

# Crowdsourcing Rare Events: Using Curiosity to Draw Participants into Science and Early Warning Systems

**Andrea H. Tapia**

Penn State University  
atapia@ist.psu.edu

**Nicolas LaLone**

Penn State University  
nick.lalone@gmail.com

**Elizabeth MacDonald**

NASA, GSFC  
eliz.macdonald@gmail.com

**Reid Priedhorsky**

LANL  
reidpr@lanl.gov

**Michelle Hall**

Science Education Solutions  
hall@scieds.com

## ABSTRACT

This research presents a centralized boundary object website and mobile app focused on allowing participants to participate in developing an early warning system through space weather and the beauty of the aurora borealis. Because of the beauty and majesty of auroral activity, people will seek information about when and where these unpredictable events occur. This activity, commonly referred to as nowcasting, can be combined with scientific data collected from observatories and satellites and serve as an early warning system with potentially far greater accuracy and timeliness than the current state of the art. We believe that long-term engagement with a citizen science tool will help bridge the many social worlds surrounding the aurora borealis and lead to the development of an early warning system that may correlate the visibility of the northern lights to violent space weather. We hope this will lead to other real time crowdsourced early warning systems in the future.

## Keywords

Citizen science, Observatories, Aurora, Space weather, Twitter, Early Warning systems.

## INTRODUCTION

Success in creating a functional early warning system based on crowd participation requires a well-trained, motivated, suitably equipped, continually aware, and continually reporting set of observers who will report the first signs of distress. Similarly, success in citizen science projects requires recruiting, training, and retaining interested and motivated participants. In this paper, we argue that curiosity can usefully motivate participation in both early warning systems and citizen science projects, and to demonstrate this notion, we use the aurora borealis (i.e., northern lights) — a curious and beautiful natural phenomenon with demonstrated broad interest.

For example, a recent weather blog's Facebook page generated over 20,000 likes in one weekend by predicting widespread aurora (MacDonald, 2012). Another example is the auroral event of October 24th, 2011, when aurora was visible as far south as Alabama. Due to the excellent timing and widespread visibility in clear skies over heavily populated areas, this storm was covered by more than 500 traditional and online news outlets from large international newspapers to blogs. In real time, tweets exceeded one per second during the event. Additionally, thousands of Twitter messages (*tweets*) containing text and images documented the visible aurora in real time (MacDonald, 2012). The aurora connects normally intangible plasma physics with a source of wide-ranging public curiosity, making this phenomenon a useful test bed for investigating crowdsourcing approaches based upon that emotion.

We find that aurora have many disparate social worlds, communities of interested parties, that are interested in auroral activity and that unifying them has significant value. We have initiated the Aurorasaurus project to investigate these questions. We argue that the Aurorasaurus website<sup>1</sup> and mobile application, currently in prototype and design stages, respectively, can serve as a *boundary object* between multiple social worlds by anchoring the creation of a new community. The core concept of boundary object theory is that there are artifacts that span the boundaries of communities — termed “social worlds” — and through which communication

<sup>1</sup> <http://aurorasaurus.org>

and collaboration can be achieved (Mark, et al 2007). Artifacts in this case are the socio-technical systems including the webpage, the network, the mobile devices, and the software and servers, which support them. Social worlds in this case are groups of people or organizations with like interests, such as space scientists, government, satellite operators, natural beauty seekers, astrophotographers, and educators. We believe that long-term engagement with a citizen science tool will help bridge the many social worlds surrounding the aurora borealis.

The novelty of our approach is to take an activity that average citizens seek to engage in as entertainment or leisure and use traces of this engagement as input to an early warning system. We find that citizens are curious about auroral activity, and this motivates them to seek information about when and where these events occur. We also find that citizens are willing to share information about auroral sightings with others. The vision of Aurorasaurus leverages this to gather real-time data about the aurora in two ways, by direct entry into a questionnaire, and by continuous scanning of Twitter for tweets about sightings. Combined with data from Earth-based and satellite observatories, this allows Aurorasaurus to offer real-time predictions (i.e. early warnings) of auroral activity in both text and map form, likely with much greater accuracy and timeliness than the current state of the art.

Strong solar storms leading to visible aurora over population centers are rare events. Solar storms strong enough to threaten human life and property are exceedingly rare. This paper is not about creating an early warning system for severe and dangerous solar storms, despite the potential value of such a system. Rather, it is about creating a proof of concept using crowdsourcing to predict and *nowcast* (estimate the current state of) a rare event. The knowledge gained by this proof of concept can then be transferred to other sudden-onset, curious events such as tornadoes and earthquakes, potentially transforming our predictive capability.

In the remainder of this paper, we describe the innovation of a crowdsourced early warning system that uses aurora sightings to nowcast solar activity. We argue that a crowd of motivated citizen observers who document auroral activity can have a large impact on timely citizen awareness of the threat of solar activity or a curious, beautiful auroral display. We examine this system through the lens of boundary objects theory with the goal of spanning social worlds and strengthening civic engagement.

## A PERFECT STORM: AURORAL ACTIVITY ON THE RISE

The years around 2013 are the latest *solar maximum* recurring on an eleven-year *solar cycle*. During this time of high solar activity, Earth is likely to experience more intense solar radiation as the result of *coronal mass ejections* (CMEs) or solar flares, events which fling huge quantities of matter and electromagnetic radiation into space (Fry 2012). These solar particles can drive particularly strong *aurorae* in large regions around Earth's magnetic poles: *aurora borealis* or northern lights in the northern hemisphere and *aurora australis* or southern lights in the south. That is, now is the time when aurora is more likely over populated areas, and this is the first solar maximum with social media — an unprecedented opportunity to engage the public, the scientific community, and the media.

This engagement is important because extreme space weather is a threat to society. In addition to causing the aurora, increased solar activity can generate electric currents in conductive materials. By this interaction, intense *geomagnetic storms* can play havoc with ground- and space-based electrical systems that modern societies depend on.

The most extreme recorded geomagnetic storm, the Carrington Event, occurred relatively recently, in 1859. The storm was so intense that auroras were visible in the tropics, and electrical currents caused by the storm disrupted telegraph communications with sparks and fire. According to a report by the National Academy of Sciences, if a Carrington-level storm were to happen today, it could cause widespread loss of power, even in backup power systems, with cascading failures in water, communications, banking and finance, government services, and other critical infrastructure, leading to profound societal and economic consequences (Lovett 2011).

Specifically, the national electrical grid system is at high risk from such solar storms. For example, in 1989, currents from a geomagnetic storm damaged transformers in New Jersey and wiped out power through much of Quebec, leaving 6 million people without electricity (Medford 1989). In 2003, the Great Halloween Storms affected power in a number of places and damaged or interrupted satellite communications (Mannucci 2005). In 2006, satellite communications with ground stations were disrupted, including the critical Global Positioning System (GPS) (Gillis 2007).

What is needed is a means to accurately forecast geomagnetic storms. This would allow operators of power stations, electric grids, and satellites to take measures, which make them less vulnerable. Current predictive

capability yields forecasts, which are coarse and unreliable. For example, forecasts derived from state-of-the-solar wind models and satellite-based observations of the Sun estimate the arrival of the CME material at Earth to a typical accuracy of  $\pm 8$  hours. Space weather observations are simply not plentiful enough to accurately forecast the 93 million miles of interplanetary plasma between the Sun and the Earth. However, a network of observers on the ground — i.e., crowdsourced citizen science — can provide nowcasts, which are an order of magnitude more accurate by augmenting existing observations and models.

## CROWDSOURCING EARLY WARNING FOR CRISES

Early warning systems provide timely information regarding a hazard, which allows threatened individuals, communities, and organizations to prepare for that hazard, mitigating its effects (International Strategy for Disaster Reduction (ISDR), United Nations (UN), 2006; Kappes 2012).

In a general sense, disasters are grouped into two types, slow onset and sudden onset. Slow onset disasters are defined as one that does not emerge from a single, distinct event but one that emerges gradually, over time based on a confluence of different events. Examples of these are sometimes flooding, droughts, and some epidemics and military/civil crises. The United Nations Office for Coordination of Humanitarian Affairs finds that slowly unfolding emergencies can be mitigated by early response (OCHA 2011). If early warning and early response systems are fully functioning, coordinated and integrated, the longer lead time means the humanitarian community can step in early enough to reduce human suffering and help prevent the downward spiral of increased vulnerability to future hazards.

On the other hand, sudden-onset disasters cannot be predicted or anticipated and even with the best preparation. These capabilities are critical for sudden-onset disasters such as earthquakes, tornadoes, tsunamis, volcanic eruptions, and tropical cyclones, which cannot be predicted in the long term and which are affecting an increasing total population per event. Current early warning systems for these events can provide tens of seconds, minutes, minutes to hours, days to hours, and hours to weeks of warning, respectively (Petz 2011). In many cases, even warnings with tens of seconds of lead time can save lives.

Several scholars have recently turned to social media to test its potential for early warning, focusing on Twitter in particular due to its real-time nature (Sakaki, T., Okazaki, M., Matsuo, Y., 2010). “Earthquakes are one thing you can bet on being covered on Twitter (Twitter) first, because, quite frankly, if the ground is shaking, you’re going to tweet about it before it even registers with the USGS and long before it gets reported by the media” (Ostrow 2009).

In this work, we regard each Twitter user as a *soft sensor* — a human knowledge provider who processes information in qualitative and ambiguous ways and conveys that knowledge to an aggregating structure for the purposes of discussion, collaboration, information dissemination, reporting and decision-making. Soft sensors may act as individuals or as part of an ordered or ad-hoc network (Hall, McNeese, Llinas, & Mullen, 2008). Soft sensors are able to provide contextualized information and create ad-hoc information links that are consistent with knowledge elements and patterns. In other words, soft sensors are situationally aware. This research has produced terms such as *participatory sensing*, *urban sensing*, *human centered sensing*, *opportunistic sensing*, *people-centric sensing*, and *citizen sensing*.

However, humans vary, so soft sensors are unpredictable at best. Additionally, soft sensors are noisy compared to ordinal physical sensors because they are not always sensing (Sakaki, T., Okazaki, M., Matsuo, Y., 2010). The variability of expertise and the ad hoc nature of soft sensing, however, undermines the validity of the information shared and the trust that is subsequently elicited, limiting their overall efficacy (Stephenson, 2005).

To argue for the efficacy of Twitter as part of an early warning system, several scholars have demonstrated its limited power to forecast (see Arias, M., Arratia, A., and Xuriguera, R. 2012). Twitter has been used for real-time notifications such as large-scale fire emergencies and downtime on services provided by content providers (Motoyama, et. al 2010). Ginsberg et al. showed that the frequency of Google search terms can be used to build a linear model which accurately estimates influenza incidence (2009). Chen, et al. introduced a continuous data collection engine which combines the detection and prediction capability of social networks in discovering real world flu trends (Chen et al. 2010; Achrekar et. al. 2011). Wolfram attempted to predict the price of NASDAQ stock quotes by using Twitter as an additional source of information (2010). Another area in which this has been tried is in the prediction of movie ticket sales (Mishne and Glance 2005, Asur and Huberman 2010). Using similar techniques other researchers have used Twitter data for prediction within presidential polls (O’Connor et al. 2010), TV ratings (Wakamiya et al. 2011), and influenza rates (Lampos et al. 2010). Finally, Sakaki et al. (2010) used built an autonomous earthquake reporting system in Japan using Twitter users as sensors.

