

A Typology of Virtual Research Environments

Iftekhhar Ahmed
University of North Texas
Iftekhhar.ahmed@unt.edu

Marshall Scott Poole
University of Illinois
mspoole@illinois.edu

Ashley B. Trudeau
University of Illinois
abt2@illinois.edu

Abstract

Virtual Research Environments (VREs) are online spaces that support communication and collaboration among scientists. Hundreds of VREs have been constructed using various configurations of research tools and information and communication technologies (ICTs) to serve many disciplines and interdisciplinary inquiry. This study characterizes a large sample of VREs in terms of the research and ICT resources they incorporate and derives a typology of VREs based on their particular ICT configurations. The four types are correlated with previous VRE typologies and disciplinary domains. Results indicate that there are correspondences, but that types of ICT configurations also exhibit complex relationships with function and discipline.

1. Introduction

In contrast to the myth of the lone scientist tenaciously pursuing truth, the majority of scientific research is done by groups rather than by individual scientists [1]. Increasingly common are extended research groups, “very large, unified, cohesive, and highly cooperative research groups that are geographically dispersed yet coordinated as though they were at one location and under the direction of a single director” [2, p. 407]. These dispersed, often large, research groups must coordinate their work via communication at a distance, which requires novel tools for conducting science and for collaboration [3]. One resource for extended research groups are sites for virtual scientific collaboration. These sites, which we will call “Virtual Research Environments” (VREs) have also be variously referred to as “collaboratories,” “cyberenvironments,” and “virtual laboratories.” William Wulf envisioned a collaboratory as a “‘center without walls’, in which the nation’s researchers can perform their research without regard to geographical location - interacting with colleagues, accessing instrumentation, sharing data and computational

resources, and accessing information in digital libraries” [4, p. 40].

VREs have become a key part of national research policy, and they have received significant amounts of support and resources as a result. In the U.S., the landmark Atkins Report (2003) advocated the development of *cyberinfrastructure* for research, which refers to infrastructure based on distributed computer, information, and communication technology [5]. The U.S. National Science Foundation (NSF) played a major role in the development of VREs in the United States, through its Office of Cyberinfrastructure [6]. Similar initiatives have been initiated in the United Kingdom, Australia, India, France, Germany, Brazil, and Japan, among other nations, as well as through international collaborations [7].

The first VREs were established to support what are traditionally called the “hard” sciences but VREs soon began to develop for all areas of research. Anandarajan and Anandarajan (2010) stated, “Without the physical boundaries of traditional social networks, online social networks replicate and enhance the benefits of traditional social networks across time and space and accelerate and globalize the process” of research [8, p. 7]. VREs are of particular scholarly interest, because the ICTs to support various forms of work and communication are fundamental to the effectiveness of the research process in VREs.

VREs vary in terms of the type of activities they promote, from simple downloads of data by a single research group to online collaborative projects. As such, research on virtual collaboration [9, 10, 11] has potential to inform the design and facilitation of research in VREs. Research on online communities also provides a useful frame for understanding VREs [12]. However, due to the unique nature of scientific inquiry and the specialized knowledge and expertise required of scientific team members, scientific teams differ from traditional user or interest-based communities and from the business and lay decision-making teams that is the focus of the virtual collaboration literature. Scientific projects typically involve much longer time scales than are envisioned in virtual collaboration models. The stages of scientific

inquiry (e.g., design, data collection, etc.) require specialized activities that require specific types of support. Scientific communities are based on disciplines with long and deep histories and so have prefigured structures that are more determined than those of most online communities. The institutional frameworks of the various sciences and the size and complexity of modern scientific projects set up requirements that are more specific than the categories of the extant literature on virtual teams or online communities.

There is a growing literature on VREs specifically, how they operate, and what makes them effective [13, 14]. However, a key gap in the current literature is the relative lack of attention to the technological infrastructure of VREs. While the literature clearly acknowledges—even celebrates—the critical role of technology in enabling VREs, it tends to foreground scientific and social processes and treats technology as the “handmaiden” of scientific activity in the VRE [5, 15]. Bos, Zimmerman, Olson, Yew, Yerkie, Dahl, and Olson’s (2007) comprehensive typology [16] identifies seven types of “collaboratories” that can support scientific communities based on their organizational configuration and the scientific functions they serve:

- Shared Instrument: remote access to expensive scientific instruments, such as a telescope.
- Community Data System: data and information repository created and maintained by a geographically-distributed research community.
- Open Community Contribution System: an open project that aggregates efforts of many geographically separate individuals toward a common research problem, such as determining protein folding.
- Virtual Community of Practice: a network of individuals who share a research area and communicate about it online, but are not actually undertaking joint projects.
- Virtual Learning Community: cyberinfrastructure to support knowledge sharing in an area.
- Distributed Research Center: a space to aggregate scientific talent, effort, and resources unified by a topic area of interest and to support joint projects on that topic.
- Community Infrastructure Project: a distributed project to develop common resources that facilitate science, such as software tools, standardized protocols, or new types of scientific instruments to further work in a particular domain.

Bos et al.’s typology specifies the purposes to which VREs can be put, and some of the collaborative and community processes and structures involved. But if we consider that VREs are sociotechnical systems, it is also important to consider the ICT side of the

equation. Configurations of technologies and ICTs shape the ways in which scientists and technologists work and communicate during the research process, the nature of the research that can be carried out in a VRE, and how research results are disseminated. Investigation of what Leonardi (2012) calls the materiality of a sociotechnical system, “the arrangement of an artifact’s physical and/or digital materials into particular forms that endure across place and time” [17, p. 42], is an important complement to study of its organization, practices, and processes. Understanding the intersection of scientific work and collaboration and IT in VREs would provide an opportunity to develop a collaboration engineering approach to the design of VREs [9].

There are many case studies of ICT platforms in VREs, with specific descriptions of ICT development and its challenges and opportunities [18, 19]. These suggest that there is considerable variability in the design and implementation of VREs. On a practical level, limitations in time and resources and in the imaginations of developers and users are likely to result in variation. Moreover, developers of VREs differ in their opinions about what ICTs best support research. As a result, VREs incorporate unique, specific configurations of technologies that designers and users believe are necessary or desirable depending on the task and context at hand. To date there has been no broad and systematic survey of ICT configurations in VREs that attempts to map these variations. This study addresses this gap by characterizing the configurations of ICTs in a substantial sample of VREs from multiple disciplines spanning the sciences, social and behavioral sciences, and the humanities. Our goal is to develop a general typology of VREs based on the ICTs they incorporate, with particular emphasis on communication, collaboration, data-sharing, and other ICTs that have the potential to support research collaborations.

The nature of collaboration could be very different for different teams and also at different phases of teamwork. Collaboration is a joint effort towards group goals that “involves multiple individuals who combine their efforts to achieve mutually desired outcomes” (9, p.122). By linking ICT configurations to functions and disciplines of the VREs, we hope to provide a base-level understanding of VREs that can be a foundation for more in-depth investigation. This study can also contribute insights into the “implicit theories” of designers of VREs and move toward a more systematic collaboration engineering approach for VREs.

The next section discusses ICTs in Virtual Research Environments and advances several research questions that guide this study. Following this, we consider disciplinary differences that might result in

different VRE configurations being preferred by different disciplines and advance some related research questions. Then we describe the methodology for this study and report results related to the research questions.

2. Virtual Research Environments and Their ICTs

Building on a framework advanced by Kouzes, Myers, and Wulf (1996), we can distinguish five general types of ICT resources VREs may provide:

Instruments, tools that enable the observation, collection, gathering, preparation, and other operations involved in producing data for scientific or scholarly inquiry. Examples include shared telescopes and microscopes, and online environments for experiments and surveys.

Data, including primary source data (the direct outputs of instruments), compilations of data from multiple primary sources, and additional information from computational analysis, which might include meta-data and provenance records. One example of primary source data is CAT scan recordings, which might be compiled into a collection, analyzed to yield a derived image from computational processing of the “raw” record, and then tagged with meta-data to enable search and retrieval.

Computational Resources, which include resources for ingestion, storage, processing, and analysis of data and also for the operation of instruments and support of publishing. Physical and virtual machines and software are included in this category.

Publishing resources support the documentation of activities in the VRE and the preparation of and curation of data, scholarly articles, and reports. Publishing includes documentation of instruments, datasets, and computational resources, especially those specially developed for the VRE. Publishing in VREs is done not only by scientists and scholars, but also by developers, data curators, and managers associated with the VRE.

Communities, which include groups of scientists/scholars that have developed relationships mediated by the VRE. Most immediate is the user community that utilizes the VRE either individually or collaboratively. Surrounding this may be a larger community of practice that provides a culture that informs practices in the VRE and the disciplines involved in the VRE, which have their own specific practices and collaborative tendencies.

VREs vary in their specific configurations of ICTs based on what their designers and users believe are

necessary or desirable. Hence, different VREs are likely to display different degrees of emphasis on the five types of resources and may also differ on the specific ICTs they incorporate. In order to characterize the field of VREs in terms of technology, we will address the following research question:

RQ 1: What ICTs do the VREs incorporate?

Identifying the specific ICTs that occur in a VRE gives us a technological profile of that VRE, and based on a sample of VRE profiles we can attempt to address a related question:

RQ 2: Can we identify specific types of VREs based on their technological configurations?

If distinctive types of ICT configurations are identified, the question arises as to how they relate to functional typologies of VREs. A relationship would suggest that ICT configurations are related to specific scholarly activities.

RQ 3: If we can identify specific types of VREs based on ICTs, how do these relate to existing typologies of VRES based on function or purpose?

3. Disciplinary Differences in VRE ICT Configurations

One factor that seems likely to differentiate the ICT configurations of VREs is the discipline(s) the VRE serves. There are several grounds on which disciplines might differ in terms of their ICT requirements in VREs. First, different disciplines utilize different tools. Astronomers rely on expensive and exclusive telescopes leading them to depend on access to instruments. Ecologists and also humanists are more likely to study collections of specimens, records, texts, and so depend on databases and perhaps collective efforts at gathering objects of study. Second, disciplines differ in their emphasis on scholarly collaboration. Studies have found that 80-90% of articles in leading journals in biology, physics, and chemistry were collaborations, whereas this figure is around 50-60% for social science journals (political science, sociology, anthropology, and psychology) and 10% for humanities [20, 21]. Disciplines also vary in the degree to which they collaborate remotely. For example, particle physics and mathematics have longer history of using computer-mediated communication technologies than other disciplines [22].

Number of publications expected from scholars also plays a significant role in collaborative work. For example, chemistry is one of the leading disciplines in publication per scholar [23]. We can also observe differences among disciplines in computational intensity, amount of data and data storage capacity, and technology use. Traditionally, we find social science

and humanities are low in computational requirements, whereas various ‘hard sciences’ such as astronomy and physics, are high.

In this study, disciplines are grouped into disciplinary domains, including the physical sciences (e.g., physics, chemistry), natural sciences (e.g., geology, paleontology), biological sciences (biology, biochemistry, medicine), social sciences (including behavioral sciences), and humanities. Across these disciplines we can ask the following question:

RQ 4: If we can identify specific types of VREs, do some disciplines favor particular types?

To develop more specific expectations about disciplinary relationships to VREs, we can draw on Becher’s (1989) useful classification system which defines dimensions for distinguishing disciplinary domains [24]. The *hard versus soft* dimension distinguishes “hard” scientific disciplines such as physics and astronomy, which have clear, consensually defined problems, are organized around a few widely accepted and formalized (often mathematically) theories, and build up cumulative bodies of knowledge from “soft” scientific disciplines such as biological sciences and most social sciences, which have multiple, broadly defined problems, a variety of research approaches, numerous theories that span a broad array of topics, and diverse “islands” of findings.

The *convergent versus divergent* dimension describes the organization of scholarly communities. In “convergent” disciplines such as physics and astronomy the research community is fairly tightly knit, a large proportion of members work on a few problems, and there is a connected network of scholars through which information diffuses rapidly. In “divergent” disciplines such as the humanities, the research community is comprised of many smaller clusters of scholars working on many problems and a less dense network of clustered groups with not as much information exchange among clusters as in convergent disciplines.

In general, we would expect physical and natural science VREs to have more instruments than other domains. Data resources and repositories should be quite common in biological sciences and humanities, both of which commonly rely on collections (and in the biological sciences there are also vast datasets of clinical, genomic, and other types of data), somewhat less common in the social and natural sciences (which develop diverse datasets that are often shared with other scholars), and least common in the physical sciences. Computational resources should be more common in the better developed and funded disciplinary domains (physical, natural, and biological sciences) compared to the social sciences and humanities. Community resources should be more

common in convergent disciplines (physical and natural sciences) than in divergent disciplines (humanities, social sciences, biological sciences).

4. Procedures

To address these research questions, we compiled a large sample of VREs and gathered data on the ICT features they employed. This involved developing a list of specific technological features under each of the five major categories of VRE resources described previously via inductive coding procedures. We then enumerated the technologies available within each of the VREs in the sample. Cluster analysis was employed to derive a typology of VREs based on technological configurations.

4.1. General approach

The procedure used in this study utilized a usability inspection approach for identification and analysis of characteristics and attributes of websites [25]. The method starts with identification of high-level quality characteristics (i.e., instruments, community) that provide a conceptual framework for design requirements for a website. Then these characteristics are decomposed into multiple levels of sub-characteristics, and finally, a sub-characteristic is mapped into a set of measurable attributes. We identified five general categories of VRE resources as high-level characteristics, then inductively identified list of specific ICT elements or applications as sub-characteristics in each high level group, and finally specified observable features (attributes) to facilitate systematic identification of ICT applications.

4.2. Sampling

The “Collaboratories at a Glance” (CAG) project offers a comprehensive list of VREs classified according to Bos et al.’s (2007) seven-fold scheme [26]. As our objective was to investigate research-focused VREs, we sampled only the four types of research-focused VREs identified by Bos et al: Distributive Research Center (DRS), Shared Instruments (SI), Community Data Systems (CDS), and Open Community Contribution Systems (OCCS). To supplement the “Collaboratories at a Glance” listing we conducted an extensive web-based search to identify additional research-focused VREs using the following keywords: *VRE, collaboratory, distributed research environment, e-Science, virtual research infrastructure, e-research, online research*

environment, virtual scientific collaboration, and online research collaboration.

A total of 251 research-oriented VREs were identified. Removing those VREs that were “bare bones,” and did not have enough tools and technology information available in their portal, reduced the sample to 232 VREs: 69 DRS, 62 CDS, 60 OCCS, and 41 SI Portals. In terms of disciplines, 82 VREs dealt with Biological Sciences, 38 Humanities, 33 Multidisciplinary Hard Sciences, 23 Natural Sciences, 23 Physical and Chemical Sciences, 20 Multidisciplinary Social Sciences/Humanities, and 13 Social Sciences. In order to have balanced sample sizes for all four types of research-oriented VREs, we used all 41 SI VREs and randomly sampled 41 from the other three types of VREs for a total sample of 164 VREs. Of these there were 54 VREs dealing with the Biological Sciences, 17 with Natural Sciences, 20 with Physical and Chemical Sciences, 10 with Social Sciences, 27 with Humanities, 13 with Multidisciplinary Social Sciences, and 23 with Multidisciplinary Hard Sciences.

