

Note: work in progress - please do not quote without permission

Sustainable Taxonomic Infrastructures: System or Process?

Henk Koerten¹ and Peter van den Besselaar

VU University Amsterdam

¹Corresponding author: h.koerten@vu.nl

Abstract

Infrastructures are usually being envisioned as stable and static systems, where change only brings them from one state into the other. They are hardly defined as an ever developing process. Based on a an analytical framework developed from notions provided by Bowker (2010) we have collected and analyzed ethnographic data on the development of a web-based data management environment called Scratchpads. Our findings suggest that we need to revise our conceptualization of Infrastructure from static and systematic towards a process based approach.

1. Introduction

Discussing infrastructure development we tend to think of physical infrastructures like roads, canals, tunnels, sewers and drinking water and how these systems emerged, evolved, sustained, and faded away (Latour 1996; Van Marrewijk and Veenswijk 2006; Edwards, Jackson et al. 2007). However, physical infrastructures spawned other types of less down-to-earth, more virtual infrastructures, heavily relying upon their tangible predecessors: electronically enabled, information-driven and treated as independent from any physical limitations (Kainz 2000; Leadbeater 2010; Brown, Chui et al. 2011; Koerten 2011). The emergence and development of Internet, AKA *The Web*, has been described as a phenomenon of revolutionary proportions, creating new infrastructures, aimed at the creation and dissemination of information (Berners-Lee and Hendler 2001; Mosco 2004; Noveck 2009; Lanier 2010; Leadbeater 2010).

Just like any other infrastructure, information infrastructures are constituted by technology (Barley 1986; Kainz 2000)}. Unlike physical infrastructures, information infrastructures are usually envisioned as dynamic, vivid, and ever-changing (Koerten and Veenswijk 2011). For instance it is assumed they tend to adopt every technological innovation that appears on the horizon. This aspect of innovation seems to be at odds with our classical concept of infrastructure, which is framed as a reliable, institutionalized, ubiquitous, hard to change phenomenon which you only start to miss in case of failure (Star 1999). We even could argue that information infrastructures are to be treated as such when they become just as unalterable and stable as their physical counterparts. So the question is: do information infrastructures need to be stable, predictable and unalterable in order to be reliable and trustworthy?

Social media are a vivid example of electronic facilities acting as information infrastructures. Enterprises like Facebook, Google, LinkedIn, Skype and Twitter are ever-developing, continuously adding new services and functions, while none of them has a history longer going back any further than a

decade. Some of their competitors and predecessors are already gone or have faded away (Leadbeater 2010).

While social media have made themselves inevitable in a ICT propelled boost, we also have to ask ourselves, how do these techniques affect traditional information infrastructures? Do they change through technological innovation? Is it just a matter of doing things faster through ICT or are there new products and services created? Does ICT change the way things are done in communities where these infrastructures are used? Are ever-changing, ever-developing electronic services still conceptualized as sustainable information infrastructures?

For that matter, we are going to investigate the world of taxonomists as part of the scientific world of biologists and natural history scholars studying evolution and biodiversity. In particular we are focusing at a new electronic service provided through the internet for taxonomists, Scratchpads, has been received, picked up and used by scientists. First, we are going to review literature on infrastructures and information infrastructures in particular. Next we will spend a few words on how growing attention for biodiversity and evolution has affected the biological and taxonomic field. After that, we are going to present ethnographic data consisting of interviews, and the analysis of documents and websites related to the genesis, uptake and use of Scratchpads. We conclude with an analysis, followed by a conclusion.

2. Emerging knowledge of information infrastructures

Information infrastructures are by no means just technological artifacts, they are social constructions filling specific information needs, being the product of evolutionary processes (Bowker, Baker et al. 2010). Bowker devises an array of concepts to the existence and sustainability of information infrastructures and he specifically connects them with time-aspects being conceptualized as heterochrony. He argues these concepts determine the performance of an infrastructure (Bowker, Baker et al. 2010). Funding, technology, skills, careers and datasets; all together forming information infrastructures should be harmonized with special attention for time matters to make them sustainable.

As social media have become ubiquitous, they tend to influence how people communicate, also affecting professional communication. Also business relations are maintained using these communication facilities. Especially when professionals are collaborating in international settings they are inclined to use electronic facilities to communicate (Schejter 2009). Sometimes even hybrid forms of electronic infrastructures are mentioned, combining electronic infrastructures with dedicated services with the intention to serve a professional community (Berendsohn, Güntsch et al. 2011).

This brings us to the question how electronic information handling facilities and work practices relate to one another in a professional, information-driven environment. We may expect that available services affect existing professional work practices making them easier or even creating new products or services. On the other hand, specific professional practices regarding information management may influence the way information infrastructures are conceptualized, eventually determining how dedicated or non-dedicated facilities are conceived, used and altered. We want to focus on how electronic information infrastructures and professional work practices affect one another?

2.1 Biodiversity as a hybrid field of research

In Egypt, China and Southern-America, evidence has been found indicating that already in ancient times humans started to collect and document life on Earth. Printed and written accounts with descriptions of different kinds of plants started to appear in the late fifteenth century. Especially in Europe efforts were made to discover life in other parts of the world, among others represented in the work of Charles Banks, Charles Darwin and Alexander von Humboldt. Colonial powers created circumstances where knowledge grew and collections of species and data were established. It gave this type of research a distinct Western flavor, with data collections, species collections and research communities aimed at studying global life but being concentrated in Europe and Northern-America (Hawksworth 1995).

In the 20th century, awareness grew that human life was threatening the diversity of earth's species. Specific programs, were launched, such as the United Nations Environment Program (UNEP) to assess the nature and impact of the problem (Watson, Heywood et al. 1995). At that time the notion grew that the pace of extinction of species was accelerating and that it was time to act. This growing awareness made funding agencies, research groups and the general public ready to focus on information collection as a prerequisite to assess the magnitude of the problem. It also stimulated involvement of biological subfields such as ecology and taxonomy, which together with the upcoming DNA techniques saved the latter from disappearance (Godfray 2002).

When we speak of biodiversity we generally mean 'the five generally recognized kingdoms of animals, bacteria, fungi, plants and protoctists' (Watson, Heywood et al. 1995)(p.19). This is a very broad classification which can be fine tuned using concepts like species, genera, families, orders classes and phyla. While effects of pollution and climate change are affecting our present environment, having put the very concept of biodiversity on the map, the subject is by no means limited to our time. Charles Darwin was the first to discover evolutionary patterns in life on earth (Cox and Moore 2005). As a consequence, if we want to interpret contemporary biodiversity patterns we have to understand the origin of species.

There is some consensus about what fields of knowledge should be involved in the assessment of biodiversity. Biodiversity as a professional field is often subdivided in ecological diversity (spatial species variety), organismal diversity (between species variety) and genetic diversity (within species variety) (Watson, Heywood et al. 1995; Gaston and Spicer 2004). The first strand is obviously the domain of ecologists with a focus on biomes, ecosystems and habitat, while the latter two are connected to the taxonomic profession, declaring within- and between variety of species using classification routines.

Even a generalization of this kind with a simple delineation between some distinctive aspects of biodiversity contains problematic relationships, since a panoply of concepts of species is involved, all having their aspects of data collection and implicit and explicit values attributed (Gaston and Spicer 2004).

Since biodiversity knows no boundaries, its information and data sources are potentially beneficial to anyone studying life around the world, creating an atmosphere of openness and value-free data processing (Busby 2002). At the same time, taxonomists, ecologists, bio-geographers and the like remain within their own professional realm to team up with colleagues from their own background to form consortia (Bowker 2000; Smith and Penev 2011).

Professions like taxonomists and ecologists regard standardization to be crucial when it comes to information exchange. In that respect data and information are distinguished, where data possesses connotations of being raw and crude, while information is regarded as data that has been processed by a scientist, being registered and certified through publication in scientific journals . Depending on the taxonomic subfield publications can have different connotations.

Taxonomists usually publish a description of a species, containing all relevant information on a variant of a specific species (Smith and Penev 2011). By considering species as the reference point in his work, a taxonomist runs the risk of only wanting to chase rare species, thus putting the question of spatial distribution at the background (Bowker 2000).

In turn, ecologists are trying to assess the variety of species in a certain area instead of taking the species itself as the object of research, which gives the data and information to be used in ecology inevitably a spatial component (Ingersoll, Seastedt et al. 1997; Busby 2002; Hobbie, Carpenter et al. 2003; Cox and Moore 2005).

As taxonomists are occupied with classifications of species, their first concern is how to bring order into their data and datasets (Raven, Berlin et al. 1971; Berendsohn, Güntsch et al. 2011). In order to accomplish that, taxonomists have been involved in numerous classification attempts, bringing the very essence of taxonomy to the forefront: classification (Bowker 2000).

