DATA SCIENCE JOURNAL

Ramapriyan, H K and Murphy, K J 2017 Collaborations and Partnerships in NASA's Earth Science Data Systems. *Data Science Journal,* 16: 51, pp. 1–7, DOI: https://doi.org/10.5334/dsj-2017-051

PRACTICE PAPER

Collaborations and Partnerships in NASA's Earth Science Data Systems

Hampapuram K. Ramