malas T, Dumontier M, Merono-Penuela A, Malic A, Poelen J, Hurlbert A, Centeno Ortiz E, Furlong L, Queralt-Rosinach N, Chichester C. Nanopublications: A Growing Resource of Provenance-Centric Scientific Linked Data. 2018 IEEE 14th International Conference on e-Science (e-Science). 2018 IEEE 14th International Conference on e-Science (e-Science); ; Amsterdam. IEEE; c2018.
- c. Thessen A, Poelen J, Collins M, Hammock J. 20 GB in 10 minutes: a case for linking major biodiversity databases using an open socio-technical infrastructure and a pragmatic, cross-institutional collaboration. *PeerJ Computer Science*. 2018 September 17; 4:e164-.
- d. Poelen J. Global Biotic Interactions: A Catalyst For Integrating Existing Interaction Datasets, Connecting Data Curators And Developing Data Exchange Methods. *Proceedings of TDWG*. 2017 August 11; 1:e20214-.

3. **Engaging in, and enabling of, “risky” scientific studies that re-use existing datasets to answer (global) questions.** While data re-use is a phrase that is often used in data management plans and review articles, researchers and scientific publishers are still hesitant to use and/or publish studies based on aggregate, cross-disciplinary research datasets compiled from heterogeneous, third party, data sources. I have directly and indirectly contributed to scientifically "risky" studies that use novel data re-use methods to help answer scientific questions (Grüss et al. 2019, Hayden et al. 2019). Indirect contributions can be found in the list publications that cited Global Biotic Interactions (Poelen et al. 2014) .

- a. Hayden B, Palomares M, Smith B, Poelen J. Biological and environmental drivers of trophic ecology in marine fishes - a global perspective. *Scientific Reports*. 2019; 9(1):-.
- b. Grüss A, Palomares M, Poelen J, Barile J, Aldemita C, Ortiz S, Barrier N, Shin Y, Simons J, Pauly D. Building bridges between global information systems on marine organisms and ecosystem models. *Ecological Modelling*. 2019 April; 398:1-19.

4. **Catalyzing discussions around data integration, frugal tools, data preservation, data standards and ontologies.** Re-using and publishing research data remains a challenge today. To help address the sociological, technical and economical aspects that inhibit the efficient use of our existing and future datasets, I spend a significant amount of my time to discuss, and contribute, to the building blocks of a sustainable data ecosystem: human collaboration networks, frugal tools, data preservation, standards and ontologies. At an invited workshop, I contributed to a vision to create an interoperable framework for essential biodiversity variables (Hardisty et al. 2019). As a contributor and editor to the Open Trait Network, I helped to create a socio-technical framework for a decentralized network of trait-based scientists (Gallagher et al. 2020). I've contributed to various ontologies such as the the ontology of plant stress (Meier et al. 2018) and the OBO Foundry's Relations Ontology. Finally, I've developed various tools to implement frugal data preservation workflows (Poelen 2019).

- a. Gallagher RV, Falster DS, Maitner BS, Salguero-Gómez R, Vandvik V, Pearse WD, Schneider FD, Kattge J, Poelen JH, Madin JS, Ankenbrand MJ, Penone C, Feng X, Adams VM, Alroy J, Andrew

SC, Balk MA, Bland LM, Boyle BL, Bravo-Avila CH, Brennan I, Carthey AJR, Catullo R, Cavazos BR, Conde DA, Chown SL, Fadrique B, Gibb H, Halbritter AH, Hammock J, Hogan JA, Holewa H, Hope M, Iversen CM, Jochum M, Kearney M, Keller A, Mabee P, Manning P, McCormack L, Michaletz ST, Park DS, Perez TM, Pineda-Munoz S, Ray CA, Rossetto M, Sauquet H, Sparrow B, Spasojevic MJ, Telford RJ, Tobias JA, Violle C, Walls R, Weiss KCB, Westoby M, Wright IJ, Enquist BJ. Open Science principles for accelerating trait-based science across the Tree of Life. *Nat Ecol Evol.* 2020 Mar;4(3):294-303. PubMed PMID: [32066887](https://pubmed.ncbi.nlm.nih.gov/32066887/).

- b. Hardisty A, Michener W, Agosti D, Alonso García E, Bastin L, Belbin L, Bowser A, Buttigieg P, Canhos D, Egloff W, De Giovanni R, Figueira R, Groom Q, Guralnick R, Hobern D, Hugo W, Koureas D, Ji L, Los W, Manuel J, Manset D, Poelen J, Saarenmaa H, Schigel D, Uhlir P, Kissling W. The Bari Manifesto: An interoperability framework for essential biodiversity variables. *Ecological Informatics.* 2019 January; 49:22-31.
- c. Poelen J. To connect is to preserve: on frugal data integration and preservation solutions. Annual Conference of the Society of Preservation of Natural History Collections. 2019;
- d. Meier A., Cooper L., Elser J., Jaiswal P., Laporte M.-A., Poelen J.H.. OOPS: The ontology of plant stress: A semi-automated standardization methodology. *CEUR Workshop Proceedings.* CEUR Workshop Proceedings; 2018; c2018.

D. Additional Information: Research Support and/or Scholastic Performance