4.3. Data Collection

We used grounded coding procedures to develop a list of ICT features that could be used to code the VREs. For each successive VRE, the major types of ICTs were identified and any that differed from the current list were added. These ICTs were initially recorded based on the label they were given in the VRE, along with information about the purpose and function of the ICT. Labels as well as purpose and function discerned from exploring the VRE and its features served as the attributes that indicated sub-categories. Different VREs employed different labels for the same ICT feature, so as a second step we classified each of the identified features into sub-categories based on their purpose, function, and resemblance to commonly known ICTs (e.g., blogs). As we explored further VREs, we expanded the list. When we reached a point of theoretical saturation and could no longer identify new features, 32 ICT sub-categories were identified, and this was used as the final ICT feature list for data collection. The next step was to match the sub-categories with the five main categories: instruments, data, computational resources, community, and publishing. The five general categories were sufficient to incorporate all 32 ICTs.

In coding the data, we used a binary measure (presence or absence of that particular feature) for each ICT. Three different coders analyzed the VREs. To assess inter-coder reliability, all three coded three randomly selected VREs, with 99% agreement at the specific feature level.

4.4. Data Analysis

Three ICTs—chat systems, conferencing systems, and podcasts—were excluded from analysis due to a low level of occurrence (less than 2 occurrences in all 251 VREs in the sampling frame). Our final data set consisted of 164 VREs and 29 ICTs distributed across five resource types. This sample provided sufficient diversity to support derivation of the typology, while being manageable in terms of coding.

Cluster analysis (Ward’s method) was used to derive meaningful groups of VREs based on their ICTs. Ward’s method is a hierarchical clustering approach that minimizes the loss of information, as items and clusters are merged, and has been shown to perform very well in evaluations of clustering methods. Squared Euclidean distance for binary data was used to compute measures for input to the cluster analysis. To determine if the clusters corresponded to meaningful types, we calculated the mean values on the total number of each of the five resources for each cluster. These were used to interpret the results and to assign clusters to meaningful types, addressing research questions 1 and 2.

To answer research questions 3 and 4, we performed cross-tabulations of the ICT types with Bos et al.’s types and with the disciplinary types.

5. Results

Research question 1 asked what ICTs the VREs incorporated. Our observations identified a great deal of variation in the ICTs incorporated in VREs. Table 1 summarizes the frequencies of occurrence of the 29 ICTs. In all, there were 1504 distinct ICTs in the sample of VREs. As Table 1 indicates, by far the most frequently occurring ICTs relate to communication and community, comprising 1003 (67.7%) of the ICT instances identified. Computational resources were identified 202 times (13.4%), Publishing 183 times (12.2%), Data resources 89 times (6.0%), and Instruments only 27 times (1.7%). If we combine the counts for resources related directly to communication—Community and Publishing—they comprise 79.9% of the ICTs included in the VREs.

In terms of specific ICTs, most VREs incorporated those related to informing newcomers or the public about the VRE: Information about the VRE (e.g. “About us”), Project information, Research information (about mission and goals of the VRE), and Collaborator lists. Ninety VREs (54.9%) have Global search tools, but otherwise, navigating the VRE is generally left up to the visitor. Access to reports and data generated by the project was also common, with

110 (67%) of the VREs having Literature depositories. The most common community ICT was the newsgroup and about 1/3 of the VREs incorporated links to social media. Scientific instruments were not common, occurring in fewer than 20 VREs out of 164.

Table 1. Features and Total Occurrences

Resource	Total
Instrument	
Scientific instrument	21
Link to external scientific instrument	6
Data	
Database	65
Link to external database	24
Computational	
Data analysis tool	30
Downloadable software	22
Link to external data analysis tool	12
Link to external downloadable software	14
Global search option	90
Specific search option	34
Publishing	
Literature depository	110
Blog	18
Wiki	6
RSS	23
Youtube	26
Community	
General VRE information	152
VRE projects information	148
Research information	124
FAQ	50
Collaborators information	97
Global E-mail	76
Specific E-mail	65
Events Calendar	60
Twitter	55
Facebook	56
Google	11
Forum	9
Newsgroup	76
Language option	24

5.2. ICT-Based Typology of VREs

Research question 2 focused on identification of specific types of VREs based on their ICT configurations. The results of the cluster analysis indicated that both four and five-cluster solutions had good fit. We selected the five cluster solution based on interpretability. Table 2 shows the mean values of the number of ICTs for each of the five major categories of ICTs for each of the clusters.

