

REVIEW

Open Access

eScience and Informatics for international science programs

Peter Fox^{1*} and Janet Kozyra²**Abstract**

Both eScience and Informatics, their conceptual approaches, methods, and extant technical solutions, are still relatively new approaches in many science fields. Collectively, they intend to facilitate the electronic conduct of science (eScience) and the application of foundational principles drawn from experience in a number of fields of Informatics (e.g., helioinformatics, geoinformatics, and climateinformatics). The Climate and Weather of the Sun-Earth System II (CAWSES II) eScience and Informatics effort had the goal of promoting an international virtual institute and several virtual observatories in order to advance system-level science investigations aligned with the four CAWSES II Task Groups. This contribution elaborates on the key elements of eScience and Informatics applicable to CAWSES II as a virtual organization in the context of the current science-data landscape. We examine what was adopted for CAWSES II and highlight the successes and challenges of the effort. Based on the lessons learned from this effort and other international communities, we present opportunities going forward that are relevant to both smaller collaborations and successor large-scale programs (such as the Variability of the Sun and Its Terrestrial Impact (VarSITI)) that are being recognized more as networks of science facilitated by a variety of modern information technologies.

Keywords: eScience; Informatics; Virtual observatory; Virtual organization; Collaboration; Interdisciplinary research

Review**Introduction: eScience, informatics for collaboration, and network science**

As laid out by Hey and Trefethen (2005), much is changing in the way science is being done today. Increasingly, science is multi-disciplinary and the data are often collected and analyzed by scientists from many organizations, often using different technical vocabularies and diverse methods and assumptions. Science is becoming more data-dependent, yet traditional data technologies were not designed for the heterogeneity and scale of the modern world. Thus, eScience, the electronic facilitation of science, is transforming many research fields across a variety of scales. For example, 'big' science projects, such as the 1,000 dollar genome (Mardis 2006), the Large Hadron Collider (LHC), or the Square Kilometer Array Pathfinder (ASKAP), will generate petabytes of data that must be analyzed by hundreds of scientists working in multiple countries and speaking many different natural

and computer languages. Clearly, the needs of such multi-disciplinary data-intensive science (Bell et al. 2009) must go beyond data warehouses and closed systems and instead strive to allow access to, and analysis of, the data to those outside the primary project teams. Such eScience infrastructures (also termed cyberinfrastructure in the U.S.) are developed to meet the current and next-generation scientist's need for scientific services that allow the use and integration of a wide range of sources of data, information, knowledge, and software tools. These services need to provide interfaces to a broad range of users, including those who are expert scientists, but not experts in data administration and computation. In turn, it is the field of Informatics (Informatics) that provides the methodological means for assessing those science requirements that define what the required eScience infrastructure must provide to researchers (individually and in collaborative teams), as well as how they operate and are useful. International science programs, such as the Climate and Weather of the Sun-Earth System (CAWSES), are prime examples of where researchers meet to identify and discuss science research topics in a collaborative manner. The aforementioned eScience infrastructures are the

* Correspondence: PFOX@CS.RPI.EDU

¹Tetherless World Constellation, Rensselaer Polytechnic Institute, 110 8th St., Troy, NY 12180, USA

Full list of author information is available at the end of the article

intended means by which data-oriented science is facilitated by the data and the people. In addition, these science programs quickly encounter even greater data-related challenges: questions such as ‘How do I use this data that I did not create?’ ‘How do I use this data type that I have never seen before with the data I use every day?’ ‘I really need data from this other discipline, but I cannot understand their terms, what do I do?’ The answers to the majority of these types of questions arise via collaborations in networks of science (Rhoten 2007) and with appropriate attention to how data are documented and made available. Informatics approaches, such as use cases, information models, and rapid prototyping (e.g., Benedict et al. 2007), are used to fully understand and document the detail of challenging research data needs.

Many eScience environments today include access to online repositories of science data. Convenient electronic access to data provides one significant step forward in allowing scientific research. However, access is one critical component to next-generation science. Another critical capability that is required (by humans or computer programs) to manage diverse data is to know what the data ‘means’. As the volume, complexity, and heterogeneity of data resources grow, scientists increasingly require new capabilities that can only be achieved if new tools and approaches that understand meaning can be deployed. Overall, the idea of virtual observatories (VOs) aims at providing virtual interconnections among heterogeneous, distributed, and international data repositories and to serve research communities. Originating in astronomy, VOs were quickly adopted by the solar and space physics communities. In order to address community goals, such as providing support for both specialist and broader non-specialist use, including lay people, of a rich set of science and information products, different approaches to VOs were devised. One such approach is that of discipline-specific virtual observatories (known as VxOs; Dalton 2007). When viewed within one discipline (‘x’), VxOs became popular and successful. For example, the National Aeronautical and Space Administration (NASA) Virtual Observatories for Heliophysical Data (VOHD) program currently funds several VxO efforts (e.g., the Virtual Magnetospheric Observatory (VMO); King et al. 2007).

As these efforts move either beyond their discipline or to non-specialist use, vocabulary challenges arise. Often, vocabularies differ, some are quite esoteric and jargon laden. Sometimes, similar terms have different meanings, and often, there are multiple terms with different meanings, and multiple terms for the same phenomenon or process. These challenges present barriers to efforts that hope to use existing cyberinfrastructure in support of interdisciplinary data query and access, especially when the interdisciplinary applications must go beyond search

and access to actual manipulation and use of the data. During CAWSES (I and II), there have been successful efforts to develop richer metadata databases to address the interdisciplinary vocabulary challenges. For example, significant progress has been achieved in both the Japanese project for the Inter-university Upper atmosphere Global Observation NETWORK (IUGONET) and its counterpart, the European Near-Earth Space Data Infrastructure for e-Sciences (ESPAS) project. Notably, there is now common use of the Space Physics Archive Search and Extract (SPASE; King et al. 2010) metadata schema by VxOs (heliospheric, energetic particle, radiation belt, etc.) and IUGONET, which is expected to further contribute to the needed interoperability.

