

# On the value of 'useless data': Infrastructures, biodiversity, and policy.

Stephen Slota, University of California, Irvine.

Geoffrey C. Bowker, University of California, Irvine.

## Abstract

As the ability to meaningfully process increasingly large quantities of data has improved, the need for systems to support the aggregation and subsequent use of disparate smaller datasets is correspondingly greater. This need is compounded in efforts attempting to understand the dynamics of global-reaching systems for the purposes of defining and guiding national and local policy efforts. The GBIF (Global Biodiversity Information Facility) is just one such project among a larger group seeking to aggregate the smaller, focused, and disparate sources of information generated for the work of science. GBIF is simultaneously an effort to coordinate and aggregate digital species occurrence data and digitize natural history collections into a single global-scale resource for biodiversity work. The GBIF database is a piece of infrastructural design work, serving as a functional basis for expanding scientific and policy work on biodiversity and conservation without functioning as a sole source for either. Management and maintenance decisions made in the creation of the database render it unsuitable as a primary source for many categories of biodiversity science and unable to provide the analytic tools for evidence-based policy decisions. However, the value of efforts like GBIF is not in their present suitability for a modern scientific and political tasks – there exist many databases and experts already who can fulfill those needs – but in providing a basis for future growth and in the development of new science and policy goals. GBIF as a social movement, as a political and scientific statement, and as the core of an infrastructure for global work in biodiversity science may well be more useful in guiding the course of global conservation policy than the data it contains. The basis for this paper is a critical study of the GBIF database and data portal as a socio-technical system.

**Keywords:** Conservation policy, biodiversity, GBIF, data infrastructure, values in design

**Citation:** Slota, S., Bowker, G.C. (2015). On the value of 'useless data': Infrastructures, biodiversity, and policy. In *iConference 2015 Proceedings*.

**Copyright:** Copyright is held by the author(s).

**Acknowledgements:** Thanks to Roderic Page for comments on the analysis.

**Research Data:** In case you want to publish research data please contact the editor.

**Contact:** [sslota@uci.edu](mailto:sslota@uci.edu), [gbowker@uci.edu](mailto:gbowker@uci.edu)

## 1 Introduction

As the ability to meaningfully process increasingly large quantities of data has improved, the need for systems to support the aggregation and subsequent use of disparate smaller datasets is correspondingly greater. This need is compounded in efforts attempting to understand the dynamics of global-reaching systems for the purposes of defining and guiding national and local policy efforts. The GBIF (Global Biodiversity Information Facility) is just one such project among a larger group seeking to aggregate the smaller, focused, and disparate sources of information generated for the work of science. GBIF is simultaneously an effort to coordinate and aggregate digital species occurrence data and digitize natural history collections into a single global-scale resource for biodiversity work. The GBIF database is a piece of infrastructural design work, serving as a functional basis for expanding scientific and policy work on biodiversity and conservation without functioning as a sole source for either. Management and maintenance decisions made in the creation of the database render it unsuitable as a primary source for many categories of biodiversity science and unable to provide the analytic tools for evidence-based policy decisions. However, the value of efforts like GBIF is not in their present suitability for a modern scientific and political tasks – there exist many databases and experts already who can fulfill those needs – but in providing a basis for future growth and in the development of new science and policy goals. GBIF as a social movement, as a political and scientific statement, and as the core of an infrastructure for global work in biodiversity science may well be more useful in guiding the course of global conservation policy than the data it contains. The worth of an infrastructural effort like GBIF, instantiated in its aggregation of wide swathes of species occurrence data, comes not from its immediate usefulness but from the creation of the social, political and technological conditions needed to support the development of such efforts at scale.

## 2 Reading the Database

Critical reflection on the design of information systems and other artifacts shows that humans embody their values and morality, often unconsciously, in the things that they create. (Winner 1980; Latour 1992; Hughes 2004; Nissenbaum, 1998) These values may be intentionally *designed into* the physical state of the artifact or system (Flannagan, Howe and Nissenbaum, 2008; Friedman, Howe and Felten 2002) or be observed resultant from and of a myriad of social factors. (Pinch and Bijker, 1987) These values can produce bias (Friedman and Nissenbaum, 1996) or otherwise be seen to have and carry politics of their own. (Introna and Nissenbaum, 2000) Successful infrastructures serve those with a variety of values, but may prioritize certain values in their design. (Knobel and Bowker, 2011) For example, mobile technology that automatically reports your location through GPS to your friends and family values connectedness and intimacy above privacy. Though these value propositions are evident in the objects themselves, often they are the result of unconscious assumptions on the part of the designer, making it quite difficult to avoid their potential negative impacts on quality of life. (Introna and Nissenbaum, 2000) Like other designed objects a database embodies a set of values in the design of its metadata, choices made about what sources of data to include, how provenance is represented, the means and experience of access and the processes of data management and preparation. A study of how the GBIF database and portal embodies a given set of values as a designed sociotechnical artifact is revelatory of the relationship between the scientific practice of biodiversity research and how it works to inform conservation policy.

