"To Infinity and Beyond!": Inner Tensions in Global Knowledge Infrastructures Lead to Local and Pro-active 'Location' Information

Ayelet Shavit

Department of Interdisciplinary Studies, Tel Hai College, Israel / ashavit@telhai.ac.il

Yael Silver

Department of Interdisciplinary Studies, Tel Hai College, Israel

Abstract

We follow two biodiversity knowledge infrastructures that hold conceptual and practical inner tensions, and we argue that some of these difficulties emerge from overlooking local information and different understandings of the term location. The ambiguity emerges from two basic concepts of space – exogenous and interactionist – that are both necessary yet readily suggest inconsistent practices – global standardization and local flexibility – to organize location records. Researchers in both infrastructures first standardized, digitized and globalized their records, then discovered inner tensions, and finally alternated between globally interoperable and locally flexible records. Our story suggests a broader lesson: since both types of 'location' information are necessary; and since vast resources were already invested in globalizing knowledge infrastructures; then investing in local knowledge infrastructures and in alternating between both types of memory practices seems the most rational option, and a good way to resist epistemic injustice afflicting local knowledge in peripheral localities.

Keywords: biodiversity, database, epistemic-injustice

A Brief Introduction to 'Location' Uncertainty

As these four special issues have argued, 'knowledge infrastructure' is a fundamental emerging concept in Science and Technology Studies encompassing a variety of definitions and case studies, with identifiable common threads across this rich diversity (Karasti et al., 2016). For the purpose of our argument here 'knowledge infrastructure' is broadly construed, as resources in a network form (Bowker et al., 2010: 98); and according to Dagiral and Peerbaye (2016: 45): "This definition departs from the conventional representation of infrastructure as a mere machinery of "tubes and wires", to include a wide range of technologies and organisations that span large-scale sites and instruments devoted to scientific research".

A number of cases studies – especially those linked to involvement of marginalized groups (Jalbert, 2016; Silver & Shavit, in press) – have considered 'knowledge infrastructures' in this broad sense, as do the two case studies elaborated in this article. One basic and necessary element within nearly any knowledge infrastructure is spatial information. Elsewhere, we have shown (Shavit & Griesemer, 2009, 2011) that 'location' – perhaps the most basic and mundane term in science – holds a basic ambiguity. While employing the same term – 'location' – rigorous records of a biological process use two different concepts of space – *exogenous* and *interactionist* – committed to different epistemic values – generalizable representativeness and comprehensive accuracy – that readily suggest inconsistent modes – global standardization and local diversity and flexibility – to organize location knowledge (Shavit & Griesemer, 2009, 2011). This basic ambiguity is especially relevant for long-term knowledge infrastructures, which have been shown to tackle inconsistent information organization on multiple aspects and dimensions (Karasti & Baker, 2008; Karasti et al., 2002, 2006, 2010). 'Location' ambiguity across long-term studies hinders reliable repeatability of an experiment or survey (Shavit & Ellison, in press) and reproducibility of its results (Ellison et al., 2006).

An *exogenous* concept of space stipulates that organisms' effects on their locality[1] - via their social structure, physiology, metabolism, behaviour and history – and can be safely ignored for successfully modelling and predicting their distribution (Guisan & Thuiller, 2005). An alternative, *interactionist* concept of space, stipulates that these aspects cannot be ignored since organisms and their environments are mutually co-determined.[2] Adopting a certain concept of space signifies a commitment, i.e. an actual expenditure of resources (Gerson, 1998: 25), to certain types of values, of a rational and social character (Longino, 1990, 2004), and entrenched working procedures to coordinate the labour by using technology, i.e. computer-supported cooperative work (CSCW) (Gerson, 2007).

An exogenous space is committed to revealing general distribution patterns, hence valuing *representative and generalizable* data; on the other hand, an interactionist bio-space values a *comprehensive* and *accurate* data-set for a particular location.

An example of an exogenous partition of space is regular quadrats according to randomly-chosen longitudes and latitudes. Human investigators define a system of grid lines – latitudes and longitudes – conventionally located with respect to the Earth's poles, equator, and Greenwich, England as prime meridians, with elevations above or below sea level decided at some arbitrary date. Organisms are located in this framework regardless of their specific behaviour or metabolism and independently of the *existence* of that conventionally imposed description. The organisms do not attend to nor can they exploit, their "lat/long".

An example for an interactionist partition is a polygon of borrows or landscape patches in accord with a gopher's activity or tree's presence. Under this concept, the organisms themselves causally contribute to the organization of the space in which they live. An organism's position will causally depend on, or bear significant relation to, its interaction with its environmental context, i.e., of places modulated or constructed by what the organisms in questions do, what their neighbouring species do, and without regard to the conventions of humans that might study them. In this sense, space becomes also the *product* of the interaction of the organisms and *their* environments.

This article emerges from an on-going involvement with two long term biodiversity case studies. The first case study was mainly conducted by following the MVZ's (Museum of Vertebrate Zoology) surveys across California between 2005 to 2008 yet our research continued until 2013; in the second case study we followed a national survey of "Hamaarag" across Israel from 2004 until 2015. There is no explicitly written method for an involved philosopher of science, but it is an active, interdisciplinary and long-term line of work that builds upon the two basic meanings of 'involvement': care and active engagement. In practice, it means a joint research process of several years, where the philosopher produces a description of a scientific research process based on active participation in routine scientific fieldwork – e.g. collecting spiders, writing trapping results etc. – and asking the scientists working beside her questions that are motivated by their mutual care – not necessarily agreement – on how best to obtain the goals of this particular scientific project and how the project's knowledge infrastructure will best represent it. In addition to fieldwork, the philosopher also sits on Principle Investigator (P.I.) meetings, recording these meetings while intervening with questions that invite the scientist to critically reflect on her description and analysis qua conceptual theoretician. For corroboration, repeated individual interviews with the scientists, each focused on practical understanding of one core concept, were recorded in addition to notes being taken. Each interview lasted one to two hours (typically the latter) and its findings were re-visited throughout the years in order to track conceptual changes. Overall, there were 9 scientists working on the MVZ's project, for which 25 in-depth interviews, 21 P.I. meetings and 6 long fieldtrips were joined. In addition – sometimes in parallel – 9 scientists leading the Hamaarag surveys were similarly followed, via 23 interviews, 10 P.I. meetings and 4 fieldtrips. Knowledge gained from the fieldwork, meetings and interviews initiated a historical examination of how a particular scientific practice and concept came to be. For example, observing a certain method being used in the field, and hearing its rationale of use, triggered a study on its original context of use and disuse. This micro-historical work was done in the MVZ's archive [3], presented online and its hard-copies located in the museum's main gallery, in addition to asking the American or Israeli participating scientists to send all their old emails and meeting minutes regarding that research project and research method. These historical results were later brought back to the scientists for short reflections upon their original thoughts and rationalizations.

Such an involved method may be relevant to scientists, HPS (history and philosophy of Science) and STS (Science and Technology Studies) scholars, as well as to any academic who seeks a more pro-active and interdisciplinary academia. Regarding pro-activity, since biodiversity researchers are often involved in conservation and public engagement, an involved method can easily lend itself to resist epistemic injustice. Epistemic injustice (Fricker, 2007) is a wrong done to someone due to a biased perception of her capacity as a knower. Production of knowledge by academic and laypeople working side by side with mutual recognition gives room for local knowledge that is often silenced, and hence it is one way of resisting epistemic injustice. We also ask whether certain ways to organize the data – e.g. top down versus bottom up – lend themselves more easily to such an involved research.

In the following case studies, the first presents a more exogenous concept of space and a more top down mode of organizing location information while the second – a more interactionist concept of space and a more bottom up mode of organization – yet the scientists in both cases thought well in advance about their knowledge infrastructure. Both found these two concepts of space necessary for an accurate and generalizable location record, yet both first invested most of their informatics resources in fitting their data to a single interoperable data model and later recognized its inherent tension. Both resolved their 'location' uncertainty data in a manner that emphasized the interplay of local workarounds alongside universal interoperability – instead of choosing one or the other – which eventually opened new possibilities for scientific research as well as for resisting epistemic injustice.

Case Study I: The Museum of Vertebrate Zoology

A History of Methodologies in a Natural History Research Museum

The Museum of Vertebrate Zoology (MVZ) was established at the University of California, Berkeley in 1908 by the patron and entrepreneur Annie Alexander and the scientific director Joseph Grinnell (Stein, 2001). Grinnell noticed the rapid demographic and economic changes in California, argued that theses trends unfold a natural experiment in species distribution and evolution (Grinnell, 1917), envisioned his museum as a supplier of facts for describing these changes, and guided by his expert advice on how best to handle them, he described an aim of: "serving as a bureau of information within our general field" (Grinnell, 1935: 2). More specifically, the museum researchers and students were to conduct a series of rigorous descriptions of species and sub-species distributions in the same location over time "with application of the 'laboratory method' out of doors as well as in the Museum" (Grinnell, 1935: 1). The laboratory provided a global method, a "placeless location" (Kohler, 2002), and applying this universal standard to specific places (Kohler, 2012) and to the idiographic narrative style of natural history research had just began. Grinnell was so keen on implementing such new technologies that he defined it as one of the duties of a museum director: "Be alert for improvement of methods in every department" (Grinnell, 1929: 5).

In line with this duty, a huge effort was devoted by Grinnell and the MVZ staff to build standardized, detailed protocols for almost every aspect of work in the museum (down to the kind of ink and paper to use). There was an 8-page written standard for recording observations in a field note journal (Grinnell, 1938) and yet another 5-page protocol specifying the structure of species information on small tags and index cards (Wythe, 1925). This minute procedural decision to distinguish between two techniques to record a species' location – open-ended field notes versus standardized cards – is a crucial point in our story, one we shall return to.

Diligent execution and updating of this distinction enabled the MVZ to function for: "the promotion of wildlife conservation and management on a biologically sound basis of fact and principle," (Grinnell, 1938) and "to establish a centre of authority on this coast" (Grinnell, 1907). The MVZ as a whole functioned in ways aptly described by Latour's (1999) 'centre for calculation', and its specimens as powerful 'boundary objects' (Star & Griesemer, 1989).

In 2001, in preparation for the museum's upcoming centennial the museum vision was re-visited and the idea of a "Grinnell Resurvey" was born (Senior staff, March 28, 2006 and May 1, 2006 interview). Studying this resurvey reveals some of the basic commitments and values entrenched in practice of MVZ researchers and information managers. The MVZ's tradition values rigorous and self-recorded work style. When a trap line is set in the field its specific setting and its method and effort of study are all meticulously recorded in one's field notebook *journal*. There one describes -– and if possible quantifies – properties of the specific locations encountered throughout that day: their landscape, weather, snow level, dominant plants, soil, sampling method and the effort of detection.

In addition, the MVZ held an extensive collection of material objects, i.e. specimens, tagged and stored in cabinets. The tag, sometimes called specimen label, is a small piece of paper attached to a specimen in the field. The tag was the crucial evidence guiding the handling of the specimen later on, upon its arrival at the museum, and its structure and content was specified and standardized (Wythe, 1925).

Once the specimens were brought in from the field, their location as indicated on the tags was entered into the MVZ's collection in the format of index cards and was never supposed to be changed or corrected, "and so, reversely the student [of today] may quickly trace back again from any particular specimen its history, by referring to the card catalogue and field notebook" (Grinnell, 1910: 35). Changing the card wording might break this chain of reference (Gannett & Griesemer, 2004; Latour, 1999). For Grinnell, a specimen without such contextual information is considered "lost. It had, perhaps, better not existed" (Grinnell, 1921:108). To add visual context, thousands of photographs were taken (of habitats, localities and specimens) and hundreds of maps were drawn. All these items were stored in the MVZ archives and all are traceable to each individual specimen stored in the collection, since, Grinnell argued, we never know what type of record will be required in the future (Grinnell, 1910: 34-35).

Grinnell stressed the need to use *both* the narrative, local description in a field notebook journal and the standardized description on a small specimen tag, yet he introduced this distinction only to facilitate the widest utility of collected material. Although standardized information might be sufficient for some taxonomic purposes, the narrative notes might be of broader significance to studies of ecology, evolution and conservation —specimens merely documenting the presence of a given species in an ecological context (Griesemer, 1990).

After Grinnell's sudden death in 1939, surprisingly, little has changed in the Museum's methodology. The primacy of an abstract, context-free point on a universal and standardized grid of longitudes and latitudes, referenced by a number with an unequivocal interpretation, began only when the museum collection was digitized. Throughout the late 1970's the MVZ collection records were entered into a computerized database and by 1998 it was the first collection of modern vertebrates in the world to go online.

One of the forces motivating computerization of records was the passage of several environmental laws in the first half of the 1970s. 'The National Environmental Policy Act' (NEPA), signed on January 1, 1970 by US President Richard Nixon, required that a statement assessing environmental impact (EIS) on species must be filed prior to any major US federal act. The Endangered Species Act (ESA), signed by Nixon on December 28, 1973, likewise created a need for information about species distributions for land developers and business entrepreneurs. Soon thereafter a boom of private companies specializing in assessing environmental impact emerged, and they started arriving at museum collections looking for information. In 1972 the American Society of Mammalogists responded by establishing a committee on Information. That committee, which included an MVZ representative, established a common set of standards for database development across all American collections. In the same year, the NSF founded a new program under which museums could apply for funding of cabinets, fumigation equipment, etc. to maintain their collections.

However, if the MVZ were to continue its role as a "centre of authority," it not only had to store information but also to supply it quickly and efficiently to the public. Luckily, the technology to do just that was already spreading in the life sciences. Mainframe computers became routinely used in the mid 1970s, and the NSF responded by expanding its existing funding program to include information technology. The director of this NSF program, William Sievers, encouraged James Patton of the MVZ and Philip Myers of the University of Michigan to jointly propose a grant to computerize the MVZ's and the University of Michigan's collections and make available a database management system for all other museums. In 1978 they received an NSF grant for retrospective capture of information on the Mammalian collection.

The grant compelled the museums to decide on the types of information to record in the database. Given that the free-text locality information of the field journal would be hard to code in a systematic way, decisions about what information to record in the database entailed trade-offs in future searchability of information about locality and required, in turn, a decision comparing the relative significance of different types of 'location'. Specifically, and practically, the question of what location information to code in the database was whether 'locality' information would be extracted from the field journal, the index card or both? It was then, for the first time, that an implicit *commitment* was made to a *single* concept of space – exogenous from the local landscape and its inhabitants rather than sensitive to it – for recording a species 'location' in the database. From then on, ever-increasing resources were allocated to recording an exogenous concept of location.