Since the mid-2000s, semantic technologies have been gaining momentum in a number of eScience areas and are being used to provide smarter query, integration, and ingest services - for example, solar-terrestrial physics (Virtual Solar-Terrestrial Observatory (VSTO), McGuinness et al. 2007; Fox et al. 2009), ecology (Science Environment for Ecological Knowledge (SEEK), Madin et al. 2007), ocean sciences (Marine Metadata Initiative (MMI)), and healthcare and life sciences (HCLS; Neumann 2005), to name but a few. It is becoming increasingly important to offer semantic-based methodologies, tools, and middleware to the developers of eScience infrastructures. In turn, these developers can facilitate scientific knowledge modeling, logical-based hypothesis checking, semantic data integration, application composition, and integrated knowledge discovery and data analysis for different scientific domains - systems that can be used by scientists, students, and, ever increasingly, non-experts.

Each stakeholder in a collaborative science network is also a repository of knowledge for her or his domain (not simply data and related documentation but also related facts via notebooks, blogs, tweets, and social network postings, in addition to more traditional exchanges via email). However, this knowledge is often incoherent and difficult to find, let alone be utilized by others. Furthermore, this knowledge is often ‘grey’ (tacit) and not accessible in a way where questions of interest can be formulated, posed, answered, assessed, and trusted. While still not a prevalent terminology, *knowledge networks* (see Diviacco 2014) provide representations of a view into such a knowledge base, with the goal of gaining insight and understanding into the various attributes of a real network, especially their collaborative nature, e.g., ‘who is doing what’. Unfortunately, promoting human presence (versus their ‘footprints’) - the central problem for successful distributed research interactions - is still a significant challenge to overcome in network communities, and one that was not addressed in CAWSES, but must be prominent in future programs.

While the role(s) and value of individuals in collaborative environments (and organizations in general) has been known for some time (McGregor and Cutcher-Gershenfeld 2005), it is relatively recently that a diverse range of individuals have become major contributors to information (and even knowledge and data) in complex research networks. Traditionally, the principal investigator (PI) and collaborators were primarily associated with the generation of knowledge artifacts.

To conclude this introduction, we turn briefly to some key underpinnings of collaboration communities (commonly known as laboratories, computer-supported cooperative work (CSCW), network science (NS), and Science of Team Science (SciTS)). These are vast fields of study, and the reader is referred to monographs on these subjects (e.g., Olson et al. 2008; Grudin 1994) for more detailed background and/or state of the art. Instead, we highlight relevant aspects of collaboration facilitated via electronic and communication and computer network means. For example, in the U.S. (to reflect the author experience; similar initiatives occurred in Europe, Asia, and Australasia), recognition for the significance of a new mode of science arose within the National Science Foundation (see Rhoten 2007), the National Institutes of Health, and the Department of Energy Office of Science. Much of the activity from U.S. funding agencies was fuelled by U.S. National Academy/National Research Council (NRC) studies that began to appear in the mid-1990s (Zare 1997). Prior to that, the entire field of CSCW, notably using the softer word *cooperation* rather than *collaboration*, had been defined a decade earlier and had become an established field of scholarly pursuit. Grudin (1994) provides perspective into that decade of evolution of CSCW and emphasizes the role of computer software, especially 'groupware', as an enabling capability that was sought. Even then, Grudin noted, CSCW 'has been criticized because "cooperative" work is often more a goal than a reality.' The CSCW community contributed importantly to the several diagrams that represent the different modes of collaboration that depend on location and time - see Figure 1 for the 3 × 3 matrix of DeSanctis and Gallupe (1987). More recently, the term virtual organization (unfortunately contracted as VO, almost the same as virtual observatories VxO) has been used. In short, a virtual organization is assembled around a common goal or task whose activities are facilitated by electronic means. As a result, specific computer applications are identified or developed to facilitate both the activities and the attainment of their goals. The overall conceptualization of virtual organization was being refined and advanced in a key workshop held during the CAWSES program period. The report (Cummings et al. 2008a) from Building Effective Virtual Organizations workshop is entitled Beyond Being There (Cummings et al. 2008b) and

informed many of the envisioned network activities in CAWSES II.

eScience and Informatics approaches have changed many fields of science, including those represented by the CAWSES and Scientific Committee of the Solar Terrestrial Energy Program (SCOSTEP) community in general. In particular, it is the data sharing and collaborative aspect that provides significant potential for international science programs to succeed. Thus, a collaborative, data-enabled eScience infrastructure must provide:

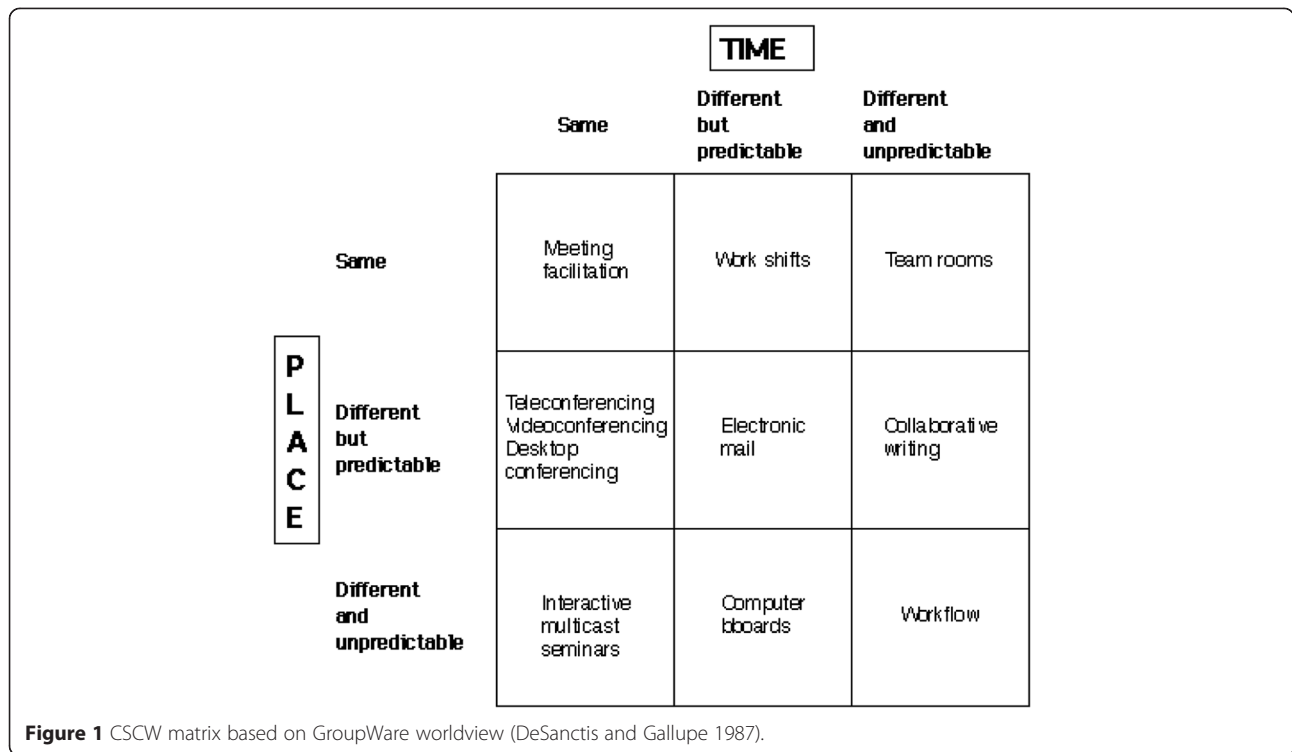
- A virtual observatory that brings data together
- A virtual organization that brings people together
- One or more collaboration modes/environments that should unite virtual observatories and virtual organizations.

In addition, Informatics provides the methodological means to define what these elements are and how they interrelate.

CAWSES I eScience activities

Before reviewing activities for CAWSES II, we re-visit the nature of eScience activities for CAWSES I. The primary focus was on experimentation with a new form of scientific communication, the virtual conference, as a tool for enhancing scientific collaboration worldwide: (1) to address the grand challenge science issues that require the combined expertise of multiple disciplines and the synthesis of data sets worldwide for progress, (2) to support multi-disciplinary and interdisciplinary collaborations, (3) to promote science capacity building in developing countries, (4) to provide a resource for students worldwide, and (5) to celebrate the 50th anniversary of the 1957 International Geophysical Year (IGY). Fifty years after IGY, the complexity of the Sun-Earth system emerged as a primary focus of the CAWSES program with new discoveries increasingly found at the intersection and boundaries between regions. Continuing progress required access to a vast developing cyberinfrastructure of large international data sets, high-performance computing, and advanced visualization, and the development of new types of interdisciplinary and international research interactions (the human element). The integration of technology with human modes of interaction, and in fact, the injection of the 'human presence,' is one of the most difficult challenges for virtual interactions, and it can be the key to their success or failure. The more those exchanges that occur within the virtual conference resemble those in traditional research collaborations, the more likely that a valuable exchange of information and ideas will result.

The CAWSES program offered a unique environment within which to perform this experimentation. CAWSES contributed scientific oversight by committees of



international scientific leaders, a loosely connected multi-disciplinary community formed in association with CAWSES largely funded by national programs, and a set of new and developing CAWSES global data products that required international collaboration to produce. This environment was further enriched by a number of sponsors and scientific advisors, including the International Heliophysical Year (IHY), Electronic Geophysical Year (eGY), Interhemispheric Conjugacy Effects in Solar Terrestrial and Aeronomy Research (ICESTAR), NASA/Living with a Star (NASA/LWS), National Science Foundation (NSF) Atmospheric Sciences, and developers of the Solar Extreme Events workshop series.

To support this effort, a virtual conference environment was developed through collaboration between the CAWSES program and Johns Hopkins University Applied Physics Laboratory. The focus of the development was an interface to support the human interaction required for collaborative research in a worldwide setting that spanned nations and discipline areas. Virtual conference is a very different medium from face-to-face (F2F) conference, with strengths that appear particularly well suited to support interdisciplinary worldwide interactions: it is non-exclusionary. The properties of virtual conference allow maximum participation across disciplines and nations, easy sharing of information, more time to pose thoughtful questions and supply well-considered answers, ability to ‘attend’ presentations in all discipline areas, and

new opportunities to explore interesting collaborations across disciplines and nations (Figure 2). Part of the effort focused on tutorials relevant to the topic of the conference aimed at communicating frontier research areas in one discipline to the others.

Virtual conference was designed with careful attention to the requirements of the participants based on lessons learned in other communities (c.f., Green 1998a, 1998b; Shimabukuro 2000; Wieman 2004). The goal was to construct a stimulating and compelling online experience in which conversations were friendly and peer to peer, and ideas could be freely exchanged. Particular attention was devoted to removing access barriers with a design that had elements resembling an F2F conference and a visible and active help desk for technical issues. The plan was to support a series of virtual conferences with each addressing a single Sun-to-Earth science topic, sufficiently broad to engage all disciplines, but also sufficiently focused to ensure those same disciplines intersected, rather than operated independently.

The conference software was adapted from previously developed software that is Common Gateway Interface (CGI) script-driven, stores the database in the eXtensible Markup Language (XML), is fully customizable, has style that is independent of content, and is surprisingly versatile. The presentations were text- and graphics-based and asynchronous in order to be accessible to a worldwide audience regardless of time zone. The communications

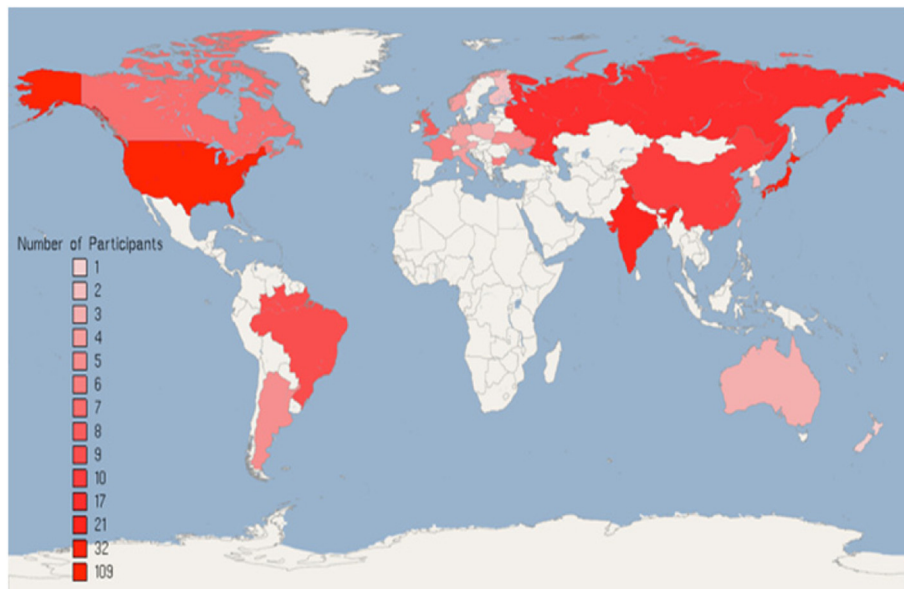


Figure 2 Remarkable worldwide participation in CAWSES I virtual conference is shown in this map (Barnes et al. 2006). More than 270 researchers from 20 countries representing all Sun-Geospace disciplines participated.

were designed to be relatively low bandwidth in order to support the range of worldwide capabilities. A File Transfer Protocol (FTP) site was employed to upload presentations and a website to host them. A data commons area contained conference data products, shared resources, and links to cyberinfrastructure, including virtual observatories, runs-on-demand of community models, and advanced visualization. Discussions occurred on message boards attached to individual papers, sessions, and conference-level overview topics.

A number of elements were adopted to create a 'human presence' in the conference environment. A friendly welcome and logistics email that summarized conference developments was sent daily to participants. Another key element was the use of moderators to encourage, clarify, and integrate the discussions that occurred on message boards but, equally important, to 'weave together' the input from the conference participants into a global perspective each day. The contributions of the moderators were critical for allowing participants to exit the conference, return later, and be updated on progress and new developments. They provided a starting point for the next round of discussions, a documentation of accomplishments, and the basis for deciding future directions. These system-level insights were the primary output of the first conference. The conference was archived for continuing collaboration and future reference. The entire effort rested on voluntary contributions from developers, programs, organizers, moderators, and participants. Without international programs such as CAWSES,

this impressive orchestration of scientific collaboration and focused resources could not have occurred.

Though only one virtual conference was accomplished because of funding and time constraints rather than as the series envisioned, the response of the international community provided an intriguing glimpse into the potential for this type of virtual interaction to introduce new ways of doing research. Participation in the virtual conference far exceeded expectations. The 'live' portion of the conference lasted eight workdays. This was followed by a 4-day ramp-down for final comments and updates, which was added in response to requests by conference participants. The conference structure was surprisingly versatile. Many suggestions by participants for improved features were incorporated in real time. Presentations developed in response to questions raised were uploaded and integrated into the conference. By the final day, the virtual conference had logged 272 registered participants from 20 countries and more than 120,000 total hits. The key ingredient, by far, was the enthusiastic collaboration of participants across missions, data sets, disciplines, and nations to produce a fusion of knowledge and experience focused on understanding the state of the Sun-Earth system during extreme space weather.

One particularly promising suggestion for future interactions is the combined use of small F2F workshops that interact with the virtual conference in order to provide an in-depth exploration of key topics within the broader context provided by virtual interaction. A natural

extension of this effort in CAWSES II was the attempt to develop a virtual institute with virtual conferencing as one element.