One striking aspect of biodiversity is that immense quantities of data and information are involved in mapping biodiversity (Olivieri, Harrison et al. 1995). Both ecologists and taxonomists have to manipulate and handle large sets of data in order to draw conclusions and to turn these conclusions into publications.

We conclude this section with the observation that mapping biodiversity requires spatially and species-related classifying elements, which are both regarded as essential. They both have distinctive features, research methods and end products. At the same time ecology and taxonomy are partially overlapping, mutually related and professionally linked. For this research endeavor we are going to focus on taxonomy as the science describing species.

2.2 Scratchpads as an online tool for taxonomists: Case profile

This section broadly describes how an electronic taxonomic data- and information exchange platform has been envisioned, created and valued. Using EU grants it was collaboratively developed by a consortium of European Natural History Museums. With the slogan 'Create your own website to connect, work with and share your data', taxonomy-oriented scientists were encouraged to 'build a virtual community based on a your specific or wider research field'ⁱ. Where similar endeavors presented themselves as taxonomic databases or libraries, Scratchpads was deliberately profiled as establishing a community to enable collaboration.

Initiation of Scratchpads

Scratchpads was mentioned firstly in the EDIT project (2006-2011) which was described as a 'platform for taxonomy', 'a collection of tools and services which together cover all aspects of the taxonomic workflow'ⁱⁱ. The main goal of EDIT was to develop the Common Data Model (CDM) , to become 'a repository for every conceivable type of data produced by taxonomists in the course of their work, and

the backend for most EDIT components'. A common development platform was desirable, to 'foster a community-wide culture of collaboration' by offering 'tools to institutions without the latest-and-greatest technical know-how at their disposal to develop their own web applications for the CDM (or indeed for any web applications they would like to make available to the EDIT community)'. It was decided to select an open-source content management system called Drupal , allowing 'the creation of flexible websites adapted in look and functionality to the needs of each community'. Scratchpads was endorsed as 'a website where taxonomists can communicate and jointly work on their data'ⁱⁱⁱ.

In an dateless communication on the EDIT website, the ViBRANT project (2010-2013) has been mentioned as following-up on EDIT. Scratchpads was described here as being a success, having produced over 260,000 web pages, while it was also mentioned that Scratchpads as such was still experimental and had to be developed into a dynamic site for collaboration. While still under construction, its importance was underlined with the phrase 'London's Natural history Museum has underwritten the continuation of the Scratchpad service for the next five years'^{iv}.

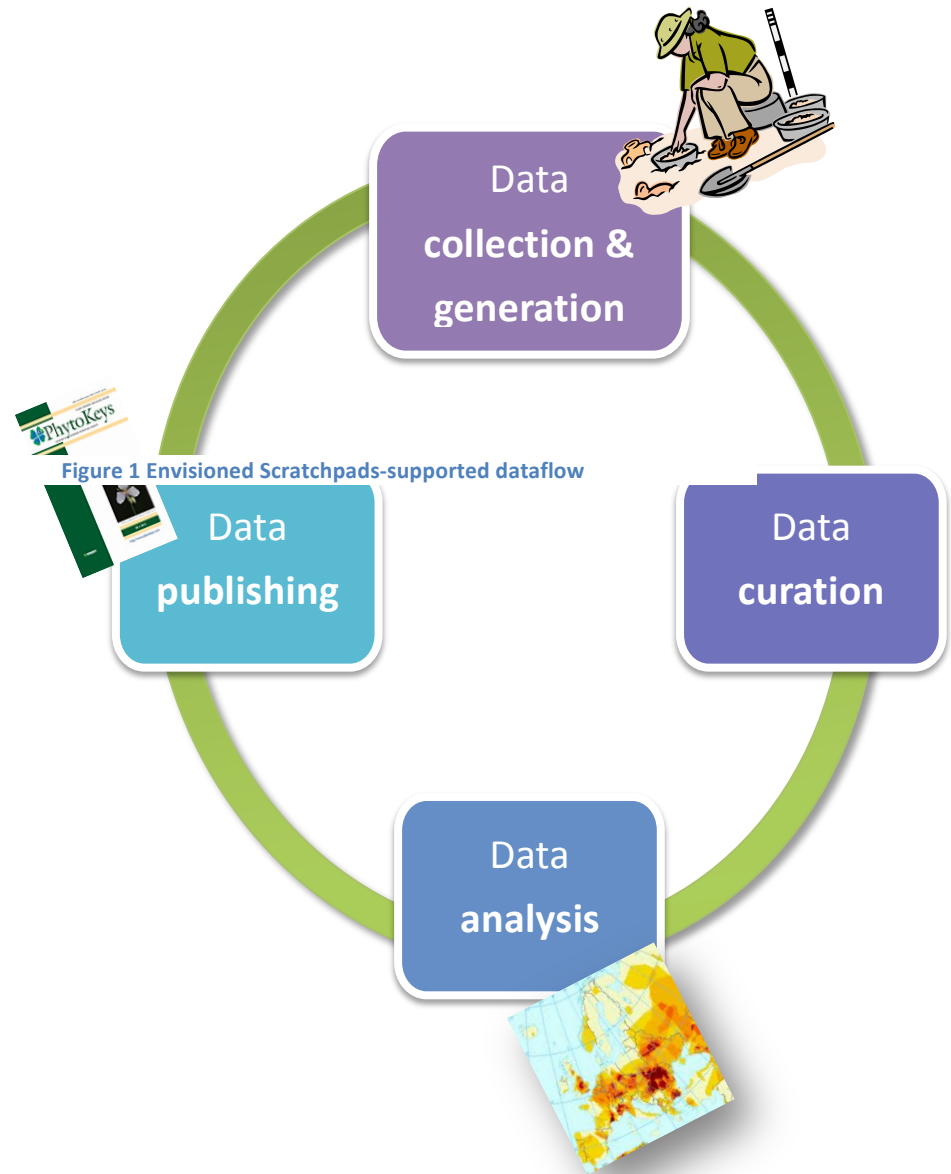
Uptake of Scratchpads

Becoming embedded in the ViBRANT project, Scratchpads became intensively endorsed and promoted. Where EDIT apparently described itself as a project focusing on developing standards for taxonomy, the ViBRANT project presented Scratchpads as a vehicle to develop a taxonomic research community. With the slogan 'Supporting biodiversity research communities', it developed services such as data publishing, data mining, literature mark-up and user training. In the 'Ambassador Introductory Presentation' which was also used at training sessions, the common procedure of taxonomic data-flow was envisioned as follows:

The four nodes of data workflow:

1. We collect and **generate** data
2. We **curate**, link and structure data
3. We **analyse** data
4. We **publish** data^v

seamless workflow



In policy documents, Scratchpads was seen as an efficient tool to connect these nodes . Bottlenecks were identified in the data flow process regarding structuring and formatting of data to be discoverable and usable to the rest of the research community and also the data publishing process as an inhibitor to make raw, unpublished and yet meaningful data available to the research community. These two bottlenecks ought to be eliminated to ensure seamless data workflow allowing better use of all available datasets allowing researchers to take credit for all their research volume (see figure 1).

Therefore, data, information & knowledge should be Digital, openly accessible, linked-up (to evolutionary data), to be analyzed and properly documented to conduct comparative analyses, studies of evolutionary process and biodiversity analyses.

Scratchpads today

Right now (June 2013) Scratchpads has become a flourishing online environment for taxonomists, allowing them to publish taxonomies together with related publications, data and pictures. The homepage of Scratchpads tells us that there are 506 Scratchpads, used by 6,468 active users, covering 72,621 taxa in 507,719 pages. There are Scratchpads ambassadors, located in 16 countries on all continents except Antarctica. Scratchpads training sessions have been organized across Europe and new sessions are coming up in Africa and Asia. Even though funds are still missing to secure Scratchpads after the expiration of the Vibrant project (ultimo 2013), it seems to be a lasting online environment which is likely to become an infrastructure for taxonomic research. The question then is, does it have, or is it likely to get the qualities needed for being an infrastructure? This question will be answered in the next section.