One reason for that choice was informatics-based. The information that the database software (TAXIR: Taxonomic Information Retrieval) could query needed to be highly standardized and organized within a single table ("flat file"), in addition to taking as little space as possible, given the processing power and storage limitations of 1970s mainframe computers. The short, standardized descriptive locality recorded on the specimen tag fitted that technical demand nicely, while the intertwined, context-dependent, free-text record in the field journal could only be stored but not searched or queried in a flexible manner. However, the main reason to leave aside the localized field notes did not involve software or hardware. It was the legal and economical burden the EIS's and ESA's put on the protection of species (rather than niches or habitats as Grinnell and others recommended (Grinnell, 1910), hence the NSF's explicit interest – and consequently Patton's and Myer's explicit focus in their proposal – in the *specimen* collection, which – by Grinnell's own stipulation– was available first and foremost from the specimen tag record. For that purpose, the field journal lacked information – such as museum catalogue or accession number – and held vast ecological information that was time consuming to retrieve.

In 1980 the MVZ's database became operable. That is, a person sending a question by mail – e.g. which species were found in Yosemite National Park – could receive a written answer within a few days after his query was entered into the mainframe computer. As a result, queries about a taxon – e.g. genus, species, sub-species – found at a certain point on a map could be answered quickly, while all the environmental, geographical and historical information contained in and distributed among the field journals about that species at that time/space point could not, because it was not machine searchable. De facto, this meant, according to anecdotal comments of current MVZ staff members, that queries about the extensive locality records stored in the field note journals were reduced from that point on.

"Backgrounding" this large source of ecological information did not raise any complaint from most database users concerned with species distribution questions. This implied that an abstract point locality became not only necessary but also *sufficient* for many queries utilizing the museum collection. To be sure, some behavioural ecologists and systematists interested in small-scale questions still routinely read field journal information – typically photocopied and mailed to them by an MVZ curator – yet most queries relied on the database as the primary, and sometimes only, way to describe species location.

In 1997 a new programmer analyst presented a new, relational data model for the collection. This database defined not only multiple search attributes for each specimen record – e.g. its location and name of collector – but also defined relations between these attributes, such as: when, where and who collected that specimen. A relational database allowed flexible queries, and was designed to be complete, i.e. contain records of all specimen tags alongside field journal entries, photos, maps and more. Yet, however ambitious and carefully planned, the database's data model could not interoperate with such open-ended records as the field journals.

In 1998 an online database system was jointly developed with the Alaska museum. "Arctos" is still the largest multi-institutional database of natural history research museums, integrating data from thirteen universities. Now that anyone with internet access could quickly and efficiently query the collection, many more did so, yet only queries about location that assumed a regular grid with standardized meanings for each term, unequivocally (and automatically) assigned to a set of data fields defined by the data model, could be answered by Arctos. The specimen tag records, along with lat/long coordinates, fitted these requirements, while the field journal descriptions did not. As seen, Grinnell's original tags did not mention lat/longs and typically referred to the area around the campground (sometimes even to a whole county). To improve the resolution of these location records in the database, the programmer analyst developed a sophisticated georeferencing algorithm and protocol, which allowed one to assign a GIS map point with a maximum error distance (degree of uncertainty) to each historical locality in the collection (Wieczorek et al., 2004). Finally, a standardized location point seemed to be comparable with current and future location recorded by GPS lat/long methods. It was hoped that whatever uncertainty remained could be reduced by reading the field journals (by now scanned and posted online, but still not searchable), applying auxiliary information to the georeferencing procedure, and thus shrinking the error distance around each point.

Thirty-five natural history museums worldwide record localities via this georeferencing protocol created at the MVZ, attesting to the overwhelming entrenchment of one concept of space as sufficient for recording a location outdoors: an abstract, universally standardized and biologically-exogenous point on a GIS map. Problems arose, however, when someone had to actually *replicate* a visit to the same outdoor location years later by following these lat/long coordinates. This line of fieldwork at first did not turn 'location' into a problem, but only meant more work for those diligent researchers who went the extra mile and interviewed old collectors or read old field notes. What MVZ staff often called "the problem with locality" (Shavit, observation during weekly Resurvey meetings during 2005-2008) did not arise until 'replication' became an institutional problem, i.e. until the "Grinnell Resurvey" project demanded in the spring of 2003 an actual return of various researchers to hundreds of survey sites across California after nearly a century.

From an informatics infrastructure perspective, the late 1990s and early 2000s seemed like the right time for such a move, as new computer technologies became available in the field. For measuring a locality, GPS receivers had become cheap enough to replace the heavier combination of map, compass, and altimeter; and for recording locality information, Palm Pilots and laptops equipped with spreadsheet software increasingly replaced handwritten field journals. The new technologies produced mostly numbers and abbreviations instead of narrative free-text descriptions. These new tools became extensively used in the Grinnell Resurvey project, and consequently the protocols for recording 'locality' in the MVZ were changing in important ways, some of them creating new challenges.

One must record new GPS data fields, e.g. precise longitude and latitude, datum, and device accuracy. This makes sense: without such GPS data-fields, using GIS mapping systems is unreliable, and without GIS maps computers are limited in power to represent and predict species distribution. However, this can also produce a common – and often unnoticed – problem. An MVZ senior naturalist explains: "…if a locality couldn't be located at a [GPS] geographic scale sufficient to be usable by the scale of the GIS layer [representing the spatial distribution of variables such as temperature, precipitation or elevation], then the model derived by the combination of those different data would likely be in error, *the extent of which would not be known*. Georeferenced localities can thus give a false sense of security, unless they are located at a scale appropriate to the other information with which they are associated" (Information manager, interview on September 3, 2008).

To allow interoperability between the georeferenced and the field journal's 'location' descriptions, the journal's information was mined and transformed to a standardized format. Locality information that was sensitive to a given species in a particular ecological and social context was transformed into a set of tables and data fields, each with a standardized meaning and structure. Moreover, location information previously readily integrated with species locality information, such as habitats across the trap line, is now separately mined in order to be incorporated into the database. The increasing prevalence of data standardization in current museum work led most MVZ researchers to record what they regarded as their most important data, in private spreadsheets – the analogue of the old field journal, although they were aware that such data are very likely to become inaccessible after a few years due to obsolete software or lack of metadata.

The net effect of these technology-induced changes in practice and in protocols for data-mining the field journal, actually deepened the gap between these two concepts of space, one exogenous to the research subjects but readily coded in the museum's online information infrastructure, the other sensitive to the subjects and their context, but hard to code and not interoperable between information systems. The result of this data-mining process was several databases on different locations (e.g. Yosemite National Park, Lassen Volcanic National Park, etc.), which, in contrast with implicit expectations, did not successfully link to the main MVZ database. Why? Because history matters: these local databases originated from the notebook narrative culture while the data model of the database originated from structured tags; each type of record was recorded at different stages of the field work, for different objectives, suggesting different data fields for recording locality data, different part/whole relations between data fields, leading to different, non-interoperable formats. Mining information from field journals thus did *not* bring about data interoperability, yet, it *did* further marginalize the concept of space embedded in the journal by rendering researchers even less compelled to invest time and effort in the original field journals.