CAWSES II virtual institute and related collaboration

The science of CAWSES II was formed around four Task Groups: (1) - What are the solar influences on the Earth’s climate? (2) - How will geospace respond to an altered climate? (3) - How does short-term solar variability affect the geospace environment? and (4) - What is the geospace response to variable inputs from the lower atmosphere? Based on the success demonstrated in CAWSES of more concentrated virtual activities, a

virtual institute was conceived for CAWSES II. The schematic in Figure 3 displays the many facets and functions of the institute. As noted in (http://cawses.org/wiki/index.php/Task_Groups):

This institute, which will be organized around the principles that progress in system-level investigations, requires: researchers committed to the value of pursuing science at the interface between disciplines, strongly-focused science topics that provide a common theme around which disciplines are able to interact, means of educating researchers about the key scientific issues in other disciplines and the connections between

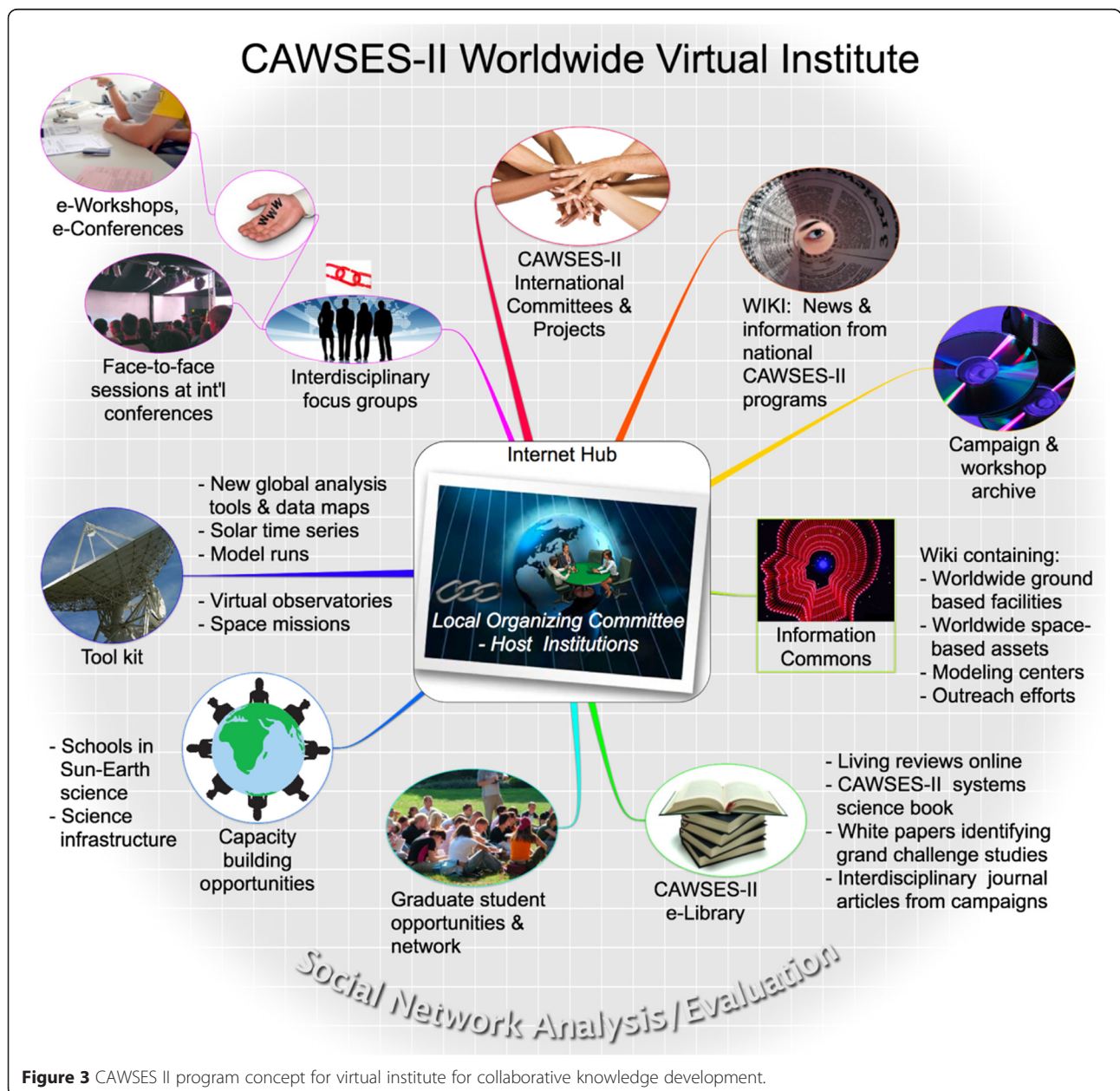


Figure 3 CAWSES II program concept for virtual institute for collaborative knowledge development.

disciplines, access to scientific publications in other discipline areas, new forms of scholarly publishing, and the structure needed to bring researchers into contact with data sets, models, and each other, across disciplines and national boundaries.

CAWSES-II focuses on the fundamental processes of the Sun-Earth system during the rising phase of solar cycle 24. These processes interact in nonlinear ways to produce effects that impact life and society. To address these topics, CAWSES-II uses a research strategy that includes comparisons with other stellar and planetary environments to inform investigations into solar-terrestrial science.

The primary surrounding constraint for the program was that all participation and effort should be voluntary, including the leveraging of technical infrastructure and data and information resources. Several unsuccessful efforts were made to fund a cohesive approach to technical resources for collaboration and dissemination for CAWSES II. As a result, the virtual institute was only partially embraced and implemented between 2009 and 2013. The intent was to hold several virtual conferences around Task Group themes, inter-Task Group interests, or specific geospace phenomena (e.g., extreme events). With these constraints in mind, a simpler and inexpensive approach was adopted for CAWSES II collaboration: a primary informational Web site (www.cawses.org), electronic mailing lists, and a Mediawiki (www.cawses.org/wiki - see <http://en.wikipedia.org/wiki/MediaWiki>) configured around the CAWSES task and working activities. A Mediawiki covers many of the functions in the right column and bottom row of the matrix shown in Figure 1.

These functions are noted to be 'different, but unpredictable' in the sense of familiarity and collaborative/co-operative effectiveness. A further consequence of the voluntary nature of CAWSES II also meant that no formal evaluation of CAWSES II as a virtual organization, effectiveness of collaboration, use of tools, and virtual observatories was conducted. The following narrative is based on the authors' direct involvement in CAWSES II activities and observations of interactions and collaborations among participants (science and leadership).

The informational Web site, a necessity in the current information age, served its purpose in the sense that its intent was to contain very basic, and mostly static, content that characterized the CAWSES II program, science objectives, and language translations (French, German, and Russian) of the CAWSES II program background and description, and numerous links to the CAWSES II wiki where the working activities would be captured.

The majority of the email communication for science activities was performed on Task Group maintained lists, and similarly, the organizational information exchanges (among Task Group chairs/co-chairs, and other program leaders) were largely informal and not immediately visible to the broader CAWSES II virtual organization. The consequences of this lack of visibility are difficult to assess, but it is reported that virtual organizations do feature significant amounts of informal communication, and because of less formal reporting agreements, more communication is required to establish the state of the organization, or at the very least, the state of progress and/or activities/outcomes (for example, in task areas for CAWSES II) and the different types of data resources developed or used in the research tasks.

It is in the latter sense that the wiki capability for CAWSES II provided registered participants (i.e., wiki accounts) the ability to contribute and collaboratively add material, such as documentation and links, to key data resources on the CAWSES II wiki pages. Of the four Task Groups, groups 4 and 3 made substantially more use of the wiki environment than the other two. These wiki pages could be read by anyone and thus provided a sometimes very detailed view into the activities and outcomes of the Task Groups, thereby providing leaders of CAWSES II the opportunity to assess progress toward group goals, as well as entice additional participants. This one view of CAWSES II indicates the bias to which a viewer of the wiki pages is subject. The other Task Groups performed activities and attained important and valuable outcomes, but the record was not immediately visible to those not directly involved. Though the wiki provided a much clearer indication of the virtual organization side of CAWSES than it did for the data component (i.e., via virtual observatories), at an organization level (CAWSES as part of SCOSTEP), the International Council for Science-World Data System (ICSU-WDS) was re-established in 2008 as reformation of the precursor data bodies of the World Data Centre (WDC) and Federation of Astrophysical and Geophysical Services (FAGS) to which scientists in the SCOSTEP fields have contributed. In a SCOSTEP bureau meeting in 2011 (SCOSTEP 2011), the importance of relationship renewal between SCOSTEP and WDS was noted and, in particular, how to be an advisor to ICSU and ICSU-WDS.

Though not the primary topic of this paper, it is very important to note that CAWSES II had two in-person meetings (one at the SCOSTEP 12th Assembly in Berlin in 2010 and the closing CAWSES II symposium in Nagoya in 2013). This same-time, same-place (upper left of Figure 1) opportunity for CAWSES II provided the participants with an opportunity to be fully informed on the achievements of the program. It was also at these meetings that more explicit discussion of the use of

virtual observatories arose, i.e., the data sources accessed and where CAWSES II-inspired generated data were being stored (and not).

To briefly expand Task Group 4's use of the wiki as an example of facilitated on-line collaboration, its overall group table of contents page (http://cawses.org/wiki/index.php/Task_4) highlights the diverse information assembled on the wiki pages, including the organizational structure, introduction to the challenge and scientific issues (topical areas such as gravity waves, tides, planetary waves, and thermospheric disturbances generated by auroral processes), how the group intended to meet the challenges, outcomes and benefits (including success criteria), existing work and plans, key linkages, and milestones. With regard to documenting activities, the second major section entitled 'News, progress reports and business meetings' captures the details of the workshops, meetings (business and science), newsletters, and progress reports.

The 'different, but unpredictable' characterization of online applications, such as wikis (a well-studied form of technology not elaborated upon here; for details, see Cress and Kimmerle 2008), and who in a community adopts them or can adapt previous modes of collaboration to use them (at all, let alone effectively). It is the author's *opinion* that wikis are both established technology and sufficiently 'simple' that their use could be more fully exploited, perhaps with training or video-based help modules targeted to a specific program or community.

Conclusions

Because an evaluation (even an informal one) study of CAWSES II is unlikely, the following is intended to stimulate ideas for future programs that similar communities may undertake. As noted earlier, the suggestions are far from complete or comprehensive, given that they are based only partially on the CAWSES II program, and more on the accumulated experience of the authors from similar global research programs.

