The Geological Society of America
Special Paper 492
2012

Extreme dynamic mapping: Animals map themselves on the "Cloud"

Eugene Potapov*
Bryn Athyn College, College Drive, Bryn Athyn, Pennsylvania 19009, USA

Valery Hronusov*
"Inform++ Ltd," Shatrova St, 20-25, Perm 614064, Russia

ABSTRACT

Animal tracking data are routinely delivered in the form of e-mail messages with an attachment or in the main text of an e-mail that includes satellite-telemetry data provided by Argos services. Downloading these data onto a computer, transferring them into shapefiles, filtering, processing, and displaying them consumes considerable end-user time and energy. In this paper, we demonstrate that freely available "Cloud"-based services are sufficient to take over this workload and fast enough to deliver spatial data to an end-user without a considerable investment of time. The animal-generated spatial data we present come in two forms: satellite data from the Argos service and GPS data delivered as text messages using a Short Message Service (SMS). We suggest a simple mail-to-map system, which automatically archives data (coordinates, time, telemetry) and displays it dynamically on various internet applications such as Google Maps/Google Earth or Google Graphs. We use the Gmail service to filter messages, a free blog service (e.g., blogger.com or wordpress.com) for unlimited-time data storage and the Google spreadsheets to dynamically assemble the KML (Keyhole markup language) files. To demonstrate the utility of our mail-to-map system, we apply the approach to two contrasting wildlife case studies—the highly endangered Steller's Sea Eagle (*Haliaetus pelagicus*) of northeast Asia and White-tailed Deer (*Odolescens virginianus*), which is ubiquitous in the eastern United States—and discuss conservation implications of the near-real-time data publication opportunities that our system provides.

INTRODUCTION

The conventional methods of obtaining geospatial data from various PTT/GPS (Platform Transmitter Terminal/Global Positioning System) data loggers involve downloading the data from a company provider's server (e.g., Argos website, Followit, NorthStar Telemetry) or extracting the data from provider-generated e-mails. Once downloaded, extensive processing on the client's computer is necessary to make these data available on an Internet web page. Failure to download the data in time generates serious gaps in the data flow, as some providers do not store the data forever. For example, to this day, the Argos system, which has a worldwide monopoly on satellite-relayed wildlife data, keeps only 10 days of the data available for online access (Argos User

*Eugene.Potapov@brynathyn.edu; xbbster@gmail.com.

Potapov, E., and Hronusov, V., 2012, Extreme dynamic mapping: Animals map themselves on the "Cloud," *in* Whitmeyer, S.J., Bailey, J.E., De Paor, D.G., and Ornduff, T., eds., Google Earth and Virtual Visualizations in Geoscience Education and Research: Geological Society of America Special Paper 492, p. 1–XXX, doi:10.1130/2012.2492(10). For permission to copy, contact editing@geosociety.org. © 2012 The Geological Society of America. All rights reserved.

Manual, 6.3.3, p. 38). Data downloads that have to be done every 10 days require the effort and time of a researcher and often are incompatible with fieldwork. These problems can be overcome by using automated scripts on dedicated corporate servers. However, this requires substantial overhead costs, including system administrator salaries and software and hardware expenses. In this paper, we suggest a simple and cost-efficient alternative to the conventional (human-to-computer intensive) and corporate (high cost) models. We suggest accumulating and storing data on the theoretically indestructible and free "Google-cloud"–based web service, which allows near-real-time data processing for generating outputs on Google Earth/Google Map services.

Prior to 2004, none of the existing wildlife tracking systems was designed to provide continual real-time data access (Clark et al., 2006). In 2004, there was an attempt to design a real-time tracking system for moose in Sweden (Dettki et al., 2004), which was based on ArcIMS server (ESRI, 2004). A new philosophy of data storage and access in the form of "software as an online service" arrived with the launch of Google Maps and Google Earth in 2005 and Google Spreadsheets in 2006. The availability of these highly integrative products has opened new possibilities for the presentation of the dynamically changing geolocations of animals.

