**Creation of Integrated Name Resolution Service**

Dmitry Mozzherin, Richard Pyle, Yury Roskov

Scientific names are the metadata which allowed to aggregate and exchange biological information for the last 250 years. Such role of scientific names became even more important with advances in informatics. Biology is transitioning fast into the realm of “Big Data”. However connecting information via scientific names is not trivial, because of many spelling variants of the same name, evolution of binomial names via creation of new genus-species epithet combinations, homonyms, etc.

The resolution process of a scientific name consists of three stages. The first stage is lexical — it is necessary to figure out if a given spelling of a name belongs to a specific lexical group of name-strings. If yes – we are able to reconcile the given spelling with the previously collected information. Second stage includes finding of all “objective synonyms” of a name, which traces to all known nomenclatural events the name went through and determining its “basionym” corresponding to original description of the name in the literature. The third stage consists in determination of “subjective” synonyms and determination of currently used name according to a recognized taxonomic authority.

Global Names Architecture, which went through 2(3?) rounds of funding by NSF was created to solve first two stages — Global Names Index does lexical analysis and placement of a name-string into a lexical group, and Global Names Usage Bank traces nomenclatural events of a name to find “basionym” and all subsequent objective synonyms for a lexical group. Global Names does not have the final third component — taxonomical resolution which determines “subjective” synonyms of a name and determination of currently used name. The most advanced, state of the art project that addresses taxonomic resolution is the Catalogue of Life. It is estimated that Catalogue of Life contains taxonomical information about of all known species.

We are excited to propose an integration of Global Names lexical and nomenclatural reconciliation of names with taxonomical approach of Catalogue of Life to create a complete reconciliation/resolution workflow. We plan to cross-map data of Catalogue of Life with Global Names Usage Bank, build tools which will significantly speedup the completion of Catalogue of Life. We want to create a workflow which will allow both Global Names and Catalogue of Life exchange missing data and provide much richer services to their customers.

**Intellectual Merit**

**Broader Impacts**

**Results from Prior NSF Support**

**Previous Award Title** *award number* ; dates, $amount

Research carried out goes here

**PROJECT DESCRIPTION**

# Introduction

Text goes here

# A Section

More text.

# Another Section

More text. Cite an example

# Yet Another Section

# Time Line and Management Plan

# Summary: Significance of proposed work

## Intellectual Merit

## Broader Impacts

**BUDGET JUSTIFICATION**