Reading the database is a proposed method for interrogating the affordances and inhered values in the structural elements of the database itself. Important to this discussion in particular is the way in which the database speaks of itself and represents its content and capability. The database as an object is simultaneously the site of storage of information as well as a representation of that information as it acts in the world. The boundary of the database is the boundary of living data. What is and can be added to the database? What is being and can be queried from the database? How is information modified as it comes into contact with the database? What is the ecology of data use and creation for the database? All this work is relative to an imagined moment of retrieval, where the database is queried, and the data is accessed and in some way used or modified. Much as the database itself might be the site of certain forms of narrative, (Manovich, 2001) it can be bounded as a series of events revolving around a set of material and non-material objects and their bearers (which themselves may be material or non-material objects). (Faulkner and Runde, 2011) The database as object is defined by its events: the data events – particularly the assignment of metadata – become more important than the material and non-material objects that are their focus. As the characteristics of the database itself embodies to some extent the values and morals of its creator, GBIF – as both the database itself and the conditions for inclusion and retrieval – bears relevance not only to the technology needed to develop global-scope resources but to the relationship between biodiversity scientists, conservation efforts, and the work of policy-making. Biodiversity scientists already draw from heterogeneous data sources, across disciplines, recorded by a variety of methods and stored in many different ways. (Bowker, 2000) In many ways the GBIF data portal embodies this work practice, reporting species occurrences drawn from citizen science collections such as eBird (Sullivan, et al. 2009) alongside the collections of natural history museums and formally represented field data. It is in this examination of the characteristics of the database and its events that we explore the relationship between biodiversity science, knowledge production and policy.

## 3 Biodiversity, Conservation, and Political action

Bowker (2005) identifies two ways in which biodiversity science comes to relate itself to political action: implosion and particularity. Briefly, implosion is attempt to discover and apply value to various categories of living things, and particularity is the effort to exhaustively catalog and describe those living things. The field of biodiversity science as described here is a particularly lucid example of how a single database (prosaically and in effect, the earth itself and all its living things) must be ‘made amenable’ to either scientific work or policy work. While biodiversity science, with its implicit focus on conservation as the end goal of the scientific work, resides very close to the policy it informs, there is still evident the effort of and push towards rendering knowledge amenable to policy work. The biodiversity scientists in the modality of implosion of the database of the world work to render it amenable to valuation of the sort that would be useful to politicians, with all its attendant simplifications, obfuscations and hidden moral arguments. Given the scientist’s existence in an arena of implicit knowledge, jargon and other restrictions to participations, the desire to see translational work as a simplification (shedding of context) rather than a context shift distances analysis of the data from the data itself, management of the data from its analysis, the mobile objects of knowledge from shared understandings implicit to its creation. Categorization then potentially

become subservient to valuation, with disagreements over the ecologic role of a category of living things played out in its classification because that classification carries political value. As rationality forms the basis of legitimizing major areas of political action, there is political value to a simplified, mobile object (we've lost X species last year!) – with the side effect that these simpler objects may well obfuscate more than they reveal. The GBIF data portal is a particular (in this sense) effort of biodiversity science, and though it seeks to support the work of both scientists and policy-makers the metaphoric distance of the complete database from the data actually used in science is quite far. The global data portal renders amenable not the species occurrence data itself but the concept of global biodiversity mapping, the usefulness of having a global-scope resource, and the value of working globally on global systems.

Knowledge is produced and intentionally disseminated as part of both scientific and policy work and their agendas are mutually influential. (Irwin and Wynne, 1996) Just as objects of knowledge, specific data sets and scientific outcomes move between different established conventions of evaluation and representation within scientific disciplines so to do they change moving outside of vocational knowledge-work. The management, analytic and interpretive work a scientist does on and through their recorded data is temporal and contextual, just as is the similar work of a politician. “Knowledge is located in the nexus of participants, practices, artefacts and social arrangements,” and occurs and is validated within the context of a given community of practice, according to their own standards. (Van House, 2002, pg. 111) The boundary between scientific and political work can be seen to emerge more from standards of evaluation and the implicit claims of the utility of that knowledge than exist as a firm disciplinary divide. Data collected and publicized on the behalf of government is not fully in the territory of the sciences or of the politic – its release implies a benefit to information movement across, without and into society as a whole. The claim of societal good bridges the discursive work of the politician and the scientist, and it would seem the discursive power of the organizations and terms that emerge from this work is not fully held by either. While the workers in science and policy may be vocationally separated the co-construction of knowledge, relevance and epistemological discourse cannot be quite so easily disentangled. Scientists can act as advocates for given policy regimes or political philosophies, politicians make take an active role in setting scientific agendas and supporting certain technologies over others.

Biodiversity sciences can be seen as emerging from a belief in conservation on the part of biologists that is supported in, rather than developed from, scientific knowledge work. (Takacs, 1996) One of Takacs interview subjects, on the question of whether a species should be conserved, said,

“the answer is always 'Yes!' with an exclamation point. Because it's obvious. And if you ask me to justify it, then I switch into a more cognitive consciousness and can start giving you reasons, economic reasons, aesthetic reasons. They're all dualistic, in a sense. But the feeling that underlies it is that 'yes!'...” (Takacs, 1996, preface)

The practices present in the organization, structure and data practices of the GBIF data portal are formed in at least some part towards that conservationist goal – though there is little visible connection to specific policy, many of the design characteristics of the aggregated database support re-interpretation for policy goals often seemingly before scientific. The work of biodiversity science is closely entwined with the work of conservation biology, and both are difficult to fully separate from the set of values and political ideals that lead to national and international conservation movements.

#### 4 The GBIF Database

The GBIF database itself is part of a cluster of similar efforts to develop wide-ranging, globally scaled data centers for the study of systems that operate at larger scales. The Long Term Ecological Research Network (<http://www.lternet.edu/>), the Group on Earth Observations Network (<http://www.geongrid.org/>), the Global Earth Observation System of Systems (<http://www.earthobservations.org/geoss.php>) and EarthCube (<http://earthcube.org/>) are all examples of similar efforts in unifying heterogeneous data sources on the scale of the earth itself, rather than limited to specific regions. GBIF itself is an international effort undertaken by a variety of labs, governments, and data collections to assemble all primary biodiversity data into a single database. As such, we explore the case of GBIF as part of a study of the general move towards global-scale, international, cooperative infrastructural efforts evidenced by the continued growth and preponderance of these projects. In its current state, the GBIF database includes over 15,000 data sets from more than 600 data publishers with over 500 million species occurrence records. (“What is GBIF?”, 2014)

The database itself is heterogeneous in terms of sources with given data sets are unified by a controlled metadata structure and are automatically aggregated to a single storage site. Data can be generated at any of the participating sites, but is only included in the GBIF database when it is cleaned and organized in such a way to conform to the metadata and file format standards decided by the GBIF participating members. However, the distributed nature of data production opens up a certain set of problems that characterize many informatics ventures. A particular issue is that of taxonomy: "In general, obtaining reliable, updated taxonomic authority files is a major problem for most taxa. The large taxonomic information services... remain far from complete." (Soberon and Peterson, 2004, pg. 692) And even when completed, taxonomies are not static – the information facilities must be maintained as new consensus is reached and new debates emerge. Soberon and Peterson (2004) point to several other major limitations in biodiversity informatics, including improper identification of specimens, outdated taxonomy and faulty georeferencing in electronic collections. Data submitted to GBIF undergoes an automated form of cleaning in order to account for some of these problems – particularly that of faulty georeferencing. Georeference data is excluded if it is exactly on the equator (0 for latitude coordinate), exactly along the prime meridian (0 for longitudinal coordinate), or a certain distance away from the coastline. These data points are excluded to account for what is assumed to be the common user error in providing location data on specimen sightings – accidental exclusion of one or both points in the coordinate pair, and accidental reversal of latitude and longitude. The cleaning process is automated, and can only account for good-intended errors – other georeferencing errors can be much harder to correct, or even find, without deep exploration of the data as it is being submitted to the database. The distributed nature of GBIF is such that significant trust is placed in the scientific standards of the participant nodes – which is the primary method for accounting for the other potential problems (incorrectly identified specimens and outdated taxonomy) in biodiversity informatics as discussed by Soberon and Peterson.

Rigid metadata standards and the necessity of transforming the researcher's own database in order to allow its inclusion (often while changing the way the data had originally been organized) limits the number of smaller efforts – particularly those with limited staff – that can participate. Yesson, et al. (2007), in analyzing the coverage of the GBIF database, found that it was substantially accurate, though coverage was lacking in several significant areas. The database does not cover well biodiversity hotspots, its collection is dominated by a few very large data sources, and many biodiversity "hot spots" are not covered at all. Hot spots are places with significantly wider varieties and density of species representation and overall represent areas of particularly dense biodiversity and are often pointed out as areas in most need of conservation work. (Mittermeier, et al. 2005) Important to the study of biodiversity is in the analysis of global patterns of movement, migration and growth. It is because of this that there even exists the attempts at a global database like GBIF. However, the exclusion of hot spots and some portion of oceanic data (eliminated in some areas to avoid "data shadowing" caused by accidental transposition of coordinate pairs) can serve to limit the database to already well-known information. According to Yesson et al. "it is possible to retrieve large numbers of accurate data points, but without appropriate adjustment these will give a misleading view of biodiversity patterns." (Abstract, 2007) Establishing standards for metadata creation and assignment, establishing acceptable synonyms and namespaces, and relying on the scientific authority of participant nodes eliminates some significant sources of error, but also serves to exclude smaller scientific efforts, certain classes of citizen science (though not all – see Sullivan et. al, 2009), and scientific institutions that may not use a similar standard set or work methods.

More important is the claim that the GBIF database and data portal will eventually underpin policy, and though there are a couple of efforts underway working to establish a better relationship between the biodiversity data and policy recommendations, there is little in the infrastructure itself to establish the policy or conservation relevance of certain data. The GBIF database's production is decentralized with central control of metadata and formatting standards for the data produced by participant nodes. Data produced and submitted to the database is primarily from Western Europe, North America, and Australia with no data at all from China and minimal coverage of such biodiversity hot spots such as those in South America and Africa. (Yesson, et al. 2007; Collen, et. al 2008) Even as a proof of concept, a single global database that underpins policy formation, problem framing, and decision making can be a troubling notion. The designation of a global biodiversity information source implies a singularity to the database – it implies global coverage that is not necessarily global. The vast majority of species and global biodiversity is in microbial life and oceanic life, but these are not the species of political or conservationist concern. The GBIF database does not represent microbial life in a substantial way, and traces oceanic life only so far away from the coast of a given continent due to their data sanitation methods. Consistent

with other movement towards including human action in studies of ecology, (Ribes, 2014) it is important to remember that biodiversity and conservation is at its core about the human ecology, and that is replicated in the data captured and presented.

The singular, global database has the tendency to undermine the conclusions of its scientists by oversimplifying the richness of scientific debate that generated that data. A significant portion of the GBIF metadata structure is the current (to the time of observation) taxonomy of the observed life form. The taxonomy of data points within the GBIF database is static: once data is submitted there appears to be little way to represent changes or new findings in taxonomy. Given its ultimate reliance on species and taxonomy as the primary descriptive metadata of its collections, the GBIF database is oddly silent about the work being done by biologists and geneticists who are challenging and debating the very notion of species as well as changing and making new discoveries in individual taxonomies. Once the observation is entered into the database, it is removed from the field of scientific debate that is occurring even among the scientists that generated that data. While this can be ameliorated somewhat by synonymizing and in the creation of namespaces, building these concepts into a database that is attempting to be global has the apparent effect of formalizing concepts and structures that are still far from consensus within their fields. In addition, namespaces, though various equivalents have been around for over 150 years, are very difficult to check against each other, and establishing species equivalence and distant locations represents substantial scholarly effort. (Bowker, 2005) While this might not be an explicitly political goal, the appearance of consensus that emerges lends itself to more apparent credibility in pursuing species conservation. To a large extent, the taxonomic and other scientific work does not need to be represented in calls to action in protecting certain areas or species. "The credibility of regulatory science ultimately rests upon factors that have more to do with accountability in terms of democratic politics, than with the quality of science as assessed by scientific peers." (Jasanoff, 2003)