At this point it may seem the researchers were left with the worst of possible worlds: a globally representative, standardized and mechanically objective (Daston & Galison, 2007) record is heavily used while inaccurate on multiple aspects as mentioned above; whilst a locally comprehensive and judgment-based accurate record is decreasingly accessible as researchers become accustomed to receiving their machine-based answers after 1 minute. Ironically, the harder the MVZ staff tried to apply Grinnell's vision, the faster it seemed in some respects to fade away.

Minding the Gap, Local Workarounds and Universal Interoperability

We have argued so far that examining the history of the MVZ's use of two concepts of space can explain, at least in part, how and why a lack of data and metadata interoperability emerged within the museum's informatics infrastructure. This is one reason why history and sociology can be useful for biologists: minute contingencies, historically entrenched in their routine work, brought about this conceptual gap, and it was the biologists themselves who uncovered this practical and conceptual "problem of locality" through careful study and reflection on their own historical records and documents. In this section, we discuss how their continued attention to institutional history, sociology and conceptual meaning is resolving the problem. We argue that resolution involves minding the gap for the purpose of "bridging" it rather than generally "closing" it, by a practice of "local workable alternation" rather than "universal interoperability".

An institutional response to the locality-interoperability challenge surfaced when the MVZ director, the PI for digitizing MVZ collections, the bioinformatics programmers and the georeferencing manager agreed that the way to "'connect'" the different locality records and make them less vague would *not* be to rewrite them all as various kinds of database records with GPS measurements. Instead of unifying all location descriptions, the MVZ resurvey team decided to return to Grinnell's alternating vision: "These field notes and photographs are filed so to be as readily accessible to the student in the museum as are the specimens themselves" (Grinnell, 1910: 34).

Since 2003, a large portion of the field notes and photographs have been digitized and posted online, yet posting did not make this information readily accessible in the sense one expects of queries to relational databases, because the posted notes were not linked with particular specimens. Since 2007, the GReF (Graphical Referencing Framework) project began to link every specimen in the collection with the journal field note page(s) on which it is described. Trained undergraduate students read the online field notes and whenever they come upon a specimen number, a date or a location, they tag it electronically. Later, a link is made to every place in the database where this number, date or locality is mentioned. The result is not interoperable in the strict sense, because one does not receive a machine-produced answer to one's query. However, a satisfying resolution is indeed achieved since one can click on a link from a single specimen page and reach a page in the journal narrating how it was collected. The researcher can thus quickly work back and forth – alternate – between the two kinds of information, posing structured queries in one and reading free-text descriptions that answer different questions in another.

Since both concepts of space are expressed through differing practices to organize location records – universal standardization and local flexibility – and since both these practices were necessary for re-using Grinnell's and the re-survey information, one can and must alternate between them while juxtaposing their different record types. GReF did not invent workable alternation – Grinnell alternated between tags and notebooks a century earlier – but it did exploit computer technology infrastructure to greatly speed it up and make it widely and freely accessible.

The story we have told here is *not* a part of a global transformation from the theory-driven goal of understanding species distributions to a data-driven goal of practically responding to climate change. Grinnell and his successors shared a vision of a universally useful information infrastructure that was based on their own, centralized contributory expertise (Collins & Evans, 2007) and in that sense, the successors have held true to Grinnell's legacy and initiated the resurvey as a fulfilment of that legacy. But the resurvey participants also brought new perspectives to bear, due in part to the transformations of ecological science, in part to changing technologies – especially the introduction of digital computers, relational databases, global positioning satellites and receivers (GPS), and GIS maps, and in part to changing political interests in climate change and pressures that placed a premium on rapid access to data on species distributions. These scientific, technological, and political changes led to tensions when new methods and protocols were brought to bear ostensibly on the Grinnellian project, which we explored here through the lens of 'location' meanings and records. These changes, however, should not be described as replacement of one set of practices by another, but rather as a more complex articulation of concepts and practices derived from the transformation of ecology, society, technology, and their intertwined infrastructures.

Case Study II: Hamaraag's Landscape Modulator

A Research Project Turning into a Monitoring Institution

In 2001, while the MVZ's senior staff began thinking about an NSF grant that would sustain the Grinnell resurvey, another group of prominent ecologists on the other side of the Atlantic began writing their own ISF (Israel's Science Foundation) proposal on species response to climate changes and to the presence of a landscape modulator (LM). An LM – typically a perennial primary producer – constructs a patch in the landscape that affects abiotic variables (e.g. soil moisture, temperature etc.) around its location and thus may filter the presence of other species from other locations (Shachak et al., 2008).

It began in 1999, with three ecologists, one from a research university and two from Israel's Nature and Park Authority, who agreed on a common theoretical interest: to test the LM model as a way of better explaining and managing biodiversity across different spatial scales. Thinking about the LM model required additional fields of expertise, which added four more researchers from three different academic institutions. They all knew and appreciated each other from years back, with discussions starting more than two years before the actual proposal submission. A first draft was completed and distributed within the group on September 2001, yet it mainly revealed the need for further clarification. Discussions continued and in November 2002 a formal proposal was submitted for the ISF's centre of excellence. Three paths were suggested for testing the LM model – mathematical modelling, experimental manipulation and analysis of observations along the Israeli gradient – yet only the middle section – won funding.[4] Hence, although the researchers originally planned for a national database to facilitate the information emerging from their nationwide experiment, the funding forced them to allocate their own limited private funding for the heavy task of building a group database, which meant that during the planning period and the first year after receiving the ISF grant, the data remained within private excel sheets rather than being shared. On November 2003, when the MVZ researchers set up their red truck and Sherman traps to leave for Yosemite Valley and repeat Grinnell's localities, the ISF grant number 1077/03 became operable and data production and organization began.[5]

Similar to the MVZ's conceptual and practical location-deliberation over how best to repeat their survey, whether to revisit a single trap, a transect line or a nearby habitat, these LMB (Landscape Modulator Biodiversity) researchers discussed whether to re-sample individual traps within a patch, individual patches, patch-types within a plot, a bounded box plot (1000 m^2) or an LTER (Long Term Ecological Research) station (20,000 m^2). By 2005 additional LTER stations joined the group, they re-named the project Hamaarag[6] (in Hebrew 'The Web'), and designed together an official logo with a symbol of Israel-LTER (Long Term Ecological Research). Toward 2007, when both ISF and NSF funding period were nearing their closure, the Grinnell Re-survey team utilized the MVZ's Alexandra foundation funds to maintained their original course, while Hamaarag seemed to be changing its course: some ecologists with a more theoretical stance began to miss P.I. meetings, a few conservation biologists and governmental officials joined, and other founding ecologists changed their titles to 'Board Directors' rather than 'Principal Investigators'.

As a general tendency, the project now became dependent on ministry and private funding.[7] Since January 2007, and especially after the ISF funding ended in October 2008, the field-protocols and data-models became focused on monitoring rather than experimentally testing a model, and the main mission turned out to be providing intensive, rich and reliable data (Karasti et al., 2010) for evidence-based national management rather than a basic theoretical synthesis. To symbolize this gradual change, in 2010 a director who was not a trained biologist was appointed, and in 2013 the new logo lost any mention of the LTER network, the official name slightly changed to Hamaarag, and almost all the data were collected outside the original LMB research stations.

The product of all this monitoring work is, first and foremost, an annual comprehensive report freely downloaded from Hamaarag's website and actively presented to relevant governmental officials. In addition, Hamaarag conducts many other activities to increase accessibility of its results, with an explicit aim to reach policy makers and enforcers at all levels rather than only scientists (Shavit, observation during Hamaarag's meeting on September 4th, 2012 and January 10th, 2013). Hence the dominant printing language is Hebrew for the Hamaarag and English for the LMB. After 12 years, it seems the scientific transformation is complete: from a theoretical and question-driven study to a practical and data-driven information infrastructure for national conservation.

We, the philosophers involved in this study, would agree about the result; we would also agree about the socio-political pressures mentioned above that contributed to this result. All this is relevant, yet we claim *not sufficient*, for telling the story of this project. Tracking the changing structures of the project's information infrastructure will tell us a deeper and more complicated story of memory practices (Bowker, 2005). That is, a knowledge infrastructure, in particular an online database targeting biodiversity in the face of climate change, is not a mere object but a dynamic and context-dependent network of commitments and choices, and the particular structure of organizing its 'location' information can reveal what these researchers are committed to remember and what they choose to forget (Bowker, 2005: chapter 3).

In particular, building an information infrastructure that assumes a single hierarchy of 'part'/'whole' among all LTER stations and a top-down standardization of all the different ways to describe a location, was not consistent with the LM's interactionist concept of space nor the international LTER tradition of diverse and flexible e-data structures (Karasti et al., 2006). This inner tension – to be elaborated below – can explain, at least partly, the project's continuous underuse of its databases and its shift toward a location-based monitoring program. The next section will illustrate our claim that following the structure of the project's memory practice may help to explain this chain of events, and it does not suffice to follow only the politics of funding or data-ownership.

Database Genealogy

This section will illustrate how deserting a top-down information infrastructure and instead enabling a localized bottom-up approach enabled new scientific questions, new working protocols and an opening for an involved citizen-science project within a national monitoring program. In order to support this claim about the relevance of information infrastructure for shaping scientific questions, models and practices, we will now briefly unfold its "infrastructure time" (Karasti et al., 2010).

As already mentioned, during the planning time-period of 2001-2002 the question of organizing the data for analysis of multiple users was raised and discussed, yet none of the P.I.'s had formal experience in information management nor sufficient funding for establishing a long-term, large scale online database from their own private research. In 2005 a bright young student began his third year of studying physics and computer science at the Hebrew University of Jerusalem. During that year, he decided to apply his programming skills, acquired through theoretical training in the university and practical experience in the High-Tech industry. After some time searching he found a place at the department of Evolution, Ecology and Systematics (EES), with a senior theoretical ecologist who was also one of the project's P.I.s (Principle Investigators). In August 2005, the young programmer was hired to design and operate the project's database, and his MSc thesis was supposed to interoperate data from this database to help build a theoretical synthesis that would be relevant for all the P.I.'s involved in the project. He devoted a semester to taking ecological courses and meeting field ecologists, and on January 2005, the programmer presented his initial design, received comments and the general approval of all the P.I.'s present, and began to work on making the database operable.

In May 26, 2005, all P.I.'s received a short email from the senior theoretical ecologist asking them to move onto the next step, that is, send their and their students' data in order for the student programmer to "develop the database *for* the synthesis of the ISF data" (P.I.'s email exchange, May 26, 2005. Our italics). In the long email exchange that followed, concern was raised over the conceptual scale of the student's synthesis and the timetable of its publication. Regarding the scale, some were not sure whether the data currently available could answer in a satisfactory manner such a broad question (P.I.'s email exchange, May 28th, 2005). Others expressed concern over the extent of the original data left unpublished for other members of the group – especially graduate students – if the synthesis were done before their student's manuscripts were sent (P.I.'s email exchange, May 29th, 2005). When the P.I.'s were asked to identify which data sets should be left outside the synthesis, what was left was neither general nor interesting enough for the young programmer and his advisor to work on. When the programmer was asked to distinguish the database from the synthesis, it became clear the former was designed for the latter, hence a strict separation was not practically feasible.

All the senior scientists involved knew each other for years, had mutual appreciation, cared about the project, successfully overcame previous rounds of passionate theoretical debates, repeatedly declared that their disagreements were *not* personal (Email exchange on May 30th, June 2th, and during multiple P.I. meetings) and were committed enough to drive long distances for face-to-face talks on June 17th and 21st, 2005. In short, a resolution seemed certain enough to joke about: "given [assuming?] that we are dealing with a reasonable group of scientists, I am guessing that some compromise is possible" (P.I.'s email exchange on May 30th, 2005).

Yet despite all efforts, a lockdown occurred, partly because the agreed original plan was for only one interoperable database for all LTER stations to test a single unifying theoretical LM model previously discussed for nearly a year. At each LTER station a different P.I. invested much time and effort in producing information to test the same LM model while organizing its location data and metadata differently. Reasons for metadata diversity are themselves diverse: a) the LM species differed among sites (e.g. Common Oak (*Quercus calliprinos*) at the Meron station and Negev Hamada (*Haloxylon articulatum*) at the Avdat station), b) hence constructed patches that looked different and denoted different within-patch-type hierarchies (for example a three layer 'woody'-'periphery'-'open' for Oak trees and a 'woody'-'open' dichotomy for Hamada bushes, Programmer's internal report, December 27th, 2006); c) the fixed plots rendered a lat/long description not necessary[8]; d) the abilities of the project's database manager[9] and finally e) the global LTER network is characterized by a mixed bottom-up knowledge infrastructures (Karasti et al., 2010) and highly diverse data-sets are highly common in eco-informatics (Michener & Jones, 2012).

Given an agreed need for a spatial hierarchy in the database between the sampling units – trap, patch, plot, site, country – but with no agreed mechanism on how to parcel it,[10] and given a single overall synthesis as an agreed common goal but with no agreed temporal mechanism on how to parcel its part/whole relations, a spatio-temporal gap between the P.I.s seemed inevitable. Should the researchers change how their spatial data were organized for the database to be able to automatically aggregate and compare their data? Will it still answer their questions? Should the graduate students donate their data to support the overall synthesis or should the synthesis study await the publication of their results?

Given the effort already invested by the P.I.s and their students in collecting and storing the information in a certain way, it was perhaps rational on their part not to begin investing in alternative suggestions in various possible worlds: two types of synthesis, two interoperable e-infrastructures, or waiting for all the ‘partial’ or ‘small’ questions – site or taxon specific – to be answered and published before the overall synthesis would be attended to.

Therefore, although all the P.I.’s wanted to reach an agreement and render the project successful, something had to give, given their goal of a common single model and yet the theoretical and methodological diversification it facilities; and given their goal of a common standard to interoperate their e-data and yet their diverse metadata at each site-location.

Most P.I.s were linked to a LTER station, and thus could continue their work without a uniform organization to their data. Eventually, not enough data were sent to that database and it was never completed. The programming student officially left the project on August 2005, and without an interoperable database to work on, his advisor also became less involved.

It took a while to find a replacement, during which data were curated on a site-by-site basis. On November 2007, a second database manager was hired. This time, no conflict of interests was expected since the database manager had no research interests invested in the database, and data and metadata interoperability was expected since he was a well-experienced informatics person. Indeed, no conflict occurred, yet the new attempt for a single standardized, all-encompassing database, even when detached from any theoretical synthesis aspirations, still did not hold much of the project’s data and most of the data it did hold was scientifically underused (Information Manager, interview on December 23, 2008).

Why? One contributing factor might be precisely this detachment. The previous, synthesis-oriented top-bottom database was complex and time-consuming for the biologists to fill, but could potentially test the model they cared about; hence, that effort could be justified. Given that a grand and potentially high profile synthesis was no longer expected to emerge from their data, and given that much of their recorded information could only be standardized via direct communication with their Information Manager and this meant additional work for all those involved, then perhaps it was rational – or at least economic – for many researchers not to invest in changing their data organization or in describing it in detail to their new Information Manager. By 2008, when the ISF grant ended, although several noteworthy publications were indeed produced and the vision of a long-term research was still intact, the completed database was nonetheless left with relatively few data entries (Senior Information Manager, interview on February 28th, 2011). Moreover, each year the database manager received less data from the different LTER sites to standardize and store in the database. In 2010 he also left the project.

For the next three years, there was no database manager and no central database. One may expect that without a unifying information infrastructure to query from, such a national, large-scale and long-term biodiversity project would have surely dissolved. Yet this did not happen. Instead, the project successfully re-invented itself as a national, again long-term, *monitoring* program.

A new, third, database manager was hired for Hamaarag on May 2013. This time, there was no deliberate attempt to use data from the LM research project or its LTER sites. The monitoring data were organized very differently: instead of a single unifying model or a single identical recording protocol to be conducted at all locations, a different, bottom-up scheme was initiated and later coordinated to fit the goal of national monitoring. Hamaarag established teams of experts – theoretical biologists, field naturalists and sometimes policy makers – who specialized in a certain region, habitat or taxon to form a think-tank, defined their specific habitat type, its threats and biodiversity indicators and tailored the monitoring protocol for their habitat and/or geographical region (monitoring director, presentation on March 21st 2013). Hamaarag’s scientific committee refined that protocol. Some parts were standardized to fit the protocols of other regions or habitats – e.g. randomly choosing 3 settlements between the variety available as a replicated location of threat – and the resulting location data became standardized and accessible by recording the GPS coordinates of each transect line and exporting that information as a series of KML records to Google Earth.[11]

Some types of interactionist location information – e.g. patch type – were gone, while others remained – e.g. a transect-line located according to an organism's interaction – habitat choice behaviour – with a nearby settlement. Not all ecologists adopt this interactionist perspective,[12] but many do, and the visibility and impact of this project for strengthening an evidence-based national policy for biodiversity conservation is of little doubt. By October 1st, 2014, a fourth information manager arrived, this time immediately following his predecessor and continuing her line of work. Hamaarag, now an official consortium of all the relevant governmental ministries that also cooperated with all major conservation NGO's in the country, began its third year of national monitoring program, secured funds for the next five years, organized its third annual international symposia, published its second "the state of nature report", its first "the state of the sea report", and further deepened its regional approach.

Hamaarag does not seek nor pretend to supply generalizable data that represents a habitat on a national or international scale. Instead, it aims to provide accurate and comprehensive regional data across the country, measuring changes in species richness relative to specific threats and in some cases, offers regional conservation recommendations. As we have seen, this project holds a history of targeting their locations and using bottom up and diverse 'location' descriptions. The targeted location was assumed to be affected by its living inhabitants – human and non-human – and it was the local experts – often different people at different regions – who mostly decided how to characterize a locality and monitor its biodiversity. Given this perspective, it is perhaps less surprising that Hamaarag responded positively to a group of eight policy makers and scientists from a peripheral region who argued for a special monitoring protocol at the northern Hula Valley[13]. Hamaarag's scientists recognized the new questions to emerge from monitoring this small region (which holds 40% of the nation's stream water) and the value of a pro-active municipality. Yet some of the local northern researchers also asked for citizen science information to be considered as part of Hamaarag's knowledge infrastructure,[14] which committed Hamaarag to additional deliberation on the epistemic value of citizen science and of resisting epistemic injustice.

A citizen scientist is a volunteer who collects and/or processes data as part of a scientific inquiry (Bonney et al., 2014; Silvertown, 2009). The citizen science project organized by the local "River-Watch" and the regional "Town Square Academia" added social involvement to the volunteer scientific activity. This local information infrastructure was designed to facilitate a pro-active learning community that would acknowledge and preserve its local heritage. Its data sets were small and diverse yet some environmental protocols were pre-structured to fit the standards of Hamaarag, therefore, enabling a peripheral locality to donate its information to a national infrastructure, and thus receive national recognition of its local expertise and knowledge.

Support for incorporating such local knowledge enables Hamaarag to help resist epistemic injustice (Fricker, 2007), which is, in our case, the injustice inflicted by prejudging the testimony of a resident of a rural periphery, due to her locality, as not really "understanding" her own environment, and therefore not recognizing her as eligible to decide on its future. Between 2009 and 2011 a local NGO named "Nature and Landscape Charms" protested for the people's right for a clean stream running through their city, and in 2012 this local NGO became aligned with another pro-active initiative: "Town Square Academia", who also aimed to recognize the local residence's knowledge about their stream.

The objective of Town-Square Academia is to galvanize an involved and pro-active *regional learning community*. In practice, multiple free courses are conducted outside the campus walls, lead by volunteer experts – academic lecturers together with local people – aimed at conveying existing scientific knowledge as well as documenting and studying local tacit knowledge that is relevant to the community and the researchers. Some courses also develop a group project to continue the learning process and to reach an action-based knowledge within a community of practice (Wenger, 1999). One such course was "A Few Things We Might Not Know About Water" and its group project was the "River-Watch", which still monitors the Jordan River sources, a common resource that physically connects Jews and Arabs, religious and secular, underprivileged and established social groups. Tracking the condition of wildlife and water may empower conservation, build new social and political ties, and suggest an alternative, less-hierarchical and more-involved dialogue between the academia and its locality.

It may seem surprising that Hamaarag, with its nation-wide coverage and uniform standards, would even consider supporting citizen science information. Yet it was considered. An international symposium on citizen science was organized in Jerusalem on February 24th 2014, following a trip to successful European projects. Eventually it was decided that collecting and saving citizen science knowledge would be the responsibility of the SPNI (Society for the Protection of Nature in Israel) rather than Hamaarag. But even after this decision, the head of the monitoring plan at Hamaarag still reserved small funds for the Hula Valley scientists and for the "River Watch" citizen science project. Eventually, a local shortage of determination and funds kept the citizen science knowledge outside the national infrastructure, but why was this option even considered – and with a clear positive spirit – by Hamaarag?