In terms of virtual organizations, a role that was identified in CAWSES I, but not utilized in CAWSES II, was that of a 'science coordinator' for each group or even sub-group (there was a more hierarchical structure of groups in CAWSES I). Such coordinator performed a secretary general-type role that is common in scientific societies. The coordinators worked with the group leaders, program leaders, and participants to track, plan, and document group activities. Perhaps, an equally important point is that they also coordinated activities between groups, i.e., disciplinary exchanges for science and also for data and information resources. We speculate that such coordinators, with a small amount of awareness/education with regard to the attributes of successful virtual organizations and technical acuity, would allow

far greater opportunities for collaboration, tracking progress toward goals, dissemination of outcomes, and engagement of new participants. They would also serve to develop or utilize connections with virtual observatories that can help with data access and related services (i.e., lower the barriers to participation), and recommend and recognize key data repositories and archives.

Another key area for international science programs is the fully emergent recognition that the data used and generated in such programs is itself a very (and often the most) valuable outcome. Such program must be encouraged to adopt more formal positions in regard to data, i.e., data policies (for example, see http://classic.ipy.org/Subcommittees/final_ipy_data_policy.pdf - an exemplar in the polar sciences community). Among the desirable 'data' tasks are to deposit (publish) data available in a data center in a way that others can cite it; whenever possible, obtain the data required from a virtual observatory and credit the VxO and original provider; provide sufficient documentation and relevant provenance and quality information to maximize the appropriate use of opportunities; share tools, results, and intercomparisons, both positive and negative, in a publically accessible way (e.g., on a Wiki); and participate in eScience collaborations so that individual presence, communication, and contributions are known and valued.

More advanced opportunities to bridge data into the scientific process include the emergence of data journals (e.g., the *Geoscience Data Journal*; [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)2049-6060](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2049-6060)), the ability to identify (and cite) data (e.g., <http://datacite.org>) and people (e.g., <http://orcid.org>), and numerous online platforms for data science (analysis, analytics, and visualization) that are free or very low cost. Stepping into the world of social media, the opportunities for science exchanges and dissemination via mechanisms such as Twitter, Figshare, and Research Gate, to name a few, are changing many fields of science and altering the means and modes of collaboration.

In summary, the CAWSES II scientific program concluded with many scientific advances across and among the four Task Group areas. The degree of effectiveness of eScience and Informatics efforts varied widely across the CAWSES II activities, covering mostly conventional means of collaboration, and certainly utilized the 'predictable' modes previously discussed (email, in person, and some teleconferencing). However, there was a sub-community within CAWSES II (we can speculate, but not confirm, that this community featured both a younger generation and innovators across generations) that embraced newer forms of collaboration, such as the Mediawiki, in what appeared to be a very effective way. We encourage future international science programs to consider the benefits and resource costs of enhancing

the virtual participation opportunities offered by the current generation of computing technologies. In the end, the CAWSES II virtual institute was an idea 'before its time' but should be retained for consideration in future programs if suitably funded and staffed.

Abbreviations

CAWSES: Climate and Weather of the Sun-Earth System; CGI: Common Gateway Interface; CSCW: computer-supported cooperative work; eGY: Electronic Geophysical Year; F2F: face-to-face; FAGS: Federation of Astrophysical and Geophysical Services; FTP: File Transfer Protocol; IHY: International Heliophysical Year; ICESAR: Interhemispheric Conjugacy Effects in Solar Terrestrial and Aeronomy Research; ICSU: International Council for Science; MMI: Marine Metadata Initiative; NASA/LWS: National Aeronautical and Space Administration/Living with a Star; NSF: National Science Foundation; NS: network science; SciTS: Science of Team Science; SCOSTEP: Scientific Committee of the Solar Terrestrial Energy Program; SEEK: Science Environment for Ecological Knowledge; VarSIT: Variability of the Sun and Its Terrestrial Impact; VO: virtual organization; VOHD: Virtual Observatories for Heliophysical Data; VxO: virtual observatory; VSTO: Virtual Solar-Terrestrial Observatory; WDC: World Data Centre; WDS: World Data System; XML: eXtensible Markup Language.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PF proposed the topic. PF and JK outlined the article. PF researched the virtual observatory and collaboration literature. JK reported on the CAWSES II eScience activities. PF and JK analyzed the effectiveness of virtual collaboration in CAWSES II and provided the interpretations. PF and JK collaborated in the construction of the manuscript. Both authors responded to reviewer comments and then read and approved the final manuscript.

Authors' information

The authors co-chaired the eScience and Informatics effort for CAWSES II.

Acknowledgements

The authors acknowledge the efforts of the CAWSES II program leadership, especially the Task Group chairs in relating experiences with the program. The authors also appreciated the comments from the reviewers to improve the manuscript. Participation of one of the co-authors (Kozyra) was funded in part by NASA grant NNX11AO60G and NSF grant AGS-1414517.

Author details

¹Tetherless World Constellation, Rensselaer Polytechnic Institute, 110 8th St., Troy, NY 12180, USA. ²AOSS Department, University of Michigan, 2455 Hayward St., Ann Arbor, MI 48109-2143, USA.