METHODS

Configuration of E-mail Clients

The majority of tracking systems (Argos or GPS/GSM [Global System for Mobile Communications]) offer e-mail as a vehicle to deliver the data, a convenient choice since e-mail systems can be easily automated. We recommended Google's Gmail service for mailing telemetry data since it is not affected by corporate e-mail server rules and has sufficient capacity to store the data. It is also possible to set individual e-mail accounts for an Argos program or subprogram, or even for an individual animal in the case of high-density GPS/GSM data streams. Gmail accounts are an attractive option since these can be configured with ease and have virtually no size limits. The size of an incoming e-mail associated with telemetry data can jeopardize data flow in the corporate model because servers in these environments might use unpredictable firewall settings and require a third party to make changes. A filter should be set up to forward e-mail messages to the user's data-storage system. It is recommended to filter the messages by the sender e-mail address and by subject line.

Data Storage

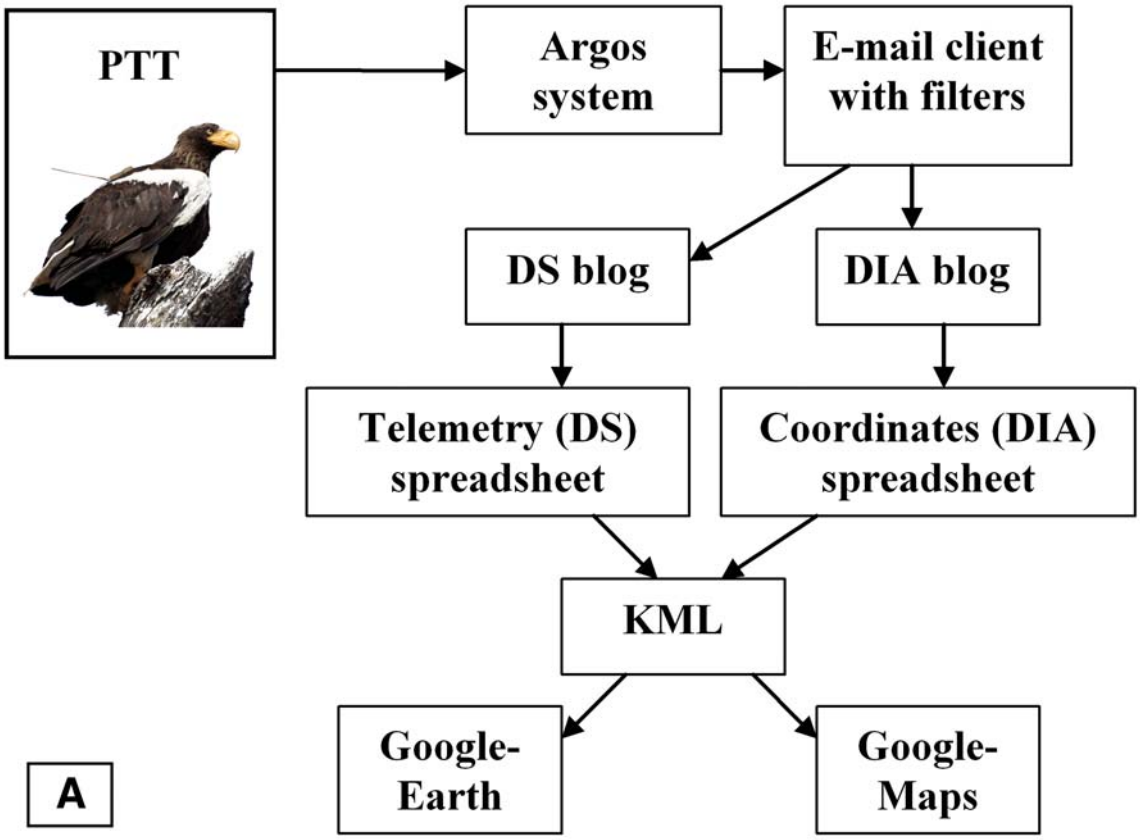
We suggest storing the data in a "blog," originally an online journal for sharing diary entries (Fellbaum, 1998). Most blogging software provides the option to send posts via e-mail. The free, cloud-based, blog service run by Google (www.blogger.com) is appealing, since it is virtually indestructible and has no size limit. The data is forwarded, as described above, from Gmail (or another mailer) to the blog by automated filters. For GPS/GSM monitoring it is recommended to have one blog per animal (http://goo.gl/EnKCl). For the Argos system, filters should be set-up to send diagnostic data (DIA; http://wp.me/2egoY) and telemetry data (DS; http://wp.me/2efLy) files to separate blogs (Fig. 1). The configuration of the blog is critical, as the data stored there will be accessed by spreadsheets for data import. The amount of the data displayed on the front page of the blog determines the amount of data subsequently displayed on the associated map. A blog does not set limits on data storage time (can be indefinite), and it also allows for the exporting of the entire blog, i.e., the entire data set. This is a significant improvement over other systems, for example, the Argos system, which limits the online data access to "one current and past 9 days" (Argos User Manual, 2011, chapter 6.3.3, p. 38).

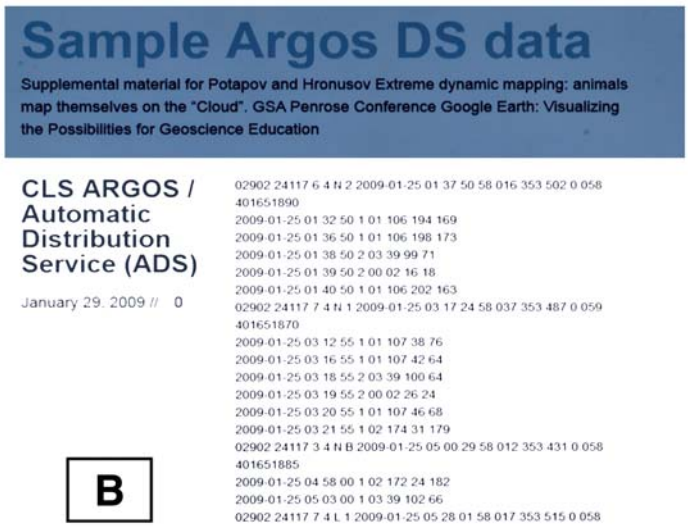
Data Processing

Once the data arrive at the designated blog(s), they become available for automated reading. This can be achieved by the import command within online spreadsheets (e.g., Google Documents or editgrid.com). Since Google spreadsheets are linked to Gmail, we highly recommend using this service for its convenience. The data from the blog gets imported into the Google spreadsheet by the command:

=importxml("http://*yourblog*.blogspot.com/","//p[1]/text()")

The values of "//p[1]/text()" depend on the way the blog is configured. In our case studies we deployed "Tic-Tac" and "Simple" templates of the www.blogger.com service. The *importxml* function reads the data into a spreadsheet with a periodicity of ~30 min (Google spreadsheet function list, 2011). The read data are broken up into values using "split()" command, and are cleaned up of unnecessary bits of text using the command "filter." The data are sorted chronologically, checked by simple error-tests that eliminate erroneous fixes, and are assembled into dynamic Keyhole markup language (KML) file(s) for the near-real-time display of the data on Google Maps/Google Earth/Google Graphs (see example at http://goo.gl/EnKCl). For the GPS/GSM service the error test is a simple elimination of the lines with the GPS time-out error, the elimination of duplicates, and optionally removing noisy fixes, whereas for Argos the error test includes not only the elimination of duplicates, but also a switch between solution 1 and solution 2 (see Argos User Manual, 2011, for terminology and example spreadsheet http://goo.gl/gGVPU). When using Argos-based PTT (Platform Transmitter Terminal, also known as satellite tag), it is very important to have the readings of telemetry sensors (battery voltage, temperature, etc.). This can be achieved automatically by reading the blog containing DS data (http://wp.me/2efLy), and decoding the sensors using the formula supplied by the device's manufacturer (http://goo.gl/gGVPU). The end result is a graph with the telemetry readings (Fig. 2), which can be embedded into KML file.