Obviously, we cannot give a definitive answer, yet one possibility is that one's infrastructure can also help entrench a certain theoretical path to be upheld downstream, whether one explicitly agrees with the contingent results of that path or not. In our case, given Hamaarag's rooted use of an interactionist 'location' and its tradition of reliance on local experts holding local information infrastructures; and given the locally oriented request to monitor the Hula Valley alone, with a proactive, human-environmental interactionist concept of space; it became much easier for Hamaarag's governing committee to make place for such an involved location information in their knowledge infrastructure. A similar move would have been much more difficult within a top down, universally and uniformly standardized database. For example, the MVZ's knowledge infrastructure was envisioned and structured, since its establishment in 1907, to be "a centre of authority" in the west coast, i.e. a research institution that spread its own standards of collection and recording rather than absorbing local standards (Shavit & Griesemer, 2011). MVZ researchers systematically relied on local information and opinion, yet a manifestation of this knowledge was never part of the MVZ collection goals and practice, hence it would have been much more difficult to incorporate into its 21th century online database (Shavit & Griesemer, 2011). Given the MVZ's top down approach and Hamaarag's bottom up approach, perhaps it is not so surprising that adding local, pro-active knowledge to the database was positively considered in the latter rather than the former.

Conclusion

This was a story of the various attempts at building and operating a long-term and large-scale knowledge infrastructure, by two influential scientific projects in California and Israel. Both projects thought in advance about how to organize their location information, both have found that two different concepts of space – exogenous and interactionist – are necessary for producing location records that are accurate and generalizable, and both were somewhat surprised to discover inner tensions if both ideals are employed for organizing the same data at the same level.

At first, both invested most of their effort and resources in making their data and metadata even more standardized and globally representative, yet later recognized the inherent tension of recording 'location' only exogenously. Both projects found workarounds to resolve the problem by frequently alternating local and flexible records with global standards – instead of choosing one or the other – and opened new possibilities for scientific research as well as for explication of local memories found – yet not recognized – in the community. Based on these stories we argued that it is justifiable to invest more in the infrastructure to sustain local memories of a locality, and in alternating between the local and global memory practices – both rationally and (sometimes) morally. This argument goes against the mainstream practice of many biologists and database managers, who keep investing ever more increasing funds into streaming and standardizing local data into global databases, as well as keep mentioning its inherent and widespread problems of accuracy (Vanderbilt and Blankman, in press). Following these long term information infrastructures revealed why and how they not only facilitated the preservation of collected data, but also theoretically problematized the foundation of these data and perhaps also directed the future course of its collection and analysis.

Acknowledgments

We deeply thank the MVZ's and Hamaarag's researchers and technicians for their rigorous work and for making this article more accurate. At the MVZ, we especially thank Steve Beissinger, Carla Cisero, Chris Conroy, Karan Klitz, Michelle Koo, Craig Moritz Juan Luis Parra, James Patton and John Perrine. At Hamaarag, we especially thank Nirit Asaf, Naama Berg, David Blankman, Yuchay Carmel, Eli Groner, Avi Pervolotzki Yael Lubin, Ronen Kadmon, Gidi Naaman and Moshe Shachak. Finally, we wish to deeply thank the pro-active people who work with the community and try to transfer that knowledge to the academia: Iris Eshed and Anat Kolumbus. This work could not have been done without the support of the ISF (Israel Science Foundation) grant 960/12.

References

Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ & Parrish JK (2014) Citizen Science: Next Steps for Citizen Science. *Science* 343(6178): 1436–1437.

Bowker GC (2005) *Memory Practices in the Sciences.* Cambridge MA: MIT Press.

Bowker GC, Baker KS, Millerand F & Ribes D (2010) Toward information infrastructure studies: Ways of knowing in a networked environment. In: Hunsinger J, Klastrup L & Allen MM (eds) *International Handbook of Internet Research*. London, UK: Springer, 97–117.

Collins H & Evans R (2007) *Rethinking Expertise*. Chicago IL: The University of Chicago Press.

Dagiral É & Peerbaye A (2016) Making Knowledge in Boundary Infrastructures: Inside and Beyond a Database for Rare Diseases. *Science & Technology Studies* 29(2): 44-61.

Daston L & Galison P (2007) *Objectivity*. New York: Zone Books.

Ellison AM, Osterweil LJ, Hadley JL et al. (2006) Analytic Webs Support the Synthesis of Ecological Data Sets. *Ecology* 87(6): 1345–1358.Ellison AM, Barker-Plotkin AA, Foster DR & Orwig DA (2010) Experimentally Testing the Role of Foundation Species in Forests: The Harvard Forest Hemlock Removal Experiment. *Methods in Ecology and Evolution* 1(2): 168-179.

Fricker M (2007) *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford, UK: Oxford Univesity Press.

Gannett L & Griesemer JR (2004) The ABO Blood Groups: Mapping the History and Geography of Genes in Homo Sapiens. In: Gaudillière J-P & Rheinberger H-J (eds) *Classical Genetic Research and its Legacy: The Mapping Cultures of Twentieth-Century Genetics*. Abingdon, UK: Taylor & Francis, 119-172.

Gerson EM (1998) *The American System of Research: Evolutionary Biology 1890-1950*. PhD Dissertation, University of Chicago, IL.

Gerson, EM (2007) Reach, Bracket, and the Limits of Rationalized Coordination: Some Challenges for CSCW. In: Ackerman W.A, Halverson M.S, Erickson C. A, Kellogg Th (eds) *Resources, Co-Evolution and Artifacts: Theory in CSCW*, Dordrecht & New York: Springer-Verlag, 193–220.

Griesemer JR (1990) Modeling in the Museum: On the Role of Remnant Models in the Work of Joseph Grinnell. *Biology and Philosophy* 5(1):3-36.

Grinnell J (1907) *Letter to Alexander*, p. 1, July 7, 1907. MVZ Archive, MVZ Main Gallery, Right Cabinet, File Name: Museum Correspondence - Grinnell.

Grinnell J (1910) The Methods and Uses of a Research Museum. The Popular Science Monthly 77: 163-169.

Grinnell J (1917) Field Tests Concerning Distributional Control. The American Naturalist 51: 115-128.

Grinnell, J. (1921) The Museum Conscience. Museum Work IV: 62-63.

Grinnell J (1929) *Schedule of Curatorial Duties for Staff Members*, p. 5, August 15, 1929. MVZ Archive, MVZ Main Gallery, Top Left Cabinet, File Name: Museum Methods - Historical.

Grinnell J (1935) *Analysis of Functions*, p. 2, November 22, 1935. MVZ Archive,MVZ Main Gallery, Top Left Cabinet, File Name: Museum Methods – Historical.

Grinnell J (1938) *Suggestions as to Collection and Field Note Taking*, April 20, 1938. MVZ Archive, MVZ Main Gallery, Top Left Cabinet, File Name: Museum Methods - Historical.

Guisan A & Thuiller W (2005) Predicting Species Distribution: Offering More than Simple Habitat Models. *Ecology Letters* 8(9): 993-1009.

Jalbert K (2016) Building Knowledge Infrastructures for Empowerment: A Study of Grassroots Water Monitoring Networks in the Marcellus Shale. *Science & Technology Studies* 29(2): 26-43.

Karasti H & Baker KS (2008) Digital Data Practices and the Long Term Ecological Research Program Growing Global. *International Journal of Digital Curation* 3(2): 42-58.

Karasti H, Baker KS & Bowker G (2002) Ecological Storytelling and Collaborative Scientific Activities. *ACM SIGGROUP Bulletin* 23(2): 29-30.