The first cluster (n=37) had low values on all major categories, with Publishing and Community ICTs most common. The VREs in this cluster were often gateways to other online scientific tools or data and listed articles and reports, but beyond that gave only modest ICT support. They primarily functioned either to show an online presence or to refer visitors to other resources. Hence, we named this cluster the *Webpage* cluster. The LIPID MAPS Lipidomics Gateway [27] offers one example of a *Webpage* VRE.

The VREs in the second cluster (n=22) provided robust levels of Computational, Publishing, and Community ICTs. These VREs generally incorporate data archives or repositories that participants can access and in some cases publish their own data to. These VREs also include computational resources for analysis and annotation of data and the second highest number of Community ICTs among the clusters. This configuration of ICTs has the potential to foster the development of a community of scholars who use computational tools and common data repositories in their own research and are interested in sharing results and papers and in interacting with other members of the user community. We named this cluster the *Analytic Community*. One example of an *Analytic Community* VRE is *Archaeotools* [28].

Two clusters showed similar patterns of ICT resources, and differed primarily in the level of resources incorporated into the VREs. The VREs in both clusters incorporated Instruments, Data and Computational resources and had low levels of Publishing and Community resources. Based on ICT configuration, these VREs focused on “doing” scholarship, as opposed to sharing reports or community interaction and outreach. Both of these clusters were named *Scientific Support* VREs. Where the clusters differed was in the level of Instruments, Data, and Computational resources they provided, as shown in Table 2. One cluster was termed *Moderate Scientific Support* and the second *Intensive Scientific Support* in view of the fact that it had among the highest amounts of Instrument, Data, and Computational resources among the five clusters.

A good example of a *Moderate Scientific Support* VRE is NEESGrid [29]. SEEK [30], developed as part of a European initiative (SysMO) to investigate dynamic molecular processes in unicellular organisms, provides a good example of an *Intensive Scientific Support* VRE. The *Moderate* and *Intensive Scientific Support* types are similar in profiles of resources provided and differ primarily in the number of resources they incorporate. For subsequent analysis these clusters were combined into a single *Scientific Support* VRE, with the acknowledgement that they vary in intensity.

Table 2. Mean Numbers of ICT Resources for VREs in ICT-Based Clusters

Cluster	Name	Resource Category					Total
		Instrument	Data	Computational	Publishing	Community	
1	Webpage	.081 ^{ML}	.216 ^L	.568 ^L	.622 ^L	5.19 ^L	6.68
2	Analytic Community	.045 ^L	.182 ^L	1.00 ^M	1.27 ^{MH}	6.41 ^M	8.91
3	Moderate Scientific Support	.089 ^{ML}	.411 ^{ML}	1.16 ^M	1.00 ^M	5.34 ^L	8.00
4	Intensive Scientific Support	.360 ^H	1.16 ^H	1.88 ^H	.880 ^{ML}	5.68 ^L	9.96
5	Archetypal Collaboratory	.375 ^H	1.04 ^H	1.96 ^H	2.25 ^H	9.54 ^H	15.2
	Total	.165	.543	1.23	1.1159	6.12	9.17

Note: Superscripts indicate relative level of resource in the cluster compared to the grand mean for the resource. Key to Superscripts: L = Low; ML = Moderately Low; M = Moderate; MH = Moderately High; H = High

Table 3. Cross-Classification of ICT-Based VRE Types with Functional VRE Types

VRE Type	Distributed Research Center	Shared Instruments	Community Data Systems	Open Community Contribution Systems	Total
Webpage	14 _a	7 _a	7 _a	9 _a	37
Analytic Community	2 _a	5 _{a, b}	3 _a	12 _b	22
Scientific Support	21 _{a, b}	22 _b	25 _b	13 _a	81
Archetypal Collaboratory	4 _a	7 _a	6 _a	7 _a	24
Total	41	41	41	41	164

Note: Each subscript letter denotes a subset of Primary Function Classification categories whose column proportions do not differ significantly from each other at the .05 level.

Table 4. Mean Numbers of ICT Resources for VREs in Disciplinary Domains

VRE Domain	Resource Category				
	Instrument	Data	Computational	Publishing	Community
Biological Sciences	.11	.65	1.28	1.0370	5.8889
Natural Sciences	.12	.47	1.53	1.1765	7.0588
Physical & Chemical Sciences	.30	.40	1.25	1.1000	5.9500
Social Sciences	.30	.90	1.30	1.2000	6.2000
Humanities	.11	.48	.93	1.2222	6.2963
Multidisciplinary Social Sciences/ Humanities	.38	.54	1.31	1.4615	6.5385
Multidisciplinary Hard Sciences	.09	.39	1.17	.91	5.6000

Table 5. Cross-Classification of ICT-Based VRE Types with Disciplinary Domains

VRE Type	Disciplinary Domain							Total
	BIO	NAT	PHCM	SS	HM	MSS/HM	MHS	
Webpage	12 _{a, b}	1 _b	3 _{a, b}	3 _{a, b}	9 _a	4 _{a, b}	5 _{a, b}	37
Analytic Community	3 _a	2 _{a, b, c}	2 _{a, b, c}	0 _{a, c}	5 _{a, b, c}	3 _{b, c}	7 _b	22
Scientific Support	32 _a	10 _{a, b}	12 _a	4 _{a, b}	10 _{a, b}	3 _b	10 _{a, b}	81
Archetypal Collaboratory	7 _{a, b}	4 _{a, b}	3 _{a, b}	3 _b	3 _{a, b}	3 _{a, b}	1 _a	24
Total	54	17	20	10	27	13	23	164

Note: Each subscript letter denotes a subset of VRE Domain categories whose column proportions do not differ significantly from each other at the .05 level. BIO=Biological/Biomedical; NAT = Natural; PHCM = Physical/Chemical; SS = Social Sciences; HM = Humanities; MSS/HM = Multidisciplinary Social Sciences/Humanities; MHS = Multidisciplinary Hard Sciences

The fifth cluster exhibited high values for all five resources. These were fully developed VREs that offered facilities for doing science collaboratively and attempted to build communities around their emphases. We labeled these *Archetypal Collaboratory* VREs, a title that reflects the vision for collaboratories advanced by Wulf, the Olsons, and others. The mean number of ICT resources provided by these VREs is 15.2, far more than any other cluster. They provide an average of 9.54 community resources, many more than the next highest VRE, the Analytic Community (mean = 6.41). The Cochrane Collaboration [31] is an example of an Archetypal Collaboratory.

5.3. ICT-Based and Functional Typologies of VREs

Research question 3 interrogated the relationship between the ICT-based and functional typologies. Table 3 presents a cross tabulation of VREs by their typologies. There is a significant association between the two typologies (Likelihood Ratio = 18.654, $df = 9$, $p = .028$). Open Community Contribution Systems most often have the ICT configurations of the Analytic Community, which emphasizes shared data repositories, computational resources, publishing, and community support. Distributed Research Centers and, somewhat surprisingly, Community Data Systems tend to avoid the Analytic Community ICT configuration. Instead, Community Data Systems are associated with Scientific Support ICT configurations, which incorporate ICT resources such as instruments, databases, and computational tools. The four functional types are evenly distributed based on expected values across the Webpage and Archetypal Collaboratory types. This suggests that functional VREs are likely to have both low-end and high-end manifestations in terms of ICTs they incorporate.

In terms of the five basic ICT resource groups, there are significant differences across the four functional types in the number of instruments and number of databases in the VREs. As would be expected, Shared Instrument VREs had more Instrument ICTs than the other three VRE types ($F_{3,160} = 6.41$, $p < .001$). Community Data Systems had more databases than the other three VRE types ($F_{3,160} = 9.78$, $p < .001$). There were no significant differences in the number of Computational, Publishing, or Community ICTs for the four functional types.

5.4. Disciplines and VRE Types

Research question 4 focused on the relationship between disciplines and ICT incorporated in VREs.

We considered both distributions of ICT resources across disciplinary domains and the association between the ICT-based VRE types and disciplinary domains.

Mean numbers of each major category of ICT Resources for the disciplinary domains are shown in Table 4. Results show that the Physical and Chemical Sciences, the Multidisciplinary Social Sciences and Humanities, and the Social Sciences had the highest number of Instruments on average. This finding makes sense for the Physical/Chemical Sciences because they often depend on large, expensive shared tools such as colliders or experimental workflow systems. We believe the high average for the Social Sciences and Multidisciplinary Social Science and Humanities sites is an artifact of relatively smaller sample size for these domains. For Data resources, the Biological Sciences, Multidisciplinary Social Sciences and Humanities, and Social Sciences, had the highest mean. Online databases are common in Biology to allow scientists to pool taxonomic, genomic and other types of data across species and individuals within species. Social scientific VREs often house common data stores from survey research and other expensive procedures.

The Natural Sciences had the highest average number of computational resources. The Natural Sciences have a tendency to provide data analytic tools for their entire communities. The US-Virtual Astronomical Observatory, for example, created four data analytic tools for their community and a number of VREs around the world are linked to these developed resources. It is also noteworthy that there is an abundance of computational resources for the Social Sciences which often include statistical software and visualization tools; these are also common in Multidisciplinary Social Science VREs, which facilitate cross-disciplinary inquiries.

Multidisciplinary Social Science VREs had the highest number of publishing resources. Part of the reason for this may be the lack of public open access publication outlets such as arXiv for the humanities and social sciences, leading to the creation of special paper repositories and sites for publishing data. The Natural Science VREs had the highest number of community resources, but otherwise the number of Community ICTs was similar across domains.

Table 5 shows the cross-classification of the four ICT-based VRE types and disciplinary domains. The association was significant at $p = .08$ (Likelihood Ratio = 26.78, $df = 18$).

Humanities have more Webpage VREs than would be expected and the Natural Sciences fewer. There were fewer Analytic Community VREs than expected in the Biological Sciences and more in the Social Sciences and Interdisciplinary Hard Sciences. For

Scientific Support VREs there were significant differences between the Biological and Chemical/Physical Sciences—which had more than expected—and the Humanities, which had fewer. The Social Sciences and Multidisciplinary Hard Sciences differed in proportions of the Archetypal Collaboratory VREs, with Social Sciences having more than expected and the Multidisciplinary Hard Sciences less.

6. Discussion and Conclusion

This study has derived a typology of VREs based on the ICT configurations they employ. A key finding is that ICTs to support communities and to post and transfer information are the most common type of ICT found in VREs, with Databases, Instruments and Computational ICTs being much less common (but also important for more substantive VRE types).

One striking finding is that the Community ICTs in these VREs were almost entirely dedicated to providing one-way transmission of information; even potentially interactive tools like Twitter were mostly used to send out information about the projects. We read a number of blogs embedded in the VREs, and even though these have the potential to be more interactive, postings primarily consist of news and reports about the project, rather than initiating and sustaining collaborations. And as we noted, interactive tools such as chat systems and conferencing systems are almost never incorporated in VREs. It appears that collaborations and collaborative groups for the most part are assembled and work outside the VRE. It may be the case that ICTs for collaboration are used outside the VRE. It may also be the case that collaboration in VREs is only intermittent in any give project. In one of the earliest studies of distributed scientific collaboration, Finholt [32] concluded that virtual collaboration does not need to be ongoing to be useful, and members can alternate between collaboration and individual work. In this respect the VREs are not taking advantage of virtual teams to the same degree that businesses were ten years ago [10]. The exceptions are databases and published datasets, in which scholars and “citizen scientists” work on data.

The four types incorporate suites of ICTs that support different types of research endeavors. Webpage configurations do not give much direct support for research, but serve as “traffic signs” that direct scholars to other research sites and summarize previous and current projects. Analytic Community VREs incorporate a suite of ICTs designed to support upload, annotation, coding, and/or classification of data by participating scholars and the public. Scientific Support VREs focus on provision of instruments,

databases, and computational resources, with less emphasis on the community of users. Archetypal Collaboratory VREs incorporate the full range of ICT resources to enable the type of full-fledged collaboration envisioned by Wulf and other leaders of the “collaboratory movement” [3, 8, 13, 14, 16].

The ICT-based types had some of the relationships with the functional typology of VREs developed by Bos et al. (2007). However, it was evident that there was considerable variation in ICT configurations in each of the functional collaboratory types. This suggests that standardized formats have not yet developed around the functions of VREs. Our qualitative observations reflect this as well: the design of VREs in our sample seems to be *ad hoc* and idiosyncratic VREs have the potential to foster vibrant and active communities that stimulate creative insights, but only if they are designed for effective collaboration. Our study suggests there is much room for improvement in the current state of the art in VRE design. Insights from online community design [12] and collaboration engineering [9], among other frameworks, would greatly enhance the social and user experience in VREs.

One important consideration in design of VREs is the disciplines they serve. Different disciplines have different ICT requirements and disciplinary cultures [22, 23]. Factors that should be considered include; (1) type of data typical of the discipline (e.g., Astronomy has huge datasets of numerical data; Weather Science huge sets of dynamic data; Humanities large corpuses of documents); (2) typical instruments (e.g., shared telescopes for Astronomy; individual interpretation for Humanities; libraries of molecules for Chemistry); (3) typical collaboration patterns (community wide collaboration; small teams; individual scholars); and (4) disciplinary publication norms and standards (open source versus curated report collections versus refereed journals). To be effective a VRE must be designed so it fits the discipline(s) it is designed for. This implies that design of effective VREs is particularly challenging for multidisciplinary scholarship.

A limitation of this study is that it was only feasible to capture data on the presence or absence of ICTs, rather than how well the ICT was done or how well it was used by the participants. Coding of quality of ICT implementation and use requires much more intensive analysis than is feasible with a large sample of VREs. One goal of future work will be to conduct more intensive analysis of a subset of the VRE types identified in this study to explore quality of implementation, factors that led to ICT choice, design philosophies, and other questions that readily come to mind when considering the typology identified in this research.

Another subject for future research is effectiveness of the VRE. A key question is which designs fit the purposes of the VRE most effectively. Key outcomes include total publications, interdisciplinary publications, collaborations initiated, community building, and satisfaction with the VRE.

7. Acknowledgement

This work was supported by the US National Science Foundation's (NSF) Virtual Organization as Sociotechnical System (VOSS) Grant Award # 1308176 and Science Across Virtual Institutions Award # OCI 1308176. The opinions expressed herein are those of the authors and do not reflect the NSF.