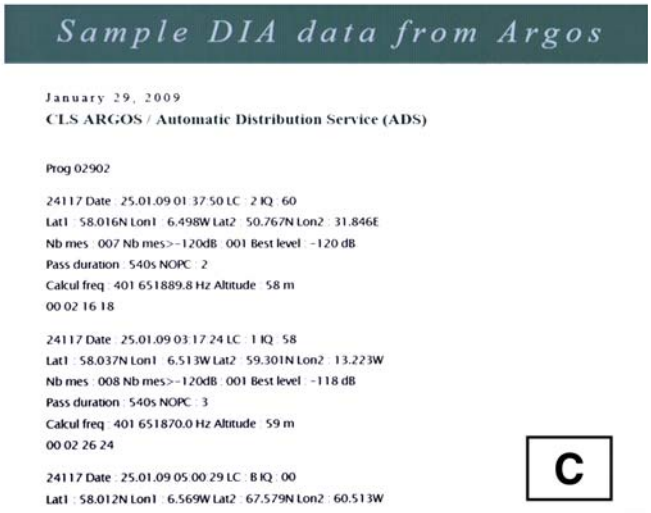


Figure 1. (A) Block diagram of the automated decoding of the Argos-based system. (B) Screenshot of the telemetry DS blog (http://wp.me/2efLy). (C) Screenshot of the DIA data blog which provides coordinates and time-stamp (http://wp.me/2egoY).

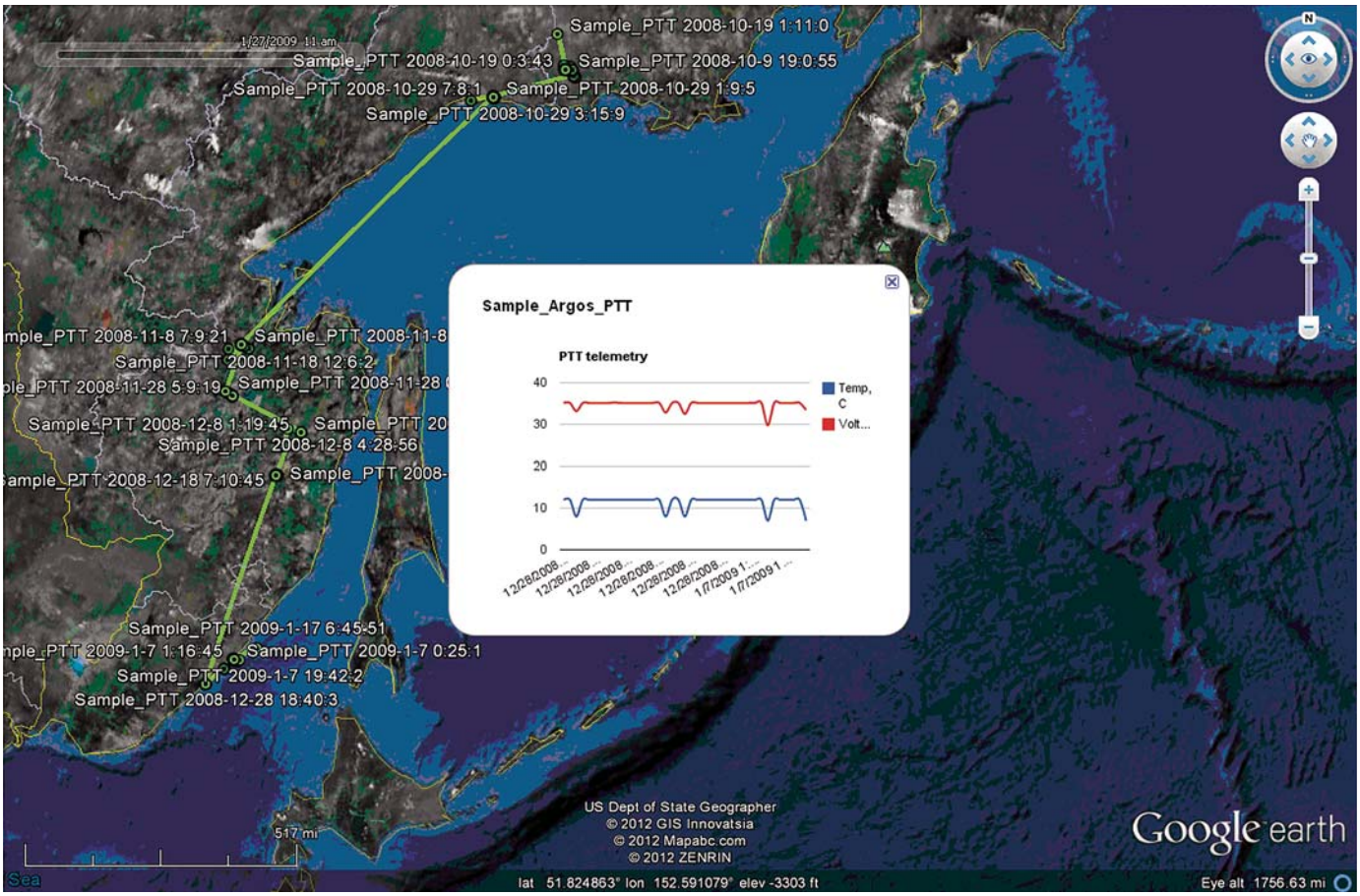


Figure 2. The trajectory of young Steller's Sea Eagle instrumented with Argos-based PTT migrating in the fall along coasts of Sea of Okhotsk, northeast Siberia. The telemetry data are decoded using spreadsheet, which reads DS data (http://goo.gl/r2OeG), the coordinates are decoded into dynamic KML using separate spreadsheet (http://goo.gl/gGVPU). Google-Earth file can be downloaded here (http://goo.gl/Kkg5d).

Such a service is exceptionally efficient when used on web-enabled phone devices as it does not involve downloading, importing, sorting, and displaying of all of the data, which would require additional computer resources and time. Once set up, the system runs unattended and is not susceptible to server shut-downs. Other advantages include free server access for data storage and accessibility from any internet connection.

CASE STUDIES

Case Study 1: GPS/GSM Biologging Collars on White-Tailed Deer

The white-tailed deer is widespread in the United States, especially in densely populated suburbs of the northeast where human/deer conflicts are common (DeNicola et al., 2000). We tracked the deer in a suburban area of Philadelphia in order to understand their spatial movements in areas of potential conflict. Tracking collars (Tellus, Followit, Sweden) transmitted a GPS location every five minutes. Once eight fixes accumulated in the collar, it transmitted the data via a cell phone network in form of a SMS text message. The text messages were then e-mailed to the end-user. The e-mail system (Gmail) has filters configured to send the e-mail messages to a blog which was set-up within the same Google account. The blog (http://goo.gl/0Kwi2) was configured to display the 15 latest messages on the front page and to archive older messages. Next, we used the import command within Google spreadsheets to read and process the data, and assemble the KML file (http://goo.gl/EnKCl). We automated our system so that each time a new e-mail is received by the Gmail account, the data appear on the blog, and this new content is read by the spreadsheet. Thus the generated KML file is continually updated with the latest data available for display in Google Earth and Google Maps (http://goo.gl/Nze33).

The high frequency of location fixes, and intuitive graphical representation of the resulting track (that can be viewed on web-enabled phones) accommodated highly sophisticated behavioral experiments. This system has been running successfully since 2008, and has been used to track 32 individual deer. A block diagram of the system is given in Figure 3 and the resulting map is given in Figure 4.

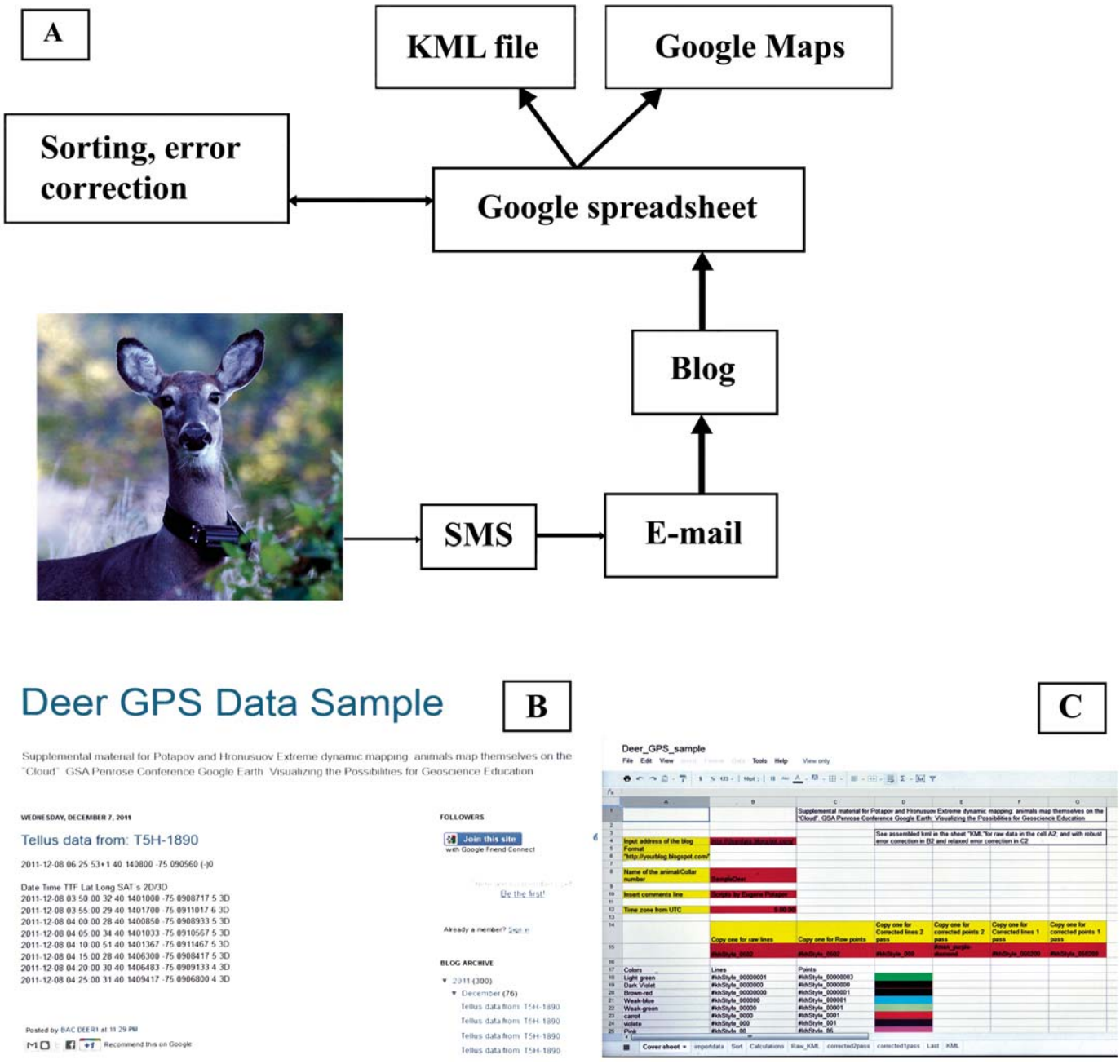


Figure 3. (A) Block diagram of the GPS/GSM-based system. (B) Screenshot of the blog which stores the data (http://goo.gl/0Kwi2). (C) Screenshot of the spreadsheet (http://goo.gl/EnKCl), which reads the blog, runs error correction, and dynamically assembles the KML file.

Case Study 2: Steller's Sea Eagles with Argos Doppler PTTs

The Steller's Sea Eagle is an endangered species, which breeds exclusively in northeast Siberia, Russia, and winters in the Russian far east, Kamchatka, and Japan. It is protected in both countries, nonetheless population numbers are on the decline. Since the breeding grounds are located in areas not yet affected by industrial society, the problems are thought to be concentrated within wintering areas and along migration routes (McGrady et al., 2003). To study the migration of the eagles, Argos-serviced PTTs were deployed. The Argos-based system is in principle no different from the GPS/GSM-based system presented above. Although the format of the Argos files is notoriously complicated, the files can be decoded with relative ease. The challenge is to decode observations that are presented in several lines.