Gaps in GBIF's coverage can, to a certain extent, be explained by the availability of funding for certain scientific efforts over others. "But the U.S. has a lot more money and media than Thailand, so a species or subspecies of vertebrate in North America is likely to get more resources and attention than the possible extinction of the entire Family Craseonycteridae." (Kinman, 2002) In addition, the amount of work necessary to verify and clean the millions of species occurrence data points by hand would be non-trivial in terms of time and opportunity cost. And in fact there is little need for such complete coverage. The structure of the database itself alongside the goals of its creators reveals a reliance on a scientific advisory system (common in the EU, UK, Australia and America) in the creation of biodiversity policy. The central political theory that is implied in the creation and accretion of the GBIF database relies on a relationship between policy and science that is somewhat simplified – it holds to the "traditional view" presented by Almeida and Báscolo (2006). The production of data (or scientific knowledge) is seen as an accumulation of useful facts – the greater the number the more useful – that policy makers are able to draw on when making decisions. In order for the GBIF database alone to underpin policy decisions policy makers must be (near-ideal) privileged rational actors capable of effectively interpreting scientific results as part of a series of strictly evidence-based rational decisions outside of economic, public or other political influences. In the field of health policy, there is (according to Habricht, et al., 1999) a three-fold requirement of adequacy, plausibility, and probability that ought to be met in allowing scientific evidence to inform policy decisions. Though I will not spend much time analyzing these concepts, it is important to remember that the three-fold requirement is deeply contextual and relative – it seeks to establish a formal structure for evaluation because one is needed as a basis from which to work. Much like the GBIF database itself, inconsistencies in coverage and need for additional expertise is seen as insufficient reason to abandon the development of such an effort – it is a basis from which to work rather than an ultimate authority of its own. The data in the database is present for interpretation, it cannot represent knowledge in its own right except in the assumption of substantial knowledge and skills required for interpretation of that data. There is little in the structure or nature of the database or data portal to indicate any translational relationship of the data to a policy maker or debate – those tasks fall once again to diplomats of science.

Calling a database global implies consensus where consensus might not exist. Expecting that database to underpin policy decisions requires a significant, but in the case of GBIF, non-obvious translational step. The GBIF data portal reduces its findings to a single number: at a certain geographic point there are a certain number of species observations. A certain species (not individual, not colony, not family unit or nest) has been observed some number of times. While there exists data structures to report the time of

observations, these are not required for entry in the database – some records contain no time of observation at all, and others include no georeference. Movement of species and change in species over time is an observation that is not built into the database itself – these things can only be interpreted through other scientific knowledge and an additional act of interpretation. A policy decision, such as whether to preserve one plot of land over another, requires that interpretative act to be performed, as well as significant amounts of other information that is not and (under current metadata standards) cannot be represented in the database. There is no place in GBIF for ecology: it does not render relationships between species, simply numbers of observations. While it may be said that this is outside the intended scope of the database, it does raise a question: if meaningful interpretations of the data require the scientists to have a sufficiency of local knowledge that is not represented within the database itself, what is the benefit of making the global connection? For policy makers and other scientists, the GBIF data can only be a starting point, a simplified visualization of what kind of biodiversity work is being done, not even the state of biodiversity globally or locally. While the data can show where a certain endangered species has been or might be seen, it cannot say what will happen if that area changes. It cannot say whether nearby changes will affect that area. It is solely descriptive, and does not represent change either over time or in scientific understanding. This lack is not an oversight, it is merely an acknowledgement of the human components of the infrastructure, the interpretations and expertise needed to allow the data to underpin policy.

## 5 Conclusion - GBIF and the “Long Now” of infrastructure development

“I am increasingly convinced that the study of biodiversity is far and away the most important endeavor in the history of humanity, certainly until now, and very possibly into the future as well.” (Pyle, 2010) Large-scale data aggregation efforts seek to develop pieces of the infrastructures that support both scientific work and policy work – they do not represent the end of the work but a tool towards its completion. While the GBIF data portal right now does not well represent either cutting-edge taxonomic work or the controversies and uncertainties that characterize scientific work it does provide a basis for future aggregation of data and representation of that work. Infrastructure work is subject to the ‘long now’ of development, where concerns for future development and maintenance are considered concurrently with daily work. (Ribes and Finholt, 2009) While the GBIF data portal may be seen as not immediately complete, its aggregation methods, partnership agreements, and overall goals exist as much as a basis for future development as they do to represent current scientific consensus. In fact, the presence and possibility of namespaces for translation (flawed though they may be) makes the database itself both responsive and somewhat resilient to shifts in taxonomic convention. Efforts like GBIF cannot be effectively evaluated in the present tense, as they are not designed for the short term needs of their communities, but rather exist as physical, designed instances of the values, goals and predictions of their creators and contributing communities. Inconsistencies or inadequacies in coverage are not a failing of the data itself, nor of the database, but a clear indication that the database exists in the long present of development – and will likely never be ‘done’ or a closed topic. The GBIF data portal exists evidently in its structure, design and organization in the future, not as a completed entity in and of itself but as a resource, map and object lesson for the development of such complete records and relationships.