Karasti H, Baker KS & Halkola E (2006) Enriching the Notion of Data Curation in E-science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network. *Computer Supported Cooperative Work* 15(4): 321-358.

Karasti H, Baker KS & Millerand F (2010) Infrastructure Time: Long-term Matters in Collaborative Development. *Computer Supported Cooperative Work* 19(3-4): 377–415.

Karasti H, Millerand F, Christine MH et al. (2016) Knowledge infrastructures: Part I. *Science & Technology Studies* 29(1): 2-12.

Kohler RE (2002) *Landscapes and Labscapes: Exploring the Field-Lab Border in Biology*. Chicago, IL: Chicago University Press.

Kohler RE (2012) Practice and Place in Twentieth-Century Field Biology: A Comment. *Journal of the History of Biology* 45(4): 579-586.

Latour B (1999) *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.

Levins R & Lewontin R (1985) *The Dialectical Biologist*. Cambridge, MA: Harvard University Press.

Longino HE (1990) *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.

Longino HE (2004) How Values Can Be Good for Science. In: Machamer P & Wolters G (eds) *Science, Values and Objectivity*. Pittsburgh, PA: University of Pittsburgh Press, 127-142.

Michener WK & Jones MB (2012) Ecoinformatics: Supporting Ecology as a Data-Intensive Science. *Trends in Ecology and Evolution* 27(2): 85-93.

Odling-Smee JF, Laland KN & Feldman MW (2003) *Niche Construction: The Neglected Process in Evolution*. Princeton, NJ: Princeton University Press.

Oyama S (2000) *The Ontogeny of Information: Developmental Systems and Evolution (2nd ed)*. Durham, NC: Duke University Press.

Shachak M, Boeken B, Groner E et al. (2008) Woody Species as Landscape Modulators and their Effect on Biodiversity Patterns. *Bioscience* 58(3): 209-221.

Shavit A & Ellison AM (in press) *Stepping in the Same River Twice: Replication in Biological Research*. New Haven, CT: Yale University Press.

Shavit A & Griesemer JR (2009) There and Back, or, the Problem of Locality in Biodiversity Research. *Philosophy of Science* 76(3): 273-294.

Shavit A & Griesemer JR (2011) Transforming Objects into Data: How Minute Technicalities of Recording Species Location Entrench a Basic Theoretical Challenge for biodiversity. In: Carrier M & Nordmann A (eds) *Science in the Context of Application*. Bielefeld, Germany: Zentrum für interdisziplinäre Forschung, 169-193.

Silver Y & Shavit A (in press) Learning from Success and the Success of Learning: "Town Square Academia." In: Orr Z, Golan D & Rosenfeld Y (eds) *Campus-Community Partnerships in Israel*. Jerusalem: Magness Press [in Hebrew].

Silvertown J (2009) *A New Dawn for Citizen Science*. Trends in Ecology & Evolution 24 (9): 467-471.

Star SL & Griesemer JR (1989) Institutional Ecology, 'Translations,' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907 - 1939. *Social Studies of Science* 19(3): 387-420.

Stein BR (2001) *On Her Own Terms: Annie Montague Alexander and the Rise of Science in the American West*. Berkeley, CA: University of California Press.

Vanderbilt K & Blankman D (in press) Reliable Metadata and the Creation of Trustworthy, Reproducible, and Re-usable Data Sets. In: Shavit A & Ellison AM (eds) *Stepping in the Same River Twice: Replication in Biological Research*. New Haven, CT: Yale University Press.

Wenger E (1999) *Communities of Practice: Learning, meaning, and Identity*. Cambridge, MA: Cambridge University Press.

Wieczorek J, Guo Q & Hijmans R (2004) The Point-Radius Method for Georeferencing Locality Descriptions and Calculating Associated Uncertainty. *International Journal of Geographical Information Science* 18(8): 745-767.

Wythe M (1925) *Suggestion for Handling Specimens Brought Fresh into Museum and Intended for Collections of the Museum of Vertebrate Zoology*, November 13, 1925. MVZ Archive, MVZ Main Gallery, Top Left Cabinet, File Name: Museum Methods – Historical.

Notes

1 'Location' and 'locality' are used interchangeably by the speakers and since this study spans decades and cultures we use an 'actor speech' approach and follow our speakers.

2 Biological models of social-environmental interaction include, for example, niche construction (Odling-Smee et al., 2003), foundational-species (Ellison et al., 2010) and landscape modulator (LM) species (Shachak et al., 2008), while social implications of this approach are explored by Levins and Lewontin's (1985) dialectical perspective and developmental systems theory (Oyama, 1985 [2000]).

3 Additional details on the MVZ's archive: http://mvz.berkeley.edu/History.html

4 We deeply thank the senior researchers for sharing their memories (Interviews on October 18th and 21nd, 2016) and for sending us their email exchanges.

5 The plots were not randomly and independently chosen, as space availability for a plot with the relevant patch types was very limited.

6 'LMB', 'MARAG','HaMARAG' and 'Hamaarag' are all names for more or less the same institution throughout its evolution. It will be referred to henceforth as 'Hamaarag' for the sake of simplicity.

7 From the start, the head biologist of Israel's NPA (Nature and Park Authority), was part of the team. During 2007 the head of the Israeli Academy of Sciences, brought his organization to take unofficial patronage of the project, and one can clearly see the shift towards organizations with a national focus: official letters of support arrived from the minister of Environmental Protection Office, the head for-

ester of KKL-JNF (Keren Kaymet LeIsrael – The Jewish National Fund) and the Heritage Program at the Prime Minister's Office. Given these assurances, the Yad Hanadiv private foundation and later the Ashkol Program, of the Ministry of Science, announced their support.

8 The location of the LTER plots was fixed by fences for a long-term duration, and given the budget constraints and the focus on the plot as the place of repeatable surveys, in all stations except one free aerial photos were used for marking the plot location rather than coordinates from GPS machines.

9 "The information manager usage of the data [is] according to his technical knowledge, for example defining a polygon is more demanding than [defining] a bounding box.**"** Interview with I-LTER database manager. February 28th, 2011.

10 In the programmer's final written report he stresses this point: "The database will *have to* take the patch type hierarchy into account…[but] note that the exact mechanism by which the database will do this is not yet defined!" December 27th, 2006. Our italics.

11 We thank David Blankman for this clarification

12 A dialogue conducted during a 'location workshop' illustrates this point: A statistician: "if we want to do a statistical estimate then in the end you should know the chance to having chosen that location. So the most objective way of doing that is listing every one by one kilometre or five by five kilometre grid cell in the region and then just using a random number generation from Excel or something like that to pick a specific one [location] and say that's your site". The response of Israel's NPA head biologist: "I want to stop you here now, because what you're suggesting is O.K. for a whole [eco-geographical] unit. But there are settlements within this unit, and we see them as the main focus of threat so we need to choose according to them. We cannot do randomly by grid!." October 22nd, 2012.October 22nd, 2012

13 The first meeting was internal to the Upper Galilee people, May 26th 2011, a proposal was submitted to Hamaarag on July 25th 2012 and on July 29th 2014 the cooperation became final.

14 On the very first meeting in May 26th 2011 two of the participants suggested children, students and lecturers as volunteers, and on August 10st 2014 Hamaarag's director sent an email agreeing to embark on a citizen science pilot.