8. References

- [1] N. J. Cooke and M. L. Hilton, (Eds.), *Enhancing the effectiveness of team science*, National Academies Press, Washington, DC, 2015.
- [2] K. Carley and K. Wendt, *Electronic Mail and Scientific Communication: A Study of the Soar Extended Research Group*, *Science communication*, 12(4), 1991, pp. 406-440.
- [3] J. H. Olson, and G. Olson, *Working together and apart: Collaboration over the Internet*, Morgan and Claypool, San Rafael, CA, 2013.
- [4] R. T. Kouzes, J. D. Myers, and W.A. Wulf, *Collaboratories: Doing science on the Internet*, *Computer*, 29(8), 1996, pp. 40-46
- [5] D. E. Atkins, K. K. Droegemeier, S. I. Feldman, H. Garcia-Molina, M. L. Klein, D. L. Messerschmitt, P. Messina, J. P. Ostriker, and M. H. Wright, *Revolutionizing science and engineering: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*, National Science Foundation, Washington, DC, 2003.
- [6] National Science Foundation, *Cyberinfrastructure*, Retrieved from NSF-Wide investment: http://www.nsf.gov/news/priority_areas/cyberinfrastructure/index.jsp, 2005.
- [7] Research Councils UK, *Previously funded programmes*, Retrieved from Research Councils UK: <http://www.rcuk.ac.uk/research/xreprogrammes/prevprogs/>, 2014.
- [8] M. Anandarajan and A. Anandarajan (Eds.), *e- Research Collaboration: Theory, Techniques and Challenges*, Springer, Berlin, 2010.
- [9] R. Briggs, G. Kolschoten, V. Gert-Jan, and D. Douglas, *Defining key concepts for collaboration engineering*, *AMCIS 2006 Proceedings*, 2006, paper 17.
- [10] A. Majchrzak, A. Malhotra, J. Stamps, and J. Lipnack, *Can absence make a team grow stronger?*, *Harvard business review*, 82(5), 2004, pp.131-137.
- [11] H.A. Smith and J. D. McKeen, *Enabling collaboration with IT*, *CAIS*, 28, 2011, article 16.
- [12] Preece, J., 2000. *Online communities: Designing usability and supporting socialbility*. John Wiley & Sons, Inc..
- [13] M. A. Fraser, *Virtual Research Environments: Overview and Activity*, *Ariadne*, 44, Retrieved from <http://www.ariadne.ac.uk/issue44/fraser/>, July, 2005.
- [14] T. Hey and A. E. Trefethen, *Cyberinfrastructure for e- Science*, *Science*, 308(5723), 2005, pp. 817-821.
- [15] Australian Government, Department of Education and Training, *National Collaborative Research Infrastructure Strategy (NCRIS)*, Retrieved from Australian Government, Department of Education and Training: <https://education.gov.au/national-collaborative-research-infrastructure-strategy-ncris>, 2015.
- [16] N. Bos, A. Zimmerman, J. Olson, J. Yew, J. Yerkie, E. Dahl, and G. Olson, G., *From shared databases to communities of practice: A taxonomy of collaboratories*, *Journal of Computer-Mediated Communication*, 12(2), 2007, pp. 652-672.
- [17] P. M. Leonardi, *Materiality, sociomateriality, and socio-technical systems: What do these terms mean? How are they different? Do we need them*, In P. M. Leonardi, B. A. Nardi, & J. Kallinikos (Eds.) *Materiality and Organizing: Social Interaction in a Technological World* (pp. 25-48). New York: Oxford University Press. .2012.
- [18] *International Cyberinfrastructure: Activities Around the Globe*, *Cyberinfrastructure Watch Quarterly*, 2(1), <http://www.ctwatch.org/quarterly/archives/february-2006.html>, February, 2006.
- [19] G. M. Olson, A. Zimmerman, and N. Bos, *Scientific collaboration on the Internet*, The MIT Press, 2008.
- [20] N. Babchuk, B. Keith, and G. Peters, *Collaboration in sociology and other scientific disciplines: A comparative trend analysis of scholarship in the social, physical, and mathematical sciences*, *The American Sociologist*, 30(3), 1999, pp. 5-21.
- [21] V. Lariviere, Y. Gingras, and E. Archambault, *Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities*, *Scientometrics*, 68(3), 2006, pp. 519-533.
- [22] J. N. Cummings and S. Kiesler, *Collaborative research across disciplinary and organizational boundaries*, *Social studies of science*, 35(5), 2005, pp. 703-722.
- [23] B. Bozeman, S. Lee, M. Gaughan, and I. Chompalov, *The impact of research collaboration on scientific productivity*, *Social Studies of Science*, 2003.
- [24] T. Becher, *Academic tribes and territories: An intellectual inquiry and the cultures of disciplines*, The Society for Research into Higher Education and Open University Press, Bristol, PA, 1989.
- [25] L. Olsina, D. Godoy, G. Lafuente, and G. Rossi, *Assessing the quality of academic websites: a case study*, *New Review of Hypermedia and Multimedia*, 5(1), 1999, pp. 81-103.
- [26] *Science of Collaboratories*: <http://soc.ics.uci.edu/>
- [27] The LIPID MAPS Lipidomics Gateway: <http://www.lipidmaps.org/>
- [28] *Archaeotools*: <http://archaeologydataservice.ac.uk/research/archaeotools>
- [29] NEESGrid: <http://www.neesgrid.org/>
- [30] SEEK: <http://www.seek4science.org/>
- [31] The Cochrane Collaboration: <http://www.cochrane.org/>
- [32] T. A. Finholt, *Collaboratories*, *Annual Review of Information Science and Technologies*, 36, 2002, pp. 73-107.