The last five years have seen the development of numerous “crowdsourcing” technologies, which leverage

networked digital technologies to enable scientists to bring research problems to non-scientists for participation and engagement. Increasingly, new tools are being developed to incorporate and facilitate participation by users who may or may not be scientists, thereby harnessing a distributed, networked labor force. Aurorasaurus is designed to present complex science in accessible ways and incentivize user engagement. In addition, users' emergent understanding of science is intertwined with the technologies that allow them to practice scientific discovery and communicate about it.

## CITIZEN SCIENCE AND CITIZEN OBSERVATORIES

With the advent of new technological tools, public participation in scientific practice has been enabled and supported to a degree unavailable even recently. Citizen science involves “organized research in which members of the public engage in the processes of scientific investigations: asking questions, collecting data, and/or interpreting results,” (from Citizen Science Central <http://www.citizenscience.org>.) Trumbull et al. found that citizen science can work on a massive scale, generating high quality data, which lead to reliable, valid scientific outcomes as well as unexpected insights and innovations (2000).

In the Aurorasaurus project, we are particularly interested in the role of technologies in supporting scientific practice rather than curiosity-driven participation (Newman et al. 2012). That is, we wish to test citizen science projects entirely mediated by information and communication technologies (ICTs), often considered a form of crowdsourcing applied to science (Wiggins and Crowston 2011). These technologies have the potential to engage broad audiences (Clery 2011), motivate volunteers (Cooper et al. 2010), improve data collection (Willett et al. 2010), improve data quality (Kelling et al. 2009), corroborate model results (Darg et al. 2011), and increase decision speed (Danielsen et al. 2010).

We engage two forms of participation: (a) collecting and managing data and (b) analyzing and interpreting data. We involve citizen scientists in the research, using technology to support our development. Our vision is to involve citizen scientists in scientific observation through contribution of tweets, structured reports, pictures, and video of aurora activity; verification of Twitter observations; and the analysis of the corrected data to inform science. That is, we will enable citizen to be intimately involved in the operation of an alert system to inform other citizens of aurora activity.

*Citizen observatories* are a form of citizen science in which crowd sourcing techniques, particularly the use of mobile devices for data collection and geographic information, are employed. One example of this is the COBWEB (Citizen OBServatory WEB) project, which allows citizens to collect environmental data suitable for use in research, decision-making and policy formation. Another such observatory is the WeSenseIt project, an observatory of water, which allows citizens and communities to become active stakeholders in information capturing, evaluation, and communication about water. The project enables citizens to make use of innovative sensor devices — citizens' collective intelligence — through monitoring social network communications and allowing citizens and communities to upload data to the observatory (see F. Ciravegna, et. al 2013).

Most importantly, we seek to create a community of citizen scientists who engage with the broader scientific picture and practice so that these crowdsourced tasks do not become repetitive and tedious and do not result in a loss of interest, in order to improve scientific longevity.

## BOUNDARY OBJECTS

*Boundary objects theory* focuses on communication and collaboration among heterogeneous communities around a particular idea — physical or metaphorical. The core concept of boundary objects theory is that there are artifacts that span the boundaries of communities — termed “social worlds” — and through which communication and collaboration can be achieved (Fleischmann, 2006a, 2006b; Star, 1989; Star & Griesemer, 1989). A social world is a community organized around specific activities, concerns, and forms of communication (Covi & Kling, 1996; Fleischmann, 2006a, 2006b). Boundary objects exist independently of any single social world, but are negotiated by each individual social world that engages them (Fleischmann, 2006a; Gasson, 2005). This dichotomy requires that boundary objects be flexible enough to be interpretable by individual social worlds, while rigid enough that they retain meaning independent of any one social world (Star, 1989; Star & Griesemer, 1989). In essence, these objects allow different types of people to negotiate on the definition of that object in different ways while uniting the groups. Boundary objects have been particularly useful for understanding collaboration among groups using technology, since a technological object is not always used as intended by its developers.

Most streams of boundary object research focus on the use and creation of artifacts rather than the properties of the artifacts themselves. This has resulted in an emphasis on capturing the negotiations and understandings that

occur around and through boundary objects; a consequence of this focus on process has been that boundary objects themselves remain largely underspecified and highly abstract. There has been research into identifying additional properties of boundary objects (Levina, 2005). For example, specific types of boundary objects are better suited for spanning specific types of knowledge boundaries (Carlile, 2002); similarly, online communities serve many different types of users, and this robust, asymmetrical (or customizable, open) environment is more inclusive and will attract many different types of users that will build consensus (Østerlund & Crowston 2011).

Boundary objects are a means of understanding communication among heterogeneous groups. By being both flexible enough to be negotiated and understood by individual groups and rigid enough to exist independent of any particular group, a boundary object facilitates communication through standardization (standards and forms), abstraction (ideal/platonic), flexibility (terrain with coincident boundaries), and collection (repositories).

Aurorae represent boundary objects themselves since many different types of people observe and apply their own meanings to these phenomena. While these groups' interaction could be analyzed through their interaction via this natural boundary object, an additional technological boundary object like Aurorasaurus will provide a closer coupling between social scientists, space scientists, and various interested citizens. This represents a novel application of the theory.