2.3 Emergent theories on Infrastructures

Quite some research has been published on all kinds of infrastructures. Subjects as diverse as Dykes, (Dicke 2001), demographic statistics (Harris, Sleight et al. 2005), spatial data (Koerten and Veenswijk 2010), communications and electricity (Edwards, Jackson et al. 2007) and postcodes (Raper, Rhind et al. 1992) have been considered as infrastructures for the benefit of human life and been scrutinized accordingly. They have been studied focusing on history (Edwards, Jackson et al. 2007), development (Koerten and Veenswijk 2010), societal impact (Raper, Rhind et al. 1992; Harris, Sleight et al. 2005) and change (Dicke 2001). They also can be divided into physical infrastructures (Dicke 2001; Edwards, Jackson et al. 2007) and electronic infrastructures (Raper, Rhind et al. 1992; Harris, Sleight et al. 2005; Koerten and Veenswijk 2010). These studies all try to grasp what an infrastructure is about, by focusing on how they originate, develop, change and even maybe fade away. The above mentioned studies lack a systematic approach. If we want the existing dispersed knowledge of infrastructures to be beneficial to emerging new infrastructures we have to bring some coherence into our thinking.

Geoff Bowker and Susan Leigh Star have tried to deepen our understanding of infrastructures by describing some common features, analyzing them as standardizing and classifying social phenomena (Bowker and Star 2000). It was also an attempt to recognize infrastructures at a more abstract level, where infrastructures help to process information through the mere fact that they function as sorting and classifying structures. Bowker and Star extend ideas initially developed by Star and Ruhleder on the conceptualization of infrastructure (Star and Ruhleder 1996) in order to make them more concrete.

In this line of reasoning, infrastructures, be it either physical or information oriented, they propose to study them using the following concepts:

- **Embeddedness**
- **Transparency**
- **Reach or scope**

- **Learned as part of membership**
- **Links with conventions of practice**
- **Embodiment of standards**
- **Built on installed base**
- **Becomes visible upon breakdown**
- **Is fixed in modular increments, not all at once or globally** (Star and Ruhleder 1996)

It follows from this list that infrastructures are ubiquitous, and yet hard to be noticed. Even if we do recognize infrastructures as such, we have trouble to distinguish aspects and parts, making it difficult to assess its boundaries and limitations. Generally we are not aware of infrastructures enabling or supporting daily life, since it invisibly supports tasks. Its existence can go back in history making it a phenomenon that has always and everywhere been around, going beyond a single event or a one-site practice. Since it is part of our daily life it is a target object to learn about. It inherits strengths and limitations from its presence, based upon functionality inhibiting modifications which can be a thrust to look for alternatives.

The attributes developed by Star and Ruhleder have guided Bowker, et al. to link them to time by mentioning temporal qualities of the discerned aspects of information infrastructures (Bowker, Baker et al. 2010). In their view, these aspects need to be aligned in time to allow information infrastructures to be successful:

[] one of the chief challenges is to move beyond a 1-6 year funding cycle of projects or a 30- year career cycle of the scientist to create baselines of data spanning multiple decades. (Bowker, Baker et al. 2010) (p. 107)

Based on notions about the US Long Term Ecological Research Program (LTER) (Hobbie, Carpenter et al. 2003), Bowker et al. suggest a successful infrastructure is only a matter of:

‘aligning what is naturally misaligned (funding cycles, scientist career trajectories, ecosystem cycles) is fundamentally an issue of distribution between technologies, communities, organizations institutions and participating individuals.(Bowker, Baker et al. 2010) (p. 107)

As Bowker et al. state in their analysis, in order create successful information infrastructures, technology, skills, careers of participants and datasets are concepts must be aligned. Reversely, if we want to assess or evaluate information infrastructures, the analysis of alignment between these concepts is a viable option. Below, we recapitulate what Bowker et al. have to say about these concepts.

Financial means are important not only for creating an infrastructure, they remain significant to its sustainability. Where project grants are a limited funding source in terms of time span, they may help to bring about certain tangible elements of an information infrastructure, like a database, a website or a metadata catalogue. These relatively short term funds are by no means an instrument to make an infrastructure really lasting. It is likely that when an information infrastructure seems to be meant to last, project funds are not able to secure that process.

Available skills may also be a determinant whether an information infrastructure comes into being and will sustain through time. Initially, tasks will concentrate on the design and construction of a

database, a website or a metadata catalogue, when it comes to actually using the infrastructure and to make it run through the veins of a community. Processes of development and enculturation have to be brought in place to embed it.

A sustainable infrastructure moves beyond the powers and means of involved individuals and their respective careers. Initial projects to launch specific actions may be successful through actions of individuals with unique capabilities and skills, in order to make these initiatives lasting they should be followed by actions to make the infrastructure a part of the work environment, letting it be a resource of which it is unthinkable that man can do without, that the user community has the idea that it has always have been there.

An information infrastructure discloses all data needed to do the job for which it is designed. It should be stored in formats to be able to comfort all actual and potential users. This does not imply that all data ideally should be stored in a universal standardized way. What is important here is to keep in mind that there may be different types of users, having different data needs, concerning nature, time and space aspects.

3. An infrastructural analysis of Scratchpads

Now we have identified some aspects and elements of what makes an infrastructure sustainable, we are going to use it as a framework for the analysis of Scratchpads, designed to support taxonomists to perform their work on describing biodiversity. In this section we will go through the project according to the themes identified in the previous section, namely funding, skills, careers and datasets. After a few notes on methodology we will present our findings , categorized using the abovementioned concepts.

3.1 Methodology

These preliminary findings are based on two focused group interviews, three single interviews and project-related documents. The focus groups consisted of 9 and 5 individuals varying in age, nationality, sex, and stage of career and were carried out in the spring of 2012. Participants were brought together to follow a training on the use of Scratchpads which could mean being a user or in the stage of considering to use it. The single interviews were with a senior researcher being an early adopter and two persons actively involved in the management of the EDIT and ViBRANT projects. The individual interviews took between 60 and 90 minutes using a topic list. The focus group interviews were held using statements to encourage discussion.

The findings presented below are based on a reduced set of data; some interview material still has to be analyzed and we are planning to do some additional interviewing, giving the research a better empirical base and to cover as much taxonomic subfields (zoology, botany, etc.).

3.2 Institutionalization and funding

Scratchpads was presented as an initiative to create an electronic environment for information exchange and storage in the EU funded EDIT program^{vi}. This program was meant to bring about ‘a collection of tools and services which together cover all aspects of the taxonomic workflow’, with a limited time horizon (2006-2011), due to the fact that only for this period of time financial means were offered by the funding agency. After the project ended, the corresponding website remains dormant.

Being the focal point during the course of the project, now it has become obsolete with just an occasional postings announcing events in related projects.

After Scratchpads was hailed by EDIT as a success story^{vii}, responsibilities were taken over in the ViBRANT project (2010-2013), which also was EU-funded. Meanwhile, with training sessions and inviting experienced users to become ambassador for the project, Scratchpads gained popularity and became known in the taxonomic community as a tool for administering and publishing data and taxonomies.

Being project driven, taxonomists have a keen eye for their vulnerable position as projects only usually have a time span of only a few years. They relate the temporal nature of their work to more institutionally related factors:

It is more an opportunistic thing, you can get funding for this if you present it in this particular way and if you think of a nice name then you can get three years of funding and then you more or less abandon it. I don't think there is a master plan for that, at all. (participant in focus-group interview, 2012)

You can spend your time, to make it nice, your website, and then after a few years, and then suddenly [fingersnapping]: 'we're sorry, no funding anymore'. (participant in focus-group interview, 2012).

Administering data within a 2-4 year project means that after the project is finished, storing and securing the data in a lasting way is not guaranteed. Some scientists describe Scratchpads as being a safe haven for collected data, including data collected by scientists from less developed countries:

I understand that it is one of the possible ways to study new unresearched territories in not so developed countries that have no money to study biodiversity. (participant in focus-group interview, 2012).

However there are also worries that maintenance of databases might become too costly:

But I must say that in general I think most of the money should still go into the research and not into developing the system even further. (participant in focus-group interview, 2012).

However, also scientists working in Western Europe see Scratchpads as an opportunity to solve inconveniences caused by poorly functioning organizational infrastructures due to insufficient financial means. A taxonomists describes the situation in her natural history institution:

Computer support in our institution is highly underdeveloped. A lot of promises, but zero results. They just don't have any budget for computer related development tasks, so we just keep on dreaming. Now we just got ourselves a GIS-lab, however with only one software engineer. That is highly insufficient as we have a collection to maintain of over 32 million objects. (Individual interview with participant working in a taxonomic institution, May 2013)

Yet, she sees on the level of joint institutions recognition of infrastructural needs that could be solved by Scratchpads:

The Consortium of European Taxonomic facilities (CETAF) has declared to guarantee budgets for upkeeping Scratchpads, so I think it is here to stay (Individual interview, May 2013)

A quite similar observation comes from a staff member of a taxonomic institution, working continuously on temporal labor contracts for various taxonomic research projects:

I have worked for this institution for about 20 years now, hopping from one temporary labor contract to another. And I am not the only one: only 30% of the total staff serve in permanent contracts. My job depends entirely on project funds, and in order to be eligible for EU funding, taxonomic institutions need to cooperate. Sometimes I even think the only reason for us to work together in projects like this is funding: without cooperation there is no funding, and consequently no work. My job is now on the line because the project expires. There are signs it will be continued with budgets coming from alternative sources, so let's hope for the best. (Project manager, personal communication, may 2013).

On the other hand, funding requirements may also determine that projects have a national signature:

Our government puts money into biodiversity research, and so we have our own taxonomic database and it is not really connected in the way that it is usually done. If you look at biodiversity informatics today, it is all about connecting and sharing and so on at the level of database systems, and that's what I think our field wants develop more and more. The way to go is the development of systems for terms and taxonomies, communicating to people of these separate databases, working towards one single database. It is not a technical problem that is hindering, it is more a political and resource thing. That people have control over this government saying: we spend our money on our national database instead of putting our data there where we don't have control over it. It is more an opportunistic thing, you can get funding for this if you present it in this particular way. (participant in focus-group interview, 2012).

Some have the feeling the taxonomic community has an obligation towards the general public and citizen scientists to explain what they are doing:

To spread the word that other people know, that things are open, and when the public is often funding us, so, they have the right to know too, and that other people are interested in this, yeah, just opening it. (participant in focus-group interview, 2012).

Explaining about the value of taxonomic research, a taxonomist speaks of species having economic value:

The ultimate goal is, they start out with species being of economic importance, like those being valuable for medical reasons, like mosquito's. I think the economic value they want to create in this particular genomic project is to describe DNA, to map genetic diversity, which does not end with only describing the variety of species. (participant in focus-group interview, 2012).

Taxonomic scientists demonstrated concern for the continuation of Scratchpads after funding ends late 2013.

What is going to happen when ViBRANT is done? Who's going to pay for Drupal version 8? O yes, there is cooperation in a technical sense, but there is also a Financial component.

You mean the upgrade of Scratchpads from Drupal version 7 to Drupal 8?

Yeah, because every three years or so, a new version of Drupal comes available, Who's going to pay for that? Now this was done through ViBRANT, but who pays next?

You know one thing for sure, they won't do it for free. So if they run out of money, the whole thing comes to an end.

They did some major investments already, so I expect that they don't just let it go.

But that's all EU project money.

All I know is the taxonomic institutions committed themselves, I know they put some effort in it to find the money.

And what about GBIF, can this be continued?

GBIF is supported by a lot of countries, signed in a multilateral treaty. (Conversation in a focus-group interview, 2012).

Our respondents definitely see an independent role for Scratchpads:

They didn't say they want to quit, and in the EDIT project it was already agreed upon in a letter of intent. I know they are actively searching for funds, because ViBRANT expires this at the end of 2013, they are looking for alternative financial resources. And they try to find out whether Scratchpads could become an independent facility, but that's too far ahead. (participant in focus-group interview, 2012).

Funding is a main issue in taxonomy which highly determines the choices people make. Scientists from taxonomic institutions, mostly part of a natural history museum, feel they are not properly supported with computers by their own institution and thus they seek other ways of getting their job done. Scientists working on temporal labor contracts on small (PhD) projects feel they have to secure results that come out of their research, otherwise their data would be gone. In countries where financial resources for biodiversity research are limited or where grant application terms require a national scope, independent services like Scratchpads offer a place to display data and an opportunity to these scientists to present themselves to the international research community without worries about budgets.

The fact that initiatives like Scratchpads are taken up by scientists from national taxonomic institutions jointly applying for EU grants gives them credibility in the work field. The team responsible for the development and revision of Scratchpads is lead by professionals having a taxonomic background affiliated with a leading taxonomic institution. This combination makes Scratchpads an initiative of taxonomists for taxonomists, being treated as highly reliable and trustworthy. It gives fellow-taxonomists the impression even though, financing is still insecure that a solution will be found to secure Scratchpads as a project.

3.3 Technology

Taxonomists are aware of the omnipotent role computers play in their work. Like in any other profession, computers are anywhere and everywhere. They are not only used in the office environment or in a laboratory to control analytical tools and devices, but also at other sites, like in museum collections and field trips:

What kind of computer device do you use?

Just my PC. (participant 1)

And on a fieldtrip?

I'll take a laptop, and usually my mobile phone to take pictures, adding coordinates, those kind of things. And of course at night, after a day of fieldwork, to analyze that day's work. In that sense a GPS device also counts as a computer. If I go to work in another museum I'll take my laptop with my database with me, together with a barcode scanner. (participant 2)

Moreover, even working in our own collection I'll take my laptop with me. (participant 3)

My database is in my computer, when I'm working on it, I'll do it in my office. When necessary, but only occasionally, I'll take my database on my laptop with me to go some place. (participant 1)

Let me put it this way, it is more efficient to take my database on a laptop with me than to do the database editing afterwards at my workplace. So, working in the collection means taking a laptop along. (Participant 3)

In our office it is often the case that you have the computer on one desk and on another you have your samples (participant 2) (conversation in focus-group interview, 2012).

Computers are seen as devices, the databases they develop and maintain are the true object of their work, that is where collected data is stored. The content is not limited to alphanumerical data, but may also contain images, photographs and documents.

Scratchpads users have the impression that technical matters regarding hardware and software are taken over by Scratchpads. It is taken up as a service rendered to the taxonomic community, offered by trusted representatives:

Scratchpads is definitely attractive to taxonomists, I remember when Vince demonstrated the application at a conference in Portugal some years ago. It is web-based, so it can be used anywhere you like, and it takes away worries about rights of data and software, you just start and that's it. In a way it is a democratic thing. By using it, I have the feeling I'm in control, you can upload any data you like. My Scratchpad contains already taxonomies of about 800 species. It solves the problem every taxonomist is facing: dispersed information. Here you bring multiple sources together, and you are able to open up your collection as well to build in restrictions, because you want to be the first to publish on that data. (excerpt from single interview, 2012).

A phenomenon that has been hailed as boosting innovation is open source software (OSS) (Gallivan 2001). OSS usually creates a platform for debugging, problem-fixing and patching (Gallivan 2001; Von Hippel and Von Krogh 2003). In the case of Scratchpads, the service has been developed with Drupal, being presented as:

. . . an open source content management platform powering millions of websites and applications. It's built, used, and supported by an active and diverse community of people around the world. (www.drupal.org)

They consider it as a service for administering data, but also as a data publication tool, moving data beyond the scope of individual taxonomic research projects. Instead of boosting innovation, together with a qualified team, Drupal ensures the reliability and robustness of Scratchpads.

Software has the habit of breeding new versions, containing updates and new features. On the one hand, software users are enthusiastic about bug-fixing updates and usually they cannot wait to use those new features. On the other hand, new software versions also have a name for causing new problems, especially when data has to be converted to a new format.

What features of Scratchpads do you want to be improved?

The actual information on how to get it done. I mean, sometimes you really have to search for it. Like how to upload data, I have been looking for it, but didn't find it. They have examples, but I just can't figure it out. (participant 1)

Just send them in London an email and you will get a swift reply. (Participant 2)

And the mere fact that if you click on help, you end up in Drupal help, then it's not about Scratchpads anymore, it's about Drupal. And all those, let me call them instructions, they are hopelessly out of date. (Participant 1)

The help menu in the old version of Scratchpads is hopeless, so I wrote a help menu myself for my Scratchpads, to help novel users to get accustomed. But the help function in the new version has improved and is quite easy to work with. But still, that improved help is so much different from what I know. (Participant 1) (conversation in focus-group interview, 2012).

Instead of thinking of computers as devices holding the data, now taxonomic scientists think in terms of databases holding all kinds of data, to be stored on either their own computer or in an internet based service. Scratchpads is seen as a computer at large, to be accessed either from an office or someplace else. It is a dedicated database, designed and operated by professionals from their own field and background, knowing the needs of the true taxonomist.

3.4 Profession and skills

Our data reveals relevant fragments of the process of data treatment from fieldwork to publication. Taxonomists generate, curate, analyze and publish data. This process is envisioned in figure 1, which is, and has been used in training sessions for Scratchpads. Our findings suggest this notion of the taxonomic profession is widespread. Without mentioning it as such, taxonomists see themselves as data

processors. However, this general, straight forward and broad observation calls for nuances, which we will describe in this section.

Taxonomists make field trips for data collection. They have to leave their office-space to go out into nature to collect, depending on their research topic, specimens and/or data on specimens. Despite the fact that field trips are mostly undertaken by a group of taxonomists, the data collection process itself is usually a personal endeavor. Having this said, it certainly does not imply that research is carried out only by individuals in isolation, but is a collaboration between colleagues that trust each other all along in the process towards publication. When they come back they take their data and specimens into the laboratory for analysis.

In the laboratory, curation entails the scrutinization, classification and preparation of data. Data needs to be ordered and to be matched with other data to make it fit in the data collection. Generally speaking, this step in the research process prepares the data for analysis. In case data collection also includes the processing of specimens taken from the research site visited on the fieldtrip, a lot of paperwork has to be processed to certify for the involved authorities that these specimens have been taken legally and only for research purposes.

Analyzing data is for taxonomists the way to make sense of it. It is a method to identify and to find evidence for new species. Again, this is a very personal way of operating, where data and preliminary results are occasionally shared with carefully selected peers. Data is collected at a secured site, where it is to be analyzed which is mainly a process of ordering and categorizing.

When there is enough data collected, it is time to move to the step of drawing conclusions. Based on findings, a new species might be successfully described, or another taxonomic publication could be placed in a new light. A concept will be offered to a peer-reviewed taxonomic journal. This publication will have the appropriate structure, usually describing some species presented using a generally accepted format. One of the main purposes is to link the research results of this particular attempt to what has been published before on that particular topic.

The moment of publication signifies that the findings in that particular publication are attributed to the author(s). It usually also marks the moment when data on which the publication is based ceases to be confidential to only the involved scientists and therefore will be shared with taxonomic community. It does not only open up the data underlying the publication for peer review, it also enables these peers to use that data for their own research.

Mapping the primary taxonomic process

A taxonomist trying to explain what the profession is all about:

What is the content of taxonomic publications?

Usually a description of more than one species or occasionally of a single one. Annually, about 30-40.000 species are being described globally, and that's what the taxonomic job basically is about. For that matter you collect data and you make sure that not just anybody can get hold of it. As your dataset has grown to the right amount and has the right quality you turn it into a publication, using DELTA (DEscription Language for TAXonomy) and Nomenclature for a species name declaration. (Excerpt from single interview, 2012).

The profession of taxonomists is personal in nature, meaning that they seek cooperation with carefully selected peers. They argue they have to collaborate to get the job done and that their work will improve through discussion:

It all begins with cooperation. Starting a project means that you have to work together. After that, you will be able to generate data. Then the manipulation comes in, that you classify what you have into a centrally operated system. For that you need also other means than just Scratchpads. Personally, I prefer to work with Googledocs because it gives you the right amount of liberty, rather than filling it out on a Scratchpad. (Participant 1)

That has to do with valorization of the data, there needs to be room for disagreement. (Participant 2)

That is not necessarily a Scratchpad feature. (Participant 1)

Well, at least it should be part of Scratchpads. Not only databases, but also when you see one of your project members proposing a determination of the species, that you have the opportunity to disagree, there has to be a discussion on these matters of some sort. (Participant 2)

Collaboration and identification of specimens, there is more to it than just that. (Participant 1)

To my knowledge Scratchpads allows you to react, to speak up in case you disagree with that taxonomy, but I don't know if that is really happening. (Participant 2)

Oh, I'm sure it does (Participant 1) (conversation in focus-group interview, 2012).

They also stress that work in progress is to be treated with delicacy:

It's is all about, also with Scratchpads, it is about research in progress. It's about an author who puts a lot of effort in his project. It is certainly not a Wikipedia-like thing where exiting knowledge is being aggregated. Within a Scratchpad, you want to generate knowledge. You are looking at a dynamic system for knowledge development. It has to be clear that it takes a qualified person to do this job and that he wants to be credited for that. These are major differences if you compare Scratchpads with Wikipedia. (participant 1)

What you are actually doing is to describe a species that hasn't been described yet. This can be dangerous at times. I know about publications where others were able to publish on data which the original data collectors also wanted to publish on. (participant 2) (conversation in focus-group interview, 2012)

There is a way to solve this. When a name has been tagged to your species by yourself, no one can claim isn't yours:

We usually upload our species descriptions in Open Nomenclature. Then, if someone wants to tag a name on it, be my guest. (Participant 1)

Yeah, if you have a lot of species you want to publish on, and you don't have the time, that is the way to go. (Participant 2)

If it really matters to you that it is you giving it a name, than that is the way to go. (Participant 2)

How does it work? (Participant 3)

Open Nomenclature is just a numbered list of species, with a description of the features of them, accompanied with sketches, photographs and site of discovery. And then you can make a claim like: number 22 has been identified by me as Palmata, it gets a tag with your information on it. Consider it as an effort to standardize species classification. The problem with Scratchpads is that you cannot make a reference. The information is dynamic, subject to change. All you want is to be cited, because that is where we will be judged upon: citations. (Participant 2)(conversation in focus-group interview, 2012).

The publishing cycle as described above is considered by taxonomists as vulnerable:

I just come back from a conference where the common opinion was that what we are about to see is that taxonomic identifications will start to wander about the internet. The matter is that you end up with identifications of specimen being entered in databases, while there is no procedure on how that specific identification has emerged. That is one of the flaws of using internet content. We have to work towards a system where you can see what identifications have been done on that species, that you can trace its development. (Participant 1)

And is that likely to happen? (Participant 2)

I just hope so, because we run the risk of being suffocated with disinformation. (Participant 1)

I observed some change at Genbank. I read an article on how determinations improved through the changing of a rule. I don't do genes, but it's important enough to be mentioned. One research priority should be the establishment of a system for specimen identification, to be annotated with reliable, or at least authorized information. You need both taxonomists and bio informationists, working in a properly financed infrastructure, with online tools for alfatonomy using central databases. Maybe it should be facilitated in open-access, because usually specimen information is not stored in closed databases. (Participant 3)

I think open-access should be promoted, to facilitate a valorization procedure. It serves as a means to reach agreement on species identification. There needs to be a way to disagree and start a discussion. (Participant 1) (conversation in focus-group interview, 2012)

It seems that taxonomists feel that discussions on proper determinations of species are essential and that they need publications as a vehicle for information exchange in that respect. It looks like

Scratchpads supports the data collection process, but does not play a role in the information exchange process:

If you talk about data, I would certainly prefer Scratchpads to hold my data. But if you seek cooperation, and you need to organize things, if you choose fellow researchers to help you out, then it is important to do that through Scratchpads. In the end, all we want is to sit in our offices with our own data and files. (Excerpt from focus group interview, 2012).

That taxonomists are after all true scientists becomes clear from the following interview excerpt:

What is your unit of work? How do you think of that?

I always think in terms of publications.

Me too, my unit of work is also publications, or, maybe, at times nice specimens.

Basic information skills of taxonomists

Taxonomic practices are data intensive. This taxonomist recognizes this job requires IT data processing skills:

If you want to be good at this job you need to be computer literate. I started just after the punchcard era, but I always managed to adjust myself to new circumstances. Some of my colleagues already get confused when they have to import an Excel sheet. (Excerpt from individual interview, May 2013)

Additionally, it requires advanced skills to transfer data into Scratchpads:

It took me a lot of effort to make our Excel version to communicate with Scratchpads. As a taxonomist you should at least understand the principle of data transfer to be able to articulate your help request to make the support team fixing it. (Excerpt from individual interview, May 2013).

These skills allow you to see the limitations of plain office automation software:

I think the main reason for starting a Scratchpad was that I got stuck with lots of Excel files. They were a good way to summarize a certain part of what I was doing, but when you get specimens, doing processing and you get results, it's really hard to keep that all in the Excel's, so you get disconnections in the data you have. When thinking about trying to follow standards, best practices and so on, that was the main reason why we started our SP. Now we have that data and added photos, we have almost everything that is available is on our Scratchpad. (excerpt from a focus-group interview, 2012)

And there are also software-legacy problems involved. There are circumstances where scientists are not inclined to use state-of-the-art software applications:

My research is on small butterflies, in which I am searching for new species, trying to describe them. The problem is, you generate a vast amount of information which you need to process in order to publish. Scratchpads really helps with that. I still have databases formatted in DBase 4. I still work with them because I have the skills to do it. I

am a bit reluctant to transfer them into Access. (Individual interview with Participant, 2013).