The most widespread option of obtaining files from the Argos provider is via e-mail. The provider sends two data files to the end user: one in diagnostic format (DIA) and one in full format (DS).

Figure 4. Snapshot of the Google Earth page with the track of the animal shown in near-real-time. The dynamic KML can be downloaded here (http://goo.gl/Nze33).

The first one is best to use for extracting coordinates, location class (LC), and date and time for Doppler locations. The second one is needed to extract telemetry information. The system in place for the eagle project is based on a Google account. The Gmail account receives Argos-distributed e-mails and, using filters, forwards these to separate blogs: one for DS data (http://wp.me/2efLy) and another for DIA (http://wp.me/2egoY), a practice that is highly recommended. The length of the blog's front page depends on the number of PTTs in the program and the number of e-mail messages received from the Argos system per day. If the number of PTTs in a program is 10, it is sufficient to set the number of posts displayed on the front page at the value of 12.

In this example, two spreadsheets read the DS (http://goo.gl/r2OeG) and DIA (http://goo.gl/gGVPU) blogs (Fig. 1). Each filters only one PTT at a time using the command "IMPORTXML." In the next step, the spreadsheet splits the lines of the messages into meaningful data using the command:

```
= FILTER(INDIRECT("A1:A"&B1) ;
(left(INDIRECT("A1:A"&B1),5) = "YOUR PTT
NUMBER")+SIGN(IFERROR(FIND("
 "&ROW(INDIRECT("A1:A"&B1))&" " ; " "&CO
NCATENATE(ARRAYFORMULA(IF(left(INDIR
ECT("A1:A"&B1),5) = "YOUR PTT NUMBER"
; (ROW(INDIRECT("A1:A"&B1))+1)&" " ;
IFERROR(1/0)))));0)))
```

This complicated formula reads two consecutive lines that are relevant to the selected PTT. Processed data are sorted by date, and filtered for desired location class (LC) accuracy. Duplicates are removed and either a dynamic KML file (http://goo.gl/Kkg5d) or dynamic graph with telemetry parameters is assembled (Fig. 2). This system has been running successfully since 2008, and has been used to track six individual Steller's Sea Eagles and also ten Tundra Swans.

DISCUSSION

As with any of the systems built on the new cloud-based architecture, the suggested automatic mapping system has some constraints. If the data stream exceeds a certain level, the blog treats the dataflow as spam and automatically evokes the internet

bot protection called "CAPTCHA." CAPTCHA is usually a distorted image of letters and numbers, which is used to prevent automated use of blogs. It is based on the fact that a computer cannot recognize distorted letters and numbers, whereas humans can recognize them with ease. CAPTCHA activation requires human intervention, and thus can potentially lead to a loss of data messages. With a low frequency of messages this is not a problem. However, if the data provider sends hundreds of messages simultaneously (say after emergency reboot), this can potentially cause a data gap, which has to be filled by manually resending the messages. As a workaround we suggest users funnel the data into a backup blog for 24 hours, allowing the original blog to recover from the "data flood." The spreadsheet has to be reconfigured manually for a new blog address. Such "overflow" conditions occurred twice in three years of our continuous monitoring of the white-tailed deer.

Other problems are built into the Google spreadsheets. Currently there is a limit of 20,000 formulas per spreadsheet (with the current settings that means 20,000 cells, as every cell contains some sort of formula), which might be a potential limit for large Argos programs. For large data sets the problem can be solved by using the alternative service EditGrid.com, which offer similar spreadsheets, but without limitations on the number of formulas. There are some other alternatives, which are rapidly developing, e.g., dabbledb.com or zoho.com. The system that we outlined above suggests that the blog displays only the latest locations, which may be a constraint for some users. The length of the time period for these displayed fixes depends on the number of PTTs in the program and varies from 2 weeks to 1 month. Thus, it does not completely eliminate the necessity to download from the blog, which is capable of archiving all the data and processes the data in a conventional manner (GIS, filters, etc.). The speed of updates of the Google spreadsheets is currently once per hour or so, which is fast enough for both Argos-based satellite telemetry data and for GPS/GSM data.

Google spreadsheet-based maps also have limits on the number of points shown in Google Maps. Currently it is 100 elements. A larger number of points can make the map difficult to read. There are no such limits in Google Earth. A prominent advantage of the suggested system is that both the blog and the final spreadsheet-generated KML file are "web-ready" and can be used within various web clients. Although the blogs are potentially accessible to all Internet users the exact address is known only to informed users. Thus, the Internet address plays the role of a password. Of course, there is a potential for concern when displaying the current locations of endangered species on the web. One possible solution is to display only delayed locations. This is easy to achieve in the spreadsheet by simply unselecting the latest filtered position, and proved to be a very effective measure. In more than a year of continuous tracking of the white-tailed deer in southeastern Pennsylvania, including the legal hunting season, a 10 hour delay was implemented, and there were no attempts to harass the instrumented animals. For an endangered species, such as the Steller's Sea Eagle, a delay of a few days is considered sufficient.

Another important feature of our system is that it can generate output for web-enabled phones, and this near-real-time output can be brought to the field for experiments and rapid checks of the tagged animals without downloads and data processing. The near-real-time mapping is especially useful when used on mobile Internet devices (e.g., iPhones) as the latter allows researchers to get close to the tracked animal during field experiments. Once set-up, the system runs unattended, is unaffected by corporate-based or other server shutdowns, firewalls, etc., and is based on free servers which are accessible from any Internet connection.

ACKNOWLEDGMENTS

We thank Dr. M. McGrady of Natural Research, Ltd., Scotland and Dr. D. Rimlinger of San Diego Zoo for their support throughout this work, and Bryn Athyn College faculty for their help and encouragement. The method was conceived while one of the authors was chasing a Tundra Swan at the Colville River, Alaska, when working for the Swans Research Program funded by CONOCO Philips. We are grateful to Dr. W. Sladen and John Whissel, Ecological Studies at Airlie, Virginia, for supporting this study. We also thank André H. Banen (aka Ahab) for assistance with tuning Google spreadsheets and Dr. J. Bailey, Matt Heavner, and an anonymous reviewer for their suggestions to improve the manuscript, and the editorial team of this Special Paper for their hard work.

REFERENCES CITED

Argos User Manual, 2011, http://www.argos-system.org/files/pmedia/public/r363_9_argos_manual_en.pdf (accessed November 2011).

Clark, P.E., Johnson, D.E., Kniep, M.A., Jermann, P., Huttash, B., Wood, A., Johnson, M., McGillivan, C., and Titus, K., 2006, An advanced, low-cost, GPS-based animal tracking system: Rangeland Ecology and Management, v. 59, p. 334–340, doi:10.2111/05-162R.1.

DeNicola, A.J., VerCauteren, K.C., Curtis, P.D., and Hygnstrom, S.E., 2000, Managing white-tailed deer in suburban environments: A technical guide: Ithaca, New York, Cornell Cooperative Extension, Cornell University, 52 p.

Dettki, H., Ericsson, G., and Edenius, L., 2004, Real-time moose tracking: an internet based mapping application using GPS/GSM-collars: Alces (Thunder Bay, Ont.), v. 40, p. 13–21.

ESRI, 2004, ArcGIS Engine developer guide: Redlands, California, ESRI Press.

Fellbaum, C., ed., 1998, WordNet: An electronic lexical database: Cambridge, Massachusetts, MIT Press, http://wordnet.princeton.edu (accessed January, 2009).

Google spreadsheets function list, 2011, https://support.google.com/docs/bin/static.py?hl=en&topic=25273&page=table.cs (accessed November 2011).

McGrady, M., Ueta, M., Potapov, E.R., Utekhina, I.G., Masterov, V., Ladyguine, A., Zykov, V., Cibor, J., Fuller, M., and Seegar, W.S., 2003, Movements by juvenile and immature Steller's Sea Eagles *Haliaeetus pelagicus* tracked by satellite: The Ibis, v. 145, no. 2, p. 318–328, doi:10.1046/j.1474-919X.2003.00153.x.

Manuscript Accepted by the Society 16 April 2012