### PROOF OF CONCEPT: PASSIVE DATA COLLECTION

During the testing of our prototype, we found that the *Kp index*, a key measure of auroral activity, correlates strongly with the number of aurora-related tweets. A comparison of peaks in tweets to the 3-hour magnetometer-derived Kp index shows a strong correlation when Kp rises above 4 (Figure 1), suggesting that volunteered Twitter data is a valid indicator of real-world events. In general, the tweets rise when auroral activity rises above background levels and when the aurora starts to be visible over more populated areas. We will continue to evaluate and refine these algorithms to improve filtering. This Twitter-based approach is interesting because while instruments can give a general notion of global aurora activity, they cannot indicate specific locations nor precise levels of activity.

Our current Twitter data pipeline consists of a simple machine-learning filter, which includes topical terms to include and exclude (for example, to remove tweets related to Aurora, Colorado or the band The Northern Lights rather than the space phenomenon). For this three-month period we have 37,000 tweets in a cleaned database. The raw collected tweets go through a proprietary machine-learning algorithm. The categories were:

- 5 - Not related to aurora borealis
- 4 - Related to aurora borealis but not specific sighting
- 3 - RT related to aurora borealis
- 2 - Related to aurora borealis - person didn't see it
- 1 - Related to aurora borealis - person saw it
- 0 - not yet classified

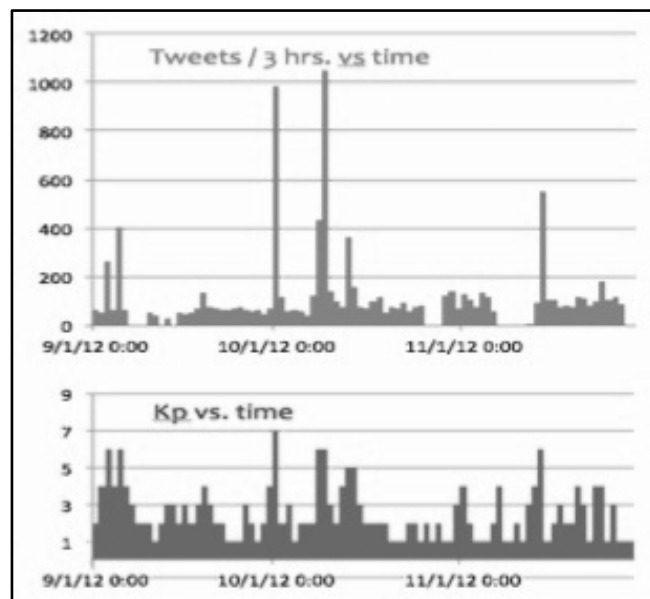


Figure 1: Kp index and tweet correlation

In the future, we plan to automate this process, leveraging user feedback on specific tweets to improve the machine classification of tweets over time. The accuracy of the tweet filtering has not been assessed yet, rather all cleaned tweets with a derived location are displayed. Over 1000 tweets from an active day were hand classified to train the machine learning algorithms. Similar to disaster response efforts (e.g. Crowdcrafting.org) the tweets can be most effectively validated by a human in the loop and that is the planned approach.

Using these new data we anticipate three major impacts on geospace science: 1) Significantly more accurate, timely, and relevant methods of auroral predictions. Predicting the visibility of aurora for the public is not the primary goal of the National Weather Service's Space Weather Prediction Center, which is tasked to predict space weather for national satellite, airborne, and ground assets. 2) New cost-effective sources of human-derived data for observing the aurora, including volunteered geographic information, images and videos, which may aid scientists in understanding their space-based data. These data also serve as an accurate ground truth of visibility

with extremely high spatial and temporal resolution. 3) The ability to study auroras at mid-latitudes during storms. A storm has never been imaged at mid-latitudes and the last global imaging satellites with significantly lower resolution ended their missions in 2006. Space scientists are still trying to connect many magnetospheric regions to their visible manifestations in the ionosphere in this data-starved, vast, and relatively young field. Other citizen science activities may yield discoveries on auroral height and the shape of the dynamic auroral oval versus time and activity due to the current lack of data on these phenomena.

## AURORASAURUS: ACTIVE PARTICIPATION AND DATA COLLECTION

A prototype version of Aurorasaurus (Figure 2) has been live on the Web for one year and has shown promising results. At its core, Aurorasaurus offers an estimate of the aurora's location and character while simultaneously seeking participatory data for improving this nowcast. User observations of the aurora — collected from Twitter as well as an entry form offered on the website itself — are displayed on a map as pins. Users can click on these pins to see more detail about an observation, as illustrated; sometimes this detail includes photos or videos. Aurorasaurus also displays a Kp-derived nowcast and 1-hour forecast of the auroral oval's location (in the future, we plan to integrate user observations as part of the estimation algorithm).

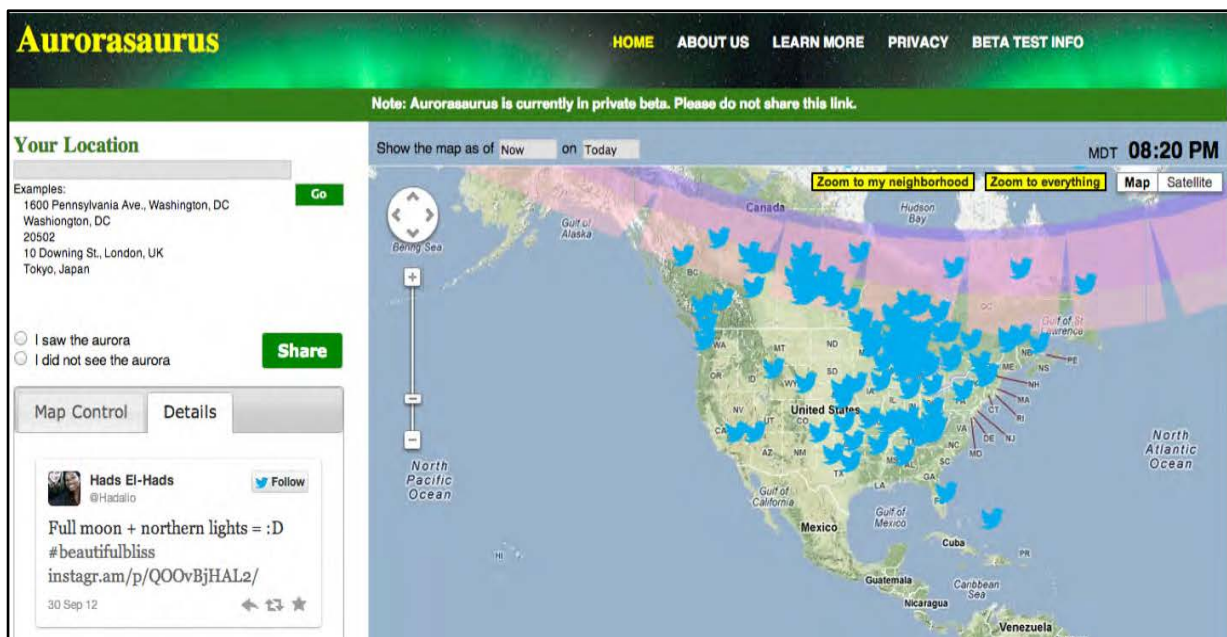


Figure 2 - Aurorasaurus during an active auroral event.

Aurorasaurus is implemented using a standard cloud-hosted web application stack including Django, Python, PostGIS, HTML5 Location API, and Google Maps. Its code is open source.<sup>2</sup> We plan to augment this web implementation with a mobile application, in order to make Aurorasaurus more fully available in the field, where its services are most useful.

Our data philosophy is to maximize the information used for auroral estimation and to present clearly to users everything the application “knows” in a mix of algorithmic integration and less-processed data which must be interpreted by the user. Specifically, data sources include:

- citizen reports submitted using a questionnaire on the website, which include at a minimum the time and location of the sighting;
- tweets filtered by keyword and geo-located, which include some combination of text, photo, and video;
- space weather data (Kp index) from NOAA; and
- cloud cover and weather data from Google.

The Twitter data stream requires a significant level of filtering to select aurora-related tweets from the daily 300-500 million tweets shared in total. Our current classification algorithm, trained on a set of tweets manually classified by us, divides a tweet stream initially filtered by keyword into the following classes: positive sighting (i.e., someone saw the aurora), negative sighting (someone looked for the aurora but failed to see it), desire to

<sup>2</sup> <https://bitbucket.org/reidpr/aurora-website>



see the aurora, not aurora related, and unclassifiable. Geo-location is accomplished either by using the (typically derived from GPS or other automated location services) geotag provided by Twitter directly or a gazetteer-based approach applied to toponyms (place names) in the tweet text or user profile. This yields approximately 400 tweets on a typical day and 1,000 or more during a non-routine event.

We plan to improve this process in two ways. First, we will bring users into the tweet filtering loop by allowing them to give feedback on the classification algorithm's judgments, perhaps using up- and down- votes similar to those found on the Reddit social news site; this will both improve the algorithm's accuracy and make it more nimble in the face of changing language. Second, we will integrate improved location inference techniques (for example, Priedhorsky et al. 2014).

Other planned features include:

- Sharing (e.g., by tweeting or retweeting) content of interest.
- Push notifications to alert users of ongoing or expected auroral activity in their locale.

Our data flow vision is summarized in Figure 3. Once fully implemented, the application will combine citizen science input with sensor-based observations of solar wind and other factors to produce highly localized, detailed nowcasts and forecasts of auroral activity, presented in text and map form on both the web and mobile devices. No other aurora forecasting system implements this level of utility for either citizens or professional space scientists.

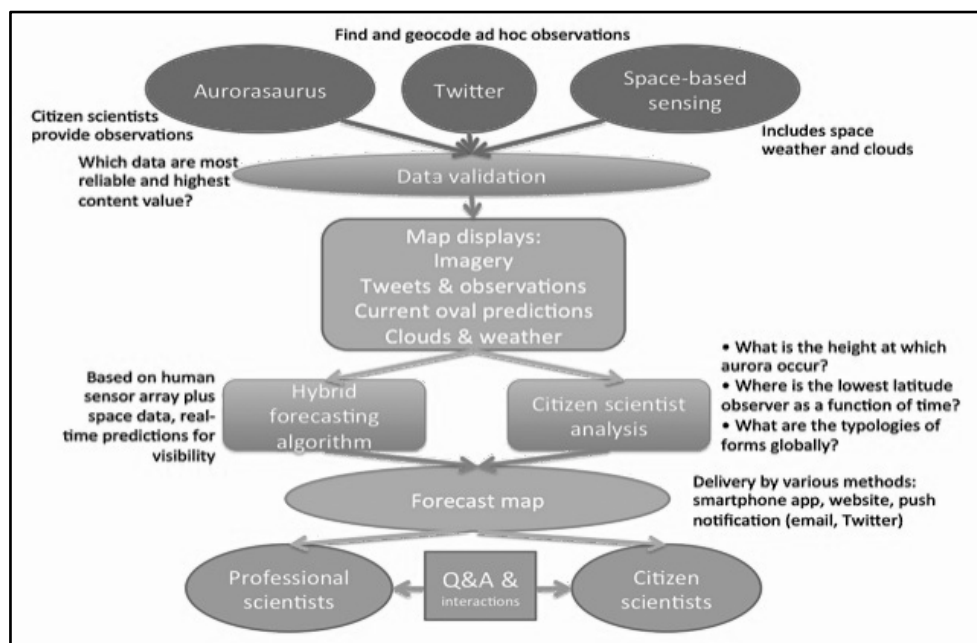


Figure 3 - Our vision for Aurorasaurus' data flow.

## CITIZEN SCIENCE AURORASARUS

As human society becomes ever more reliant on satellite-based technologies and begins to explore away from its home planet, the imperative to educate the public on the science of living with a variable star — our Sun — increases. We have initiated and look forward to continuing to build a *citizen science community*. That is, while many people have heard of the aurora and seen photographs, fewer have seen the curious and remarkable phenomenon personally, and fewer still can describe its causes in terms. This general lack of knowledge is an opportunity, one which we are taking advantage of, making a relevant understanding of space weather available to many people for the first time.

Our vision is to engage citizen scientists in every phase of the research, from scientific observations contributed as tweets, structured “lab notebook” logs, images, and video, to verification and cleaning of data, to analysis; the results will inform scientific research in space science, machine learning, and other fields. We will offer a library of educational resources to stimulate curiosity and in-depth investigations of aurora; this library will provide resources for participants from novice to expert, on topics including aurora itself, the effects of space weather on society, space weather models, data and their uncertainties; additionally, it will offer tutorials that support the tools and research activities on the website. We will offer a suite of citizen activities that advance

current understanding of space weather. Finally, we will offer learning modules to support the specific citizen science projects that can be undertaken by a wide variety of audiences for both deep and meaningful learning as well as science.

Specific scientific problems that citizens can help address include:

- Characterizing typologies of auroral forms and understanding how these typologies evolve over time. For example, citizen scientists can determine the lowest-latitude observation in real time through the course of an event, thus quantifying the speed at which aurora can move from north to south and vice versa.
- Calculating auroral parameters such as peak emission altitude by observing the angle above the horizon at which aurora is visible; triangulation of reports from different locations can yield a precise 3D location of specific light-emitting features.
- Improving aurora nowcast and forecast models by making and verifying geo-located observations of aurora activity (or lack thereof). These improved models will in turn facilitate a deeper understanding of the physics driving the aurora.

## DISCUSSION AND CONCLUSIONS

Seeing the aurora is often an emotionally meaningful moment in someone's life, one which merits membership in "bucket lists" and/or major travel at financial and time cost. In this context, the significance of a solar maximum, auroral activity moves southward, become viewable by larger populations; that is, the aurora comes to the people rather than the people coming to the aurora. However, while visible aurora becomes more likely at middle latitudes, it is still a rare event — one cannot intuitively know when and where aurora will appear. Therefore, a mechanism, which alerts interested parties to this event, is valuable in itself.

We argue that this broadly shared deep interest provides three valuable opportunities. First, it can engage public interest in the space weather and science underlying the auroral activity. Second, it can contribute to the body of scientific theory and practice around the aurora. Finally, it can provide a testing environment for a crowdsourced early warning system.

Aurorasaurus is an attempt to create a boundary object out of the disparate types of communities who engage with the aurora. While the aurora itself can be considered a gigantic boundary object *in vacuo*, technological negotiating elements such as the website and mobile application can enhance the ability of many different types of users to create value together. That is, differing types of people will shape not only various types of science, including that of early warning systems, but also simply knowledge about how people view and enjoy curiously beautiful aurora. By creating an object that unites these people, we hope to promote both observation of the aurora and knowledge about space weather and geomagnetic phenomena.

During early testing of the Aurorasaurus prototype, there was a strong solar event with very significant auroral activity (see above for details). A time series comparison of the global strength of the aurora, embodied by the Kp index, with the number of tweets related to auroral sightings yielded a strong correlation. That is, gathering and processing social media data can enable a near real-time indicator of auroral activity. However, our data processing pipeline can be improved by human feedback on its output; this is a key feature we will add to Aurorasaurus. In addition, as interest in Aurorasaurus grows, we expect to gather more observations directly on the website and mobile application, which is useful because such observations are more structured.

Auroral nowcasts and forecasts generated by Aurorasaurus and its user community are inherently geographic. Thus, map-based information is presented to users; this map forms an early warning system to help interested people identify nearby current and future auroral state in real time. For contexts where map display is inconvenient or impossible, we will also offer text-based alerts to summarize the system's estimate. Both types of alert will be offered in a push form.

Through this, we create a community that is both interested in and contributes to space weather science. Participants will be offered educational modules about space weather and training on how to filter images and tweets as they arrive. This encourages learning about space weather beyond that of common knowledge. Free text observations in existing social media, structured observations entered via questionnaire, both text and visual, comprise scientific contributions to improve understanding auroral activity and prediction. This research will have a positive impact on the concept of nowcasting and forecasting as they relate to space weather.

Key to all citizen science efforts is longevity. While interfaces may change and design aesthetics evolve, the encouragement of long-term participation in a collective scientific endeavor is paramount to a project's long-term viability; this includes recruiting, training, and retaining interested and motivated participants. We argue



that people are already interested in the aurora and would be further motivated to participate if they could be alerted to potential sightings in real time. Similarly, a functioning early warning system based on crowd participation requires a well-trained, motivated set of continuous observers equipped with tools to report the first signs of distress. We argue that aurora sightings have the potential to yield insight for both applications, serving as the motivation for participation in an early warning system as well as citizen science projects.