Somebody mentions the existence of a very simple Scratchpad:

Have you seen John Johnson's Scratchpads? (participant 1)

No. (participant 2)

It's a Googledoc! [All participants are laughing] On the homepage of the Scratchpads it says the classification has been transferred to this Googledoc! It's just a sign of when real cooperation matters, it's a lot easier to work in a Googledoc instead of Scratchpads. Using Scratchpads you can also work jointly on a page, but you need to be logged in and take a two day training course, all before you have done anything. Taking one step in Googledocs means taking 20 in Scratchpads. (participant 1)

But in Scratchpads you work on your definitive output, while in Googledocs its just mainly text. (participant 3)

I work with citizen scientists and some still prefer email, they just refuse to communicate otherwise. (participant 1)

To start a Googledoc just gives you lots of freedom, and its more convenient, than filling out a Scratchpad. (participant 2)

But how do you attract attention using a Googledoc? how do you find people? (participant 1)

No, no, no, it's not the attention, it's all about collaboration. (participant 2) (conversation in focus-group interview, 2012)

The unit of communication, being the publication in the form of scientific articles is now a matter of discussion:

On some conference I heard about the concept of nanopublication. There they asked themselves, what is the smallest unit of authored information to be published on the web which can count as a publication. My feeling was, we should stay in touch with these people. (participant 1)

Wouldn't that be an idea for Scratchpads? (participant 2)

Maybe. We all have had our moments of doubt. Then you have discovered something just too small to put in an article, but which just has to be published. That you just don't see how to turn it into a full publication. (participant 3)

And if you follow that route than you don't get it accepted, journals won't have it. (participant 2)

To find a way to get those small things published would make life easier, also for journal editors. (participant 3)

I think we are right in the middle of a publishing transition, moving from a static to more dynamic ways of publishing. (participant 1) (conversation in focus-group interview, 2012)

Taxonomists feel their world is on the move. It seems like they want to keep the publication as their unit of work, together with all its attributes. How that publication is delivered is subject of discussion and they seem to be open to evaluate alternative options.

3.5 Career and community

According to Godfray, the world of biodiversity is dominated by biologists, to be divided in camps of ecologists and taxonomists (Godfray 2002). When asked about what these professions do, a taxonomist explains:

Generally speaking, taxonomists are focussing on species and ecologists on the environment or habitat. These are separate worlds slowly moving towards each other. That is because it is a matter of: can't live with them, can't live without them. But in the end they need one another. Emerging DNA techniques definitely play a role in that. The World needs taxonomists to do the barcoding that is needed in other disciplines. Only in a later stage they started to create research projects for themselves. However, these worlds just can't be fully separated. A taxonomist should at least have some knowledge of the habitat of the place of discovery and reversely, an ecologist needs taxonomy to describe the variety of species in a habitat. As a Taxonomist, I have learned the basics of ecology and we teach our students on that.

Do these notions come to the fore, are they discussed at all?

Together with some colleagues we edited a book on biodiversity. I wrote a chapter on how these worlds are related.

Do you do that more often?

No, although it was fun editing that book, but it is simply not rewarding, at all. In fact, it keeps you away from doing your job. We taxonomists are of minor importance. You see, we publish on species in specialized journals having a very low impact factor. That's because usually our colleagues use the information without a citation. These journals don't appear in Web of Science and as authors we generally have a very low h-index. So, as a taxonomist it would be easy, and in a way even be beneficial to start your own journal, because the only thing you want is to put your publication online. Electronic publishing is the new standard. You can tell that from journals also appearing in print: all articles in one issue have different publication dates.

For my job performance evaluation they ask me to give figures on my citation index, but I just refuse to do that. Let them find somebody else to sort that out. I see it as a way for managers to get up close with what I'm doing. I think their job is to support researchers instead of judging them. My work needs to be measured, of course, however a fair reference point is missing, it simply does not fit our practices. (Individual interview with Participant, May 2013).

Apart from a relationship between ecology and taxonomy, there are also ties with molecular biology. The invention and application of DNA sequencing techniques also meant new perspectives for taxonomy, or as some have felt, bringing new life to a diminishing scientific domain (Bateman 2011).

Taxonomic subfields are not equally affected by the emergence of DNA sequencing. A taxonomist working with insects explains:

We tend to become hybrid researchers. We are taxonomists and molecular biologists, and also bio-information professionals. I try to follow that path, to become a hybrid researcher. I am really trying to approach it that way.

But if you want to pursue hybrid knowledge generation, that could also mean a taxonomist and a molecular biologist collaborating, that does not have to be combined into one person?

No, no, no, I just have to understand all those worlds. (excerpt from a focus-group interview, 2012)

In the field of paleontology it is a different situation. Here it is a combination of scientist specializing in a panoply of fields and subfields, all working together:

One of the things that is definitely going to happen is, at least in my field which is paleontology is a trend towards mega projects. And in these projects, specialists like paleontological climatologists, oceanographers, molecular biologists, biodiversity specialists, all having something to say about history. Unlike bio information specialists, I work strictly specimen-based. I am suspicious about datamining. I mean, that you only use published material without any reference, without thinking about how determinations of species may change over time without considering multiple interpretations, that really worries me. That kind of research is booming, almost a tradition. (excerpt from a focus-group interview, 2012)

A zoologist is also moving beyond the limits of her field of research:

Well in terms of my research may help or the direction that I'm thinking of now that I certainly need information from other fields. In general from, if you work in marine bellies that is working pretty well actually with oceanography, geography and such things. You need the currents, you need physical properties of different water bellies and such. (excerpt from a focus-group interview, 2012)

Somebody working on plants explains wants to remain more to herself:

I still work on my plants and I make sure not to be disturbed by colleagues of any kind. I would like to collaborate with others, but only in my own field. So I really don't work together with others and I'm not actively seeking contact with others to make that happen, maybe at a later stage I will.

What keeps you from doing that?

Well, I don't think there are that many specialists that could make a contribution to what I am doing. And the ones that could are just too busy doing other things. I want them to be involved, and that is something that eventually will happen, but right now it's only my concern. (excerpt from a focus-group interview, 2012)

There is always hope for the better. This taxonomist sees different worlds coming together:

I think, it is not a matter of ruling things out, like disqualification of others. Professional taxonomists and citizen scientist, they just can't rule out one another. Or molecular

biologists and taxonomists, even if they hate each other, they just shouldn't. They need each other. And what I do hope for, is that, simply, that it all unifies. That people do a proper taxonomy and that taxonomists consider molecular data. That's what I would like to happen. (Individual interview, May 2013).

Biology and taxonomy are scientific disciplines mainly based in and originating from the Western world. Data collection, however, is a global practice, being done in remote places all around the world. That would imply taxonomists from all over the world working together, forming a tight community. A plant taxonomist explains about his specialty:

My research is on ferns growing on one specific big island. I know who my peers are, what data they have and what they want to do with it. I am looking for a way to exchange species identifications, to add their identifications to my specimen data. (excerpt from a focus-group interview, 2012)

Someone mentions in a focus group how the taxonomic world got smaller and smaller:

You can use Google to find out who is who in your field. Sometimes we put an ad on our Scratchpad that we are in need of descriptions of some kind, or materials to be DNA sequenced. Then it's just a matter of seeking attention.

How would you do that before, using a mailing list?

Yeah, for long that was the way to go, to assemble a long mailing list to get in touch with all those specialists. The last time I did that five out of a hundred responded and we got together in a meeting and that small group still exists. One of them was somebody from Italy, who managed to be a gatekeeper for a Group of our students to do some fieldwork there. These scientists have uploaded their data on our Scratchpad which attracted other people as well who found us through Google. A lot of them were on our initial mailing list, when they saw what we did on our scratchpad, but also some who were not. The last time we did a count it were 44. (excerpt from a focus-group interview, 2012)

Generation issues are hardly mentioned, and if so, they are related to computer literacy:

Don't underestimate how many of the older generation have trouble to use IT tools. They send you an email with data and want you to copy-paste it into the system. You see the same thing in editorial management systems that have been put in place at certain journals. You just don't want to know how many editors and authors try to skirt such systems by sending emails. It drives me mad. These editorial management systems can get really complex, and their purpose should be to make life easier, not complicated. (excerpt from a focus-group interview, 2012)

Generally speaking, taxonomists are well aware of how their research is valued. They are driven by their own research interests and feel they are part of a small, internationally affiliated research community. It is obvious that DNA sequencing techniques have changed the research field and that it calls for more integration and collaboration in research project. If this is a matter of turning oneself into a multidisciplinary scientist or a collaboration of specialists seems to depend on the subfield.

3.6 Datasets and information

According to our data there are different kinds of data definitions in taxonomy. First, there is the publication, a peer reviewed piece of information on one or more species. A publication marks the process of bringing information collected by mostly collective of taxonomists to the open. After publication, its contents can be used, cited and referenced.

The open data collection Nomenclature is mostly being referenced as a database. A taxonomist explains:

Nomenclature is a non-scientific classification system. What I mean is that it's basically a system to sort the species into the proper position. It's nothing more than a sorting system, with the intention to give it a name. Zoology, botany or bacteriology, they all have their own unique systems with their own unique classification procedures. In a way it is comparable with astronomy. When a new star has been discovered there is a procedure to be followed, authored by an internationally recognized organization to give it a name. However in nomenclature, scientific relationship rules have to be followed based on taxonomic research. (Individual interview, May 2013).

The International Nucleotide Sequence Database Collaboration is holding an open access database called GenBank, with an annotated collection of all publicly available DNA sequences. The Global Biodiversity Information Facility (GBIF) is an international organization making scientific data on biodiversity available through the Internet. These are examples of sites with databases holding taxonomic data, while there are a lot more sites on the internet where comparable data can be found^{viii}. The obvious and puzzling question is: why not integrate all these different databases into one information source on biodiversity? We asked that question and here are two of the answers:

Oh yeah, that would require a sociological study. How all these systems have emerged, how they were developing in their own way and why they are unique. In one discipline you have a Nomenclature commission, in others there are congresses to reach agreement on names, they all follow different paths. Moreover, there are boundary objects that could relate to more than one domain. You have bacteria that are also recognized as animals and vice versa. The same holds true for plants and bacteria. And there are names appearing in multiple nomenclature having different meanings. So to bring that all together would be a hell of a job. (Individual interview, May 2013)

There are so many databases and it is sad they don't integrate into each other, and it is difficult because you have to maintain more than one database. And many are more or less the same, they have the same purpose or goals. So I believe it would be so much better if we have one database. (excerpt from a focus-group interview, 2012)

Some of her colleagues from other domains are more optimistic:

Yeah, I mean, there is just so much data, and there are so many different databases right now, so if you can bring those together in one place. (participant 1)

Then you need new skills and new ideas: how do you think this will work? (participant 2)

There are different types of data which we will be able to integrate them in some way. I mean, it is probably overwhelming, and without these, hopefully in the coming years, hopefully there will be tools to make it possible to actually do these things. Like climate

data and those things that's one of the things. We really have to have tools to be able to analyze that. The other part is to analyze molecular biodiversity data. Now molecular data has started flowing now, it is totally enormous, even today there aren't enough tools and people that can deal with them, so, in terms of IT that is really needed. (participant 1)

What's the logic behind it? (participant 3)

The more data you can analyze together, the more questions you can press. And without the tools, you can't press anything. Hopefully products like life watch would be matured in twenty years, that would include data from all over the globe, not just Europe or specific countries, to be integrative. (participant 1)

In this taxonomic environment, being a world of dispersed publications, information and data, scratchpads tries to find its niche. A taxonomist remembers how the project started from a rather idealistic view:

Scratchpads started as an initiative from the EDIT project, and everybody had warm feelings about it. But the initiator has a name for having good ideas that just don't materialize. Then Vince became impatient and started to think practically. He took the space created by others to find a solution and turned it into a tangible product. It was him who came up with the idea to develop it in Drupal. (Individual interview, May 2013)

Scratchpads are recognized as just one of a set of possible resources, however with specific features. Taxonomists may use it in more than one way as they see fit. First of all it is recognized as a tool to organize the data at hand:

I haven't used a Scratchpad, but I want to believe that it helps to organize my own data. You can use it as an open database to focus on one thing, on one object, right? So you can have it like a small database for yourself. It can help you to organize lab work, fieldwork, and of course to present findings. (excerpt from a focus-group interview, 2012)

Some would like to use it for their own data, but rather would like it to interactively connect it to their own database:

The question is, why would you have a Scratchpad, because my own system allows me basically to enter all the same information that I could enter into the Scratchpad until now. I mean, I'm interested in identification keys and I have a taxonomy on Scratchpads. But that never really worked, because previously I worked with my own system, importing my work on Scratchpads. And now I wonder, should I maintain my Scratchpad manually as well. It is just too much work, so what I really would like, is some way to keep it interactively updated. We have many different resources, so I don't need scratchpads in particular. (excerpt from a focus-group interview, 2012)

Somebody else mentioned Scratchpads is helping to organize his own database:

The idea was that you want to bring all information on none species together at one place. And that keeps you from having to endlessly search the literature. Well, to be

honest, I still have to browse the literature, which I have also on my scratchpad, but I don't do that very often. However, I use it frequently to check upon a taxonomy because its the only place where you have all the information available at your fingertips. Except of course my own computer where I have it also on file. And if I need additional information on say an author or the year of publication then I. . . (participant 1)

You use the search function on Scratchpads? (participant 2)

Occasionally, but mostly I prefer Google where Scratchpads ranks first in the hitlist. (participant 1)

You actually are using it as a reference to your own data? (participant 2)

Occasionally, I do. The literature in it is far from complete, so I hope the other members help to complete it. And that would also be a thrust to use it myself, because that's the whole idea. And that I would use it to organize my illustrations, but I guess I'm not ready for that. (participant 1)

How do you keep them now, in folders? (participant 2)

Yeah, they are stored in a rather dispersed way. And one of the frustrations of Scratchpads is the time it takes upload these pictures. I mean, to upload the image itself is a piece of cake, but adding metadata is like hell, it's almost like doing it manually. And I already have it in a database, but there is just no easy way uploading it. That gives a lot of frustration. (participant 1)

And have you already uploaded all your source information, one at a time? (participant 2)

Yeah, it was relatively easy. And now, looking up and checking things in Scratchpads is relatively easy, but my databases are not interconnected. I did not have a interlinked databases, so my literature has been separated from my taxonomy. So to link articles with taxonomies has to be done manually. (participant 1)

And how do you share your photos? What kind of tool do you use for that? (participant 3)

We put them on Picasa. (participant 1)

Picasa? You also see people using Facebook for that, both work real smooth. (participant 3)

Well, if you do it within Scratchpads or with a link to Facebook. Facebook has the advantage to be used on your smartphone.

This observation again highlights that Scratchpads is seen as a service to taxonomy and definitely is not the embodiment of a taxonomy itself.

4. Analysis and conclusion

Now we have presented our findings in section 3 using the framework developed in section 2 it is time to analyze and to draw conclusions.

Taxonomists used to define their scientific work as being supportive to other biological and non-biological realms. The public attention given to a diminishing global biodiversity has led to increased self-esteem of the whole scientific sphere. Taxonomy is hot again. However, it looks like ecologists are taking the lead in the biodiversity discussion, while they are aware that they depend on taxonomy in their work. Likewise, taxonomists need the microbiologists to advance their within-species taxonomic work further using DNA sequencing. For all these efforts a taxonomic infrastructure is inevitable and we are about to depict what kind of infrastructure that is and how it is and should be developing.

4.1 Analysis

Taxonomists see their world as moving towards integration. Probably under influence of the quest for biodiversity information the trend in taxonomy is towards integration. Having this said, the question is: are taxonomists coming from a world of specialists focusing on only a small part of the classification of nature, ready to collaborate in the wider and global project of knowledge creation on biodiversity?

Our data shows that taxonomists keep their ways of working firm in place. As infrastructures become increasingly electronic and web-based, they are utilized to span distances between scientists within taxonomic subfields, rather than between subfields as a whole. And if taxonomists do move beyond their professional borders, it is either aimed at colonizing new techniques within their own subfield, as it is the case with DNA sequencing, or it is to establish relationships between their own taxonomic subfield and other disciplines outside taxonomy, like paleontology or oceanography. These outcomes suggest that taxonomists are still driven by the urge to be supportive to fields outside taxonomy, like they have been for centuries.

Producing publications in scientific journals still forms the core of taxonomic academic life. Publishing an article is the product of a validation process, it means the process of going from crude data to information that has been approved by peers and which is fit for academic debate, rather than the end product as a building block for mapping biodiversity. since publications form the ultimate reference for new taxonomic work. Commencing a new taxonomic publication is in essence a process in which newly acquired, raw data is redressed and ordered to make it fit in the wider taxonomic framework, using multiple sources of data and metadata as a framework.

Taxonomy is envisioned mostly as stable system of descriptions of human life (Godfray 2002; Berendsohn, Güntsch et al. 2011). Our findings suggest that taxonomy is utterly dynamic. It is a scientific field with an intrinsic inclination for debate and discussion. The notion of a world in motion is quite common among taxonomists, but always has been related to the fact that we still don't have a complete picture of life on earth, which makes it obvious to direct all efforts towards completing that picture (King, Morse et al. 2011).

Scratchpads as a data management tool is in our view just one, rather important element of the taxonomic process. It supports a specific part of the taxonomic process: converting crude data into full-fledged scientific publications. It certainly helps to improve varying taxonomic practices as all taxonomic subfields have something to gain from improved data management tools. It's influence on bringing about a bibliography of life is according to our findings rather low.