## References

- Almeida, C., & Báscolo, E. (2006). Use of research results in policy decision-making, formulation, and implementation: a review of the literature. *Cadernos de Saúde Pública*, 22, S7-S19.
- Bowker, G. C., 2000, Biodiversity datadiversity. *Social Studies of Science*, 30, 643–683.
- Bowker, G. C. (2005). *Memory practices in the sciences* (pp. 269-298). Cambridge, MA: MIT Press.
- Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: addressing disparity in global monitoring. *Tropical Conservation Science*, 1(2), 75-88.
- Faulkner, P., & Runde, J. (2011). The Social, the material, and the ontology of non-material technological objects. *27th EGOS Colloquium*, Gothenburg, Sweden, 2011.

- Flanagan, M., Howe, D., & Nissenbaum, H. (2008). Embodying values in technology: Theory and practice. *Information technology and moral philosophy*, 322-353.
- Friedman, B., Howe, D. C., & Felten, E. (2002, January). Informed consent in the Mozilla browser: implementing Value-Sensitive Design. In *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on* (pp. 10-pp). IEEE.
- Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, 14(3), 330-347.
- Habicht, J. P., Victora, C. G., & Vaughan, J. P. (1999). Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *International journal of epidemiology*, 28(1), 10-18.
- Introna, L. D., & Nissenbaum, H. (2000). Shaping the Web: Why the politics of search engines matters. *The information society*, 16(3), 169-185.
- Kinman, Ken. (2002, August 21). Library on Fire (triage). [Electronic mailing list message] retrieved from <http://mailman.nhm.ku.edu/pipermail/taxacom/2002-August/041477.html>
- Knobel, C., & Bowker, G. C. (2011). Values in design. *Communications of the ACM*, 54(7), 26-28.
- Latour, B. (1992). Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. In *Shaping Technology/ Building Society*, edited by W. Bijker and J. Law, 225-258. Cambridge: MIT Press.
- Nissenbaum, H. (1998). Values in the design of computer systems. *Computers and Society*, 28(1), 38-39.
- Manovich, L. (2001). *The language of new media*. MIT press.
- Mittermeier RA, Gil PR, Hoffman M, Pilgrim J, Brooks T, et al. (2005) *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions*. Mexico City: Agrupacion Sierra Madre, Mexico
- Pinch, T.J. and W. E. Bijker. 1987. The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. In *The Social Construction of Technological Systems*, edited by W. E. Bijker, T.J. Pinch, and T. P. Hughes, 17-50. Cambridge: Mass.
- Pyle, Richard. (2010, January 27) [Taxacom] Was Reproducibility of phylogenetic analysis. [Electronic mailing list message] retrieved from <http://mailman.nhm.ku.edu/pipermail/taxacom/2010-January/067852.html>
- Sengers, P., Boehner, K., David, S., & Kaye, J. J. (2005, August). Reflective design. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility* (pp. 49-58). ACM.
- Soberón, J., & Peterson, T. (2004). Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1444), 689-698.
- Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10), 2282-2292.
- Takacs, D., (1996) *The Idea of Biodiversity: Philosophies of Paradise*. John Hopkins University Press, Baltimore, MD.
- Van House, N. A. (2002). Digital libraries and practices of trust: networked biodiversity information. *Social Epistemology*, 16(1), 99-114.
- Winner, L. (1980). Do Artifacts Have Politics?. *Daedalus*, 109(1), 121-136.

Yesson, C., Brewer, P. W., Sutton, T., Caithness, N., Pahwa, J. S., Burgess, M., ... & Culham, A. (2007). How global is the global biodiversity information facility?. *PLoS One*, 2(11), e1124.