Received: 25 July 2014 Accepted: 14 April 2015

Published online: 19 May 2015

References

- Barnes RJ, Kozyra J, Weiss M, Fox NJ, Morrison D, Paxton L (2006) Implementing a virtual workshop for interdisciplinary collaboration on grand challenge issues: lessons learned. *Eos Trans. AGU* 87(52), Fall Meet. Suppl., Abstract SA43A-02, A2
- Bell G, Hey T, Szalay A (2009) Beyond the data deluge. *Science* 323(5919):1297–1298
- Benedict JL, McGuinness DL, Fox P (2007) A semantic web-based methodology for building conceptual models of scientific information. *EOS Trans. AGU*, 88(52), Fall Meeting Suppl., Abstract IN53A-0950, A950
- Cress U, Kimmmerle J (2008) A systemic and cognitive view on collaborative knowledge building with wikis. *Computer-Supported Collaborative Learning* 3:105–122, 10.1007/s11412-007-9035-z
- Cummings J, Finholt T, Foster I, Kesselman C, Lawrence KA (2008a) Beyond Being There. http://web.cui.chicago.edu/events/VirtOrg2008/VO_report.pdf, Accessed online 7 April 2014
- Cummings J, Finholt T, Foster I, Kesselman C, Lawrence KA (2008b) Beyond being there: a blueprint for advancing the design, development, and evaluation of virtual organizations. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.205.6126&rep=rep1&type=pdf>. Accessed 20 Jan 2015
- Dalton R (2007) Geophysicists combine forces. *Nature* 447(7148):1037
- DeSanctis G, Galleppo RB (1987) A foundation for the study of group decision support systems. *Management Science* 33(5):589–609, doi:10.1287/mnsc.33.5.589
- Diviacco P (2014) Reconciling knowledge and collaborative e-research. In: Diviacco P, Fox P, Leadbetter, Pshenichny C (eds) Collaborative knowledge in scientific research networks. Hershey, PA, USA: IGI Press
- Fox P, McGuinness DL, Cinquini L, West P, Garcia J, Benedict J (2009) Ontology-supported scientific data frameworks: the Virtual Solar-Terrestrial Observatory experience. *Comput Geosci* 35(4):724–738
- Green L (1998a) Online conferencing: lessons learned, human resources development Canada, Hull (Quebec). Office of Learning Technologies (OLT). <http://files.eric.ed.gov/fulltext/ED433015.pdf>. Accessed 21 July 2014
- Green L (1998b) Playing croquet with flamingos: a guide to moderating online conferences. Office of Learning Technologies (OLT), Human Resources and Social Development Canada. http://www.hrsdc.gc.ca/eng/hip/ild/olt/Skills_Development/Research_reports.shtml. Accessed 21 July 2014
- Grudin J (1994) CSCW: history and focus. *IEEE Computer* vol 27, 5, pp 19–26
- Hey T, Trefethen A (2005) Cyberinfrastructure for e-Science. *Science* 308(5723):817–821, doi:10.1126/science.1110410
- King T, Walker R, Merka J, Joy S, Narock T (2007) The architecture of a multi-tiered virtual observatory. *Earth Science Informatics* 1(1):21–28, doi:10.1007/s12145-008-0006-3
- King T, Thieman J, Roberts DA (2010) SPASE 2.0: a standard data model for space physics. *Earth Science Informatics* 3(1–2):67–73, doi:10.1007/s12145-010-0053-4
- Madin J, Bowers S, Schildhauer M, Krivov S, Pennington D, Villa F (2007) An ontology for describing and synthesizing ecological observation data. *Ecological Informatics* 2(3):279–296
- McGregor D, Cutcher-Gershenfeld J (2005) The human side of enterprise. New York, NY, USA: McGraw-Hill (first published in 1960)
- McGuinness DL, Fox P, Cinquini L, West P, Garcia J, Benedict JL, Middleton D (2007) The virtual solar-terrestrial observatory: a deployed semantic web application case study for scientific research. In: Proceedings of the Nineteenth Conference on Innovative Applications of Artificial Intelligence (IAAI-07). Vancouver, British Columbia, Canada, pp 22–26
- Neumann E (2005) A life science semantic web: are we there yet? *Sci. STKE* 2005, p pe22
- Olson GM, Zimmerman A, Bos N (2008) Scientific collaboration on the Internet. Cambridge, MA, USA: The MIT Press
- Rhoten D (2007) The dawn of networked science. *The Chronicle Review*. 54:2, B12. <http://chronicle.com/weekly/v54/i02/02b01201.htm>. Accessed 19 Jan 2014
- SCOSTEP 2011, Bureau Report. http://scostep.apps01.yorku.ca/wp-content/uploads/2010/10/BM_Minutes-Oct9-10-2011_Actions.pdf. Accessed 15 Jan 2015
- Shimabukuro J (2000) The evolving virtual conference: implications for professional networking. *The Technology Source Archives*, September/October 2000. http://technologysource.org/article/evolving_virtual_conference/, Accessed online 15 July 2014
- Wieman A (2004) Organizing virtual conferences: lessons and guidelines. *Intern. Inst. for Comm. and Dev.* <http://www.iicd.org/articles/IICDnews.import1852>, Accessed online 15 July 2014
- Zare RN (1997) Knowledge and distributed intelligence. *Science* 275(5303):1047, doi:10.1126/science.275.5303.1047
- Mardis ER (2006) Anticipating the 1,000 dollar genome. *Genome Biol.* 7(7), 112. <http://www.ncbi.nlm.nih.gov/pubmed/17224040>
- LHC, The Large Hadron Collider at CERN. <http://lhc.web.cern.ch/lhc/>, Accessed online 17 July 2014
- ASKAP, The Australian Square Kilometer Array Pathfinder. <http://www.atnf.csiro.au/projects/askap/>, Accessed online 2 June 2014
- Informatics. <http://en.wikipedia.org/wiki/Informatics>, Accessed online 17 July 2014
- MMI, Marine Metadata Interoperability Project. <http://www.marinemetadata.org/>, Accessed online 1 July 2014
- HCLS, Semantic Web Health Care and Life Sciences Interest Group. <http://www.w3.org/2001/sw/hcls/>, Accessed online 16 July 2014