4.2 Conclusion

The findings of this research pose the question: what is an infrastructure. We have seen efforts to bring about an infrastructure which are aimed at the integration of systems (Ingersoll, Seastedt et al. 1997; Bowker 2000; Berendsohn, Güntsch et al. 2011; King, Morse et al. 2011). These systems are envisioned as archives for data storage, either being constituted as a collection of filing cabinets or an electronic data storage center. The taxonomic world has been tempted to integrate, systematize and categorize taxonomic output into a meaningful whole and to restrict that effort only to results.

By producing an ethnographic film of taxonomy we have brought the dynamics of the taxonomic publishing process to the fore. There is only one conclusion to be drawn: the taxonomic effort is intrinsically dynamic. Taxonomy is a process of descriptions being doubted, contested, altered, improved and supplemented. This very process forms the heart of taxonomy, not the taxonomic end products.

Bowker wrote about the challenges he saw in relationship to the development of an infrastructure for mapping biodiversity (Bowker 2000). In that, he seriously questioned the efforts of bringing about a system on biodiversity. However, it was suggested that such a snapshot of biodiversity was a viable option. Using these notions as a starting point, we suggest instead to take the taxonomic publishing process as a reference point for developing such an infrastructure. Our conclusion is that an infrastructure by all means must be taken as a process instead of a system.

If we want to assess the impact of a taxonomic data management service like Scratchpads on taxonomy as a whole we have to judge the system as such as being a part of the global taxonomic community together with all these databases and services forming the taxonomic infrastructure, trying to depict life as we know it so far.

References

- Barley, S. (1986). "Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments." *Administrative Science Quarterly* **31**: 78-108.
- Bateman, R. (2011). The perils of addressing long-term challenges in a short-term world: making descriptive taxonomy predictive. *Climate change, ecology and systematics*. D. Hodgson, M. Jones, S. Waldren and J. Parnell. Cambridge UK, Cambridge University Press. **78**: 67-95.
- Berendsohn, W., A. Güntsch, et al. (2011). "Biodiversity information platforms: From standards to interoperability." *ZooKeys*(150): 71-87.
- Berners-Lee, T. and J. Hendler (2001). "Scientific publishing on the semantic web." *Nature* **410**: 1023-1024.
- Bowker, G. (2000). "Mapping biodiversity." *International Journal of Geographical Information Science* **14**(8): 739-754.
- Bowker, G., K. Baker, et al. (2010). Toward information infrastructure studies: ways of knowing in a networked environment. *International handbook of internet research*. J. Hunsinger, L. Klastrup and M. Allen, Springer: 97-117.
- Bowker, G. and S. Star (2000). *Sorting Things Out: Classification and its Consequences*. Cambridge MA, The MIT Press.
- Bowker, G. C. (2000). "Biodiversity datadiversity." *Social Studies of Science* **30**(5): 643-683.
- Brown, B., M. Chui, et al. (2011). "Are you ready for the era of 'big data'?" *McKinsey Quarterly*(4): 1-12.
- Busby, J. (2002). Biodiversity mapping and modelling. *Environmental Modelling with GIS and Remote Sensing*. A. Skidmore. London UK, Taylor and Francis: 145-165.
- Cox, C. and P. Moore (2005). *Biogeography: an ecological and evolutionary approach*. Malden MA USA, Blackwell Publishing.
- Dicke, W. (2001). *Bridges & Watersheds; A narrative analysis of watermanagement in England, Wales and the Netherlands*. Amsterdam, Aksant Academic Publishers.
- Edwards, P., S. Jackson, et al. (2007). Understanding Infrastructure: Dynamics, Tensions, and Design; Report of a Workshop on "History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures.

- Gallivan, M. (2001). "Striking a balance between trust and control in a virtual organization: a content analysis of open source software case studies." *Information Systems Journal* **11**(4): 277-304.
- Gaston, K. and J. Spicer (2004). *Biodiversity: an introduction; Second Edition*. Malden MA USA, Blackwell Publishing.
- Godfray, H. C. J. (2002). "Challenges for taxonomy." *Nature* **417**(6884): 17-19.
- Harris, R., P. Sleight, et al. (2005). *Geodemographics, GIS and Neighbourhood Targeting*. Chichester UK, John Wiley & Sons.
- Hawksworth, D. (1995). The resource base for biodiversity assessments. *Global biodiversity assessment*. Cambridge University Press, Cambridge, United Kingdom: 549-605.
- Hobbie, J., S. Carpenter, et al. (2003). "The US long term ecological research program." *BioScience* **53**(1): 21-32.
- Ingersoll, R., T. Seastedt, et al. (1997). "A model information management system for ecological research." *BioScience*: 310-316.
- Kainz, W. (2000). The foundation technologies. *Geospatial Data Infrastructure: Concepts, cases and good practice*. R. Groot and J. McLaughlin. Oxford, Oxford University Press.
- King, D., D. Morse, et al. (2011). "Towards the bibliography of life." *ZooKeys*(150): 151-166.
- Koerten, H. (2011). *Taming Technology; the narrative anchor regulating technology in geoinformation infrastructures*. Delft/Amsterdam, Delft University Press/IOS Press.
- Koerten, H. and M. Veenswijk (2010). "The Social Construction of Public Infrastructure: The Case of the Dutch National Geoinformation Clearinghouse Project." *International Journal of Business Anthropology* **1**(2): 48-69.
- Koerten, H. and M. Veenswijk (2011). Thinking in circles: how National Geo-Information Infrastructures cannot escape from the temptation of technology. *Spatial Data Infrastructures in Context; North and South*. Z. Nedovic-Budic, J. Crompvoets and Y. Georgiadou. New York, NY, CRC Press/Taylor & Francis: 137-153.
- Lanier, J. (2010). *You Are Not a Gadget*. London UK, Penguin Books Ltd.
- Latour, B. (1996). *Aramis or The Love of Technology*. Cambridge MA, Harvard University Press.
- Leadbeater, C. (2010). *Cloud Culture; the future of global cultural relations*. London, UK, Counterpoint.
- Mosco, V. (2004). *The Digital Sublime*. Cambridge MA, MIT Press.
- Noveck, B. S. (2009). *Wiki government: how technology can make government better, democracy stronger, and citizens more powerful*, Brookings Institution Press.
- Olivieri, S., J. Harrison, et al. (1995). Data and information management and communication. *Global Biodiversity Assessment*. V. Heywood and R. Watson. Cambridge UK, Cambridge University Press: 611-670.
- Raper, J., D. Rhind, et al. (1992). *Postcodes: the new geography*. Essex UK, Longman Scientific and Technical.
- Raven, P., B. Berlin, et al. (1971). "The origins of taxonomy." *Science* **174**(4015): 1210.
- Schejter, A., Ed. (2009). *... And Communications for All; A Policy Agenda for a New Administration*. Plymouth UK, Lexington Books/Rowman & Littlefield Publishers Inc.
- Smith, V. and L. Penev (2011). "Collaborative electronic infrastructures to accelerate taxonomic research." *ZooKeys*(150): 1.
- Star, S. (1999). "The Ethnography of Infrastructure." *American Behavioral Scientist* **43**(3): 377-391.
- Star, S. and K. Ruhleder (1996). "Steps toward an ecology of infrastructure: Design and access for large information spaces." *Information systems research* **7**(1): 111-134.
- Van Marrewijk, A. and M. Veenswijk (2006). *The Culture of Project Management; Understanding Daily Life in Complex Megaprojects*. Harlow UK, Pearson education Limited.
- Von Hippel, E. and G. Von Krogh (2003). "Open source software and the "private-collective" innovation model: Issues for organization science." *Organization Science* **14**(2): 209-223.
- Watson, R., V. Heywood, et al. (1995). "Global Biodiversity Assessment; Summary for Policy-makers."

ⁱ <http://scratchpads.eu/>, accessed February 2013

ⁱⁱ All quotes in this section: <http://wp5.e-taxonomy.eu>, accessed April 2013.

ⁱⁱⁱ EDIT Information Science & Technology Commission (ISTC), minutes of the 2nd Meeting, September 10-11, 2007, Botanic Garden and Botanical Museum Berlin-Dahlem

^{iv} EDIT Success Story, ViBrant: Following up an EDIT success, Smith, V. and Roberts, D., Natural History Museum, London <http://www.e-taxonomy.eu/files/EDIT%20success%20VIBRANT.pdf>, accessed April 2013.

^v <http://scratchpads.eu/about/resources>, accessed May 2013.

^{vi} <http://wp5.e-taxonomy.eu/>

^{vii} <http://www.e-taxonomy.eu/files/EDIT%20success%20VIBRANT.pdf>

^{viii} http://vbrant.eu/sites/vbrant.eu/files/D2.3_Sustainability.pdf