Semantic Indexing of the Green Technology Patent Literature
An Application of the NamesforLife Contextual Index

George M. Garrity1,2, Charles Parker1, Dorothea Taylor1, Kara Mannor1, Catherine Lyons1
1NamesforLife, LLC, East Lansing, Michigan, US and Edinburgh, UK
2Michigan State University, East Lansing, MI

Background
Research on biodiversity, biosynthesis, and carbon sequestration moves from the laboratory into production or commercial environments, a number of important policy and business decisions must be made that demand correct information. These include establishing the patentability of a given technology, freedom to operate, and potential infringement of patents held by competitors, both in the U.S. and abroad. Failure to pay careful attention to these issues can have serious consequences beyond the payment of small penalties for infringement. These include lost opportunities for technology licensing, failure to detect and understand regional disparities, rapid growth in patent coverage of technologies by competitors and migration of technology across international borders.

N4L Nomenclature Model
To manage dynamic terminologies, we have developed a semantic model that represents names, taxa, and exemplar objects as distinct objects. NamesforLife uses a context-driven model of semantic resolution that is based on the rules of biological nomenclature, specifically bacterial nomenclature, but is generally applicable.

Indexing the Patent Space
NamesforLife, LLC has created a suite of software tools and techniques to manage dynamic terminologies using the underlying term set described in Tables 1 and 2, supplemented with links to the relevant taxonomic literature and key genetic and genomic information. The Company’s N4L tools can automatically detect and tag bacterial and archaebal names in HTML and XML documents with a high degree of precision. An interactive browser-based application (N4L Guide) provides end users direct access to correct nomenclatural and supplementary information that is served-up on demand while reading the literature. N4L tools use ISO-standard Digital Object Identifier (DOI) technology to create links at each occurrence of a validly published name in HTML documents. The company has also developed batch tools (N4L Semantic Tagger) that can embed N4L DOIs into XML versions of scientific articles that are created as part of the contemporary publishing workflow and used to create human readable content in various forms (e.g., PDF, rich-on-paper). The company has also developed a unique way of tracking the occurrence of biological names in the literature, based on the usage of our tools (N4L Contextual Index).

NamesforLife Taxonomic Model
The N4L Nomenclature Model
Manual taxonomic classifications and their corresponding terms are represented in the NamesforLife taxonomy, which is based on the published DICG (1980) and later editions of the bacterial names, the remaining exemplars are taken from Data Bases of Biological Nomenclature.

Nomenclature Project Status
At present, the NamesforLife database (N4LDB) contains 14,650 distinct names, 11,883 of which are validly published, 119 Candidates, and 47 that are illegitimate but relevant to the field. N4LDB also contains 14,939 metatags (representations of species, subspecies, strain), 9,461 of which represent distinct type strains for 11,511 taxa and 12,993 names, the remaining exemplars representing different types for the remaining 2,747 names are associated with higher taxa. The major classes of events that have occurred since publication of the Approved Lists in 1980, by event, are shown below. Less common events (journals, Revues Nominales, Accepted Names, Exemplars, Retractions, etc.) are not shown here.

N4L Contextual Index
The N4L Contextual Index is an interactive, taxonomically-aware web application is currently under development in cooperation with the Department of Energy Joint Genome Institute (DOE JGI). The tool is designed to help users detect, view and track links to existing nomenclature to current genomic bins via Crossref.

NamesforLife Semantic Services
A semantic indexing web service (NamesforLife) is available to provide an up-to-date nomenclature, strain and genome information, and a full bibliography. The screenshots below demonstrate the use of this tool on an OSM patent application. Instructions for installing and using this tool can be found at the NamesforLife services site, located at http://service.namesforlife.com.

Acknowledgments
We wish to thank B.J. Tindall, D.M. Amann, and J. Ezurie for their assistance in curating the underlying taxonomic and nomenclature information used in our models. This research is supported by the Office of Biological and Environmental Research of the U.S. Department of Energy under Phase I and II SRTR Awards DE-FG02-03ER63601 A001 - A005.