We expect that our tools will serve as a proof-of-concept for crowdsourcing other forms of early warning system. That is, we gather ad-hoc data indirectly offered by the crowd through social media; we categorize and filter this raw data with machine learning techniques; we employ a community of human analysts to further refine the data; we transform the resulting insights to map- and text-based forms; and, finally, we send these insights to interested participants in a timely manner. Such a system could be employed during other kinds of disasters such as earthquakes, tornados, tsunamis, and flooding. That is, real-time alerts informed by crowd participation could save lives.

This paper outlines a product meant to capitalize on human curiosity in ways that augment instrument-based observation of space weather. Auroral events, though rare and ancillary to the creation of this website and device, allow us to almost inadvertently create the first stages of an early warning system that incorporates soft sensor data. Aurorasaurus does this by offering a geographically grounded system for collecting, analyzing, interpreting, and redistributing data on the state and evolution of auroral events. Through ad-hoc data like tweets and structured observations entered by citizen scientists into the system we are developing a product that will not only aid in developing an early warning system for rare event detection, but a product that will aid in science literacy and participation. Finally, we facilitate longer-term participation by engaging citizen scientists in basic research of aurora.

## ACKNOWLEDGEMENT

This material is based upon work supported, in part, by the National Science Foundation (NSF) under Grant #1344296. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF. This work was supported in part by the U.S. Department of Energy through the Los Alamos National Laboratory LDRD Program. LANL is operated by Los Alamos National Security, LLC for the Department of Energy under contract DE-AC52-06NA25396. Approved by LANL for public release, LA-UR 14-20644. Twitter data and first-level filtering courtesy of SocialFlow, Inc.

## REFERENCES

1. Arias, M., Arratia, A., and Xuriguera, R. 2012. Forecasting with Twitter Data. *ACM Trans. Intell. Syst. Technol.* V, N, Article A (July 2012), 25 pages.
2. Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2010 IEEE/WIC/ACM International Conference on (Vol. 1, pp. 492-499). IEEE.
3. Bonney, R. & LaBranche, M. (2004), 'Citizen science: Involving the public in research', *ASTC Dimensions* p. 13.
4. Achrekar, H., Gandhe, A., Lazarus, R., Yu, S. H., & Liu, B. (2011, April). Predicting flu trends using twitter data. In *Computer Communications Workshops (INFOCOM WKSHPS)*, 2011 IEEE Conference on (pp. 702-707). IEEE.
5. Carlile, P. R. (2002). A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science*, 13(4), 442-455.
6. Chen, H. Achrekar, B. Liu, and R. Lazarus, "Vision: towards real time epidemic vigilance through online social networks: introducing sneft – social network enabled flu trends." *ACM Mobile Cloud Computing and Services*, San Francisco, California, June 2010.
7. Ciravegna, F., Huwald, Lanfranchi, Wehn de Montalvo Citizen observatories: the WeSenseIt vision [http://inspire.jrc.ec.europa.eu/events/conferences/inspire\\_2013/schedule/submissions/261.pdf](http://inspire.jrc.ec.europa.eu/events/conferences/inspire_2013/schedule/submissions/261.pdf)
8. Clery D. 2011. Galaxy Zoo volunteers share pain and glory of research. *Science* 333: 173–75.
9. COBWEB, <http://cobwebproject.eu/project#sthash.0AtfmUDT.dpuf>
10. Covi, L., & Kling, R. (1996). Organizational dimensions of effective digital library use: Closed rational and open natural systems models. *Journal of the American Society for Information Science*, 47(9), 672-689.
11. Danielsen F, Burgess ND, Balmford A, *et al.* 2009. Local participation in natural resource monitoring: a characterization of approaches. *Conserv Biol* 23: 31–42.
12. Danielsen F, Burgess ND, Jensen PM, and Pirhofer-Walzl K. 2010. Environmental monitoring: the scale and speed of implementation varies according to the degree of people's involvement. *J Appl Ecol* 47: 1166–68.
13. Darg DW, Kaviraj S, Lintott CJ, *et al.* 2011. Galaxy Zoo: multi-mergers and the Millennium Simulation. *Mont Not R Astron Soc* 416: 1745–55.
14. Fleischmann, K. R. (2006a). Boundary Objects with Agency: A Method for Studying the Design-Use Interface. *The Information Society*, 22(2), 77-87.
15. Fleischmann, K. R. (2006b). Do-it-yourself information technology: Role hybridization and the design-use interface. *Journal of the American Society for Information Science and Technology*, 57(1), 87.
16. Fry, E. K. (2012). The risks and impacts of space weather: Policy recommendations and initiatives. *Space Policy*, 28(3), 180-184.

*Proceedings of the 11<sup>th</sup> International ISCRAM Conference – University Park, Pennsylvania, USA, May 2014*  
S.R. Hiltz, M.S. Pfaff, L. Plotnick, and P.C. Shih, eds.

17. Gasson, S. (2005, January 3 - 6). *Boundary-Spanning Knowledge-Sharing in E-Collaboration*. Paper presented at the 38th Hawaii International Conference on System Sciences, Island of Hawaii.
18. Gillis, Carmeyia. 2007. Researchers find global positioning system is significantly impacted by powerful solar radio burst. *NOAA Magazine* retrieved on November 13, 2013 from <http://www.noaa.gov/stories2007/s2831.htm>
19. Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2008). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014.
20. Hall, D. L., Llinas, J., McNeese, M., and Mullen, T. (June/July 2008) "A Framework for Hard/Soft Fusion", The 11th International Conference on Information Fusion (FUSION 2008), Cologne, Germany, June/July 2008, pp 1-8.
21. Kappes, M. S. (2012-11). Challenges of analyzing multi-hazard risk: a review. *Natural hazards (Dordrecht)*, 64(2), 1925-1958.
22. Kelling S, Hochachka WM, Fink D, *et al.* 2009. Data-intensive science: a new paradigm for biodiversity studies. *BioScience* 59: 613–20. Kim J, Lee E, Thomas T, and Dombrowski C. 2009. Storytelling in new media: the case of alternate reality games 2001–2009. *First Monday* 14.
23. Kuznetsov, S. (2013). Expanding our visions of citizen science. *interactions*, 20(4), 26-31.
24. Lampos, V., De Bie, T., & Cristianini, N. (2010). Flu detector-tracking epidemics on Twitter. In *Machine Learning and Knowledge Discovery in Databases* (pp. 599-602). Springer Berlin Heidelberg.
25. Levina, N. (2005). Collaborating on Multiparty Information Systems Development Projects: A Collective Reflection-in-Action View. *Information Systems Research*, 16(2), 109-130.
26. Lovett, Richard A. 2011. What if the biggest solar storm on record happened today? *National Geographic News* retrieved on November 30, 2013 from <http://news.nationalgeographic.com/news/2011/03/110302-solar-flares-sun-storms-earth-danger-carrington-event-science/MacDonald>, E. A. (2012). TweetAurora. com-Citizen Scientists Experiencing the Extremes of Space Weather (No. LA-UR-12-20634). Los Alamos National Laboratory (LANL).
27. Mannucci, A. J., Tsurutani, B. T., Iijima, B. A., Komjathy, A., Saito, A., Gonzalez, W. D., ... & Skoug, R. (2005). Dayside global ionospheric response to the major interplanetary events of October 29–30, 2003 “Halloween Storms”. *Geophysical Research Letters*, 32(12).
28. Mark, G., Lyytinen, K., & Bergman, M. (2007). Boundary objects in design: an ecological view of design artifacts. *Journal of the Association for Information Systems*, 8(11).
29. Medford, L. V., Lanzerotti, L. J., Kraus, J. S., & MacLennan, C. G. (1989). Transatlantic earth potential variations during the March 1989 magnetic storms. *Geophysical Research Letters*, 16(10), 1145-1148.
30. Mishne, G., & Glance, N. S. (2006, March). Predicting Movie Sales from Blogger Sentiment. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs* (pp. 155-158).
31. Motoyama, M., Meeder, B., Levchenko, K., Voelker, G. M., & Savage, S. (2010). Measuring online service availability using twitter. *WOSN'10*, 13-13.
32. Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., & Crowston, K. (August 2012). Citizen Science Futures: Emerging Technologies and Shifting Paradigms. *Frontiers in Ecology and the Environment*, 10(6):298-307
33. O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *ICWSM*, 11, 122-129.
34. OCHA (2011). OCHA and slow-onset emergencies, OCHA Occasional Policy Briefing Series, No. 6, New York: Policy Development and Studies Branch (PDSP)/UN Office for the Coordination of Humanitarian Affairs (OCHA).
35. Østerlund, C., & Crowston, K. (2011). Boundary-spanning documents in online communities. *International Conference on Information Systems*.
36. Ostrow, Adam. 2009. Japan Earthquake shakes twitter users ... and Beyonce. *Mashable* retrieved on November 30, 2013 from <http://mashable.com/2009/08/12/japan-earthquake/>.
37. Petz, Daniel. 2011. Regional Workshop on Internal Displacement caused by Natural Disasters and Climate Change in the Pacific. *Pacific Island Forum Secretariate, Suva, Fiji*.
38. Priedhorsky, R. Aron Culotta, and Sara Y. Del Valle, “Inferring the Origin Locations of Tweets with Quantitative Confidence”. *Proc. ACM CSCW 2014*.
39. Sakaki, T., Okazaki, M., & Matsuo, Y. (2010, April). Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web* (pp. 851-860). ACM.
40. Star, S. L. (1989). The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. In M. Kuhn & L. Gasser (Eds.), *Readings in Distributed Artificial Intelligence* (pp. 36-54). Menlo Park, CA: Morgan Kaufman.
41. Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3), 387-420.
42. Trumbull, D. J., Bonney, R., Bascom, D., & Cabral, A. (2000). Thinking scientifically during participation in a citizen-science project. *Science Education*, 84(2), 265-275.
43. Wakamiya, S., Lee, R., & Sumiya, K. (2011, February). Towards better TV viewing rates: exploiting crowd's media life logs over Twitter for TV rating. In *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication* (p. 39). ACM.
44. Whiter, D. K., Gustavsson, B., Partamies, N., & Sangalli, L. (2012). A new automatic method for estimating the peak auroral emission height from all-sky camera images. *Geoscientific Instrumentation, Methods and Data Systems Discussions*, 2, 893-924.
45. Wiggins, A. & Crowston, K. (2010). Developing a Conceptual Model of Virtual Organizations for Citizen Science. *International Journal of Organisational Design and Engineering*, 1(1/2), 148–162.
46. Wiggins, A., Crowston, K. (2011) *Describing Public Participation in Scientific Research*, New York
47. Wolfram, M. S. A. (2011). *Modelling the Stock Market using Twitter*. Masters Thesis for The University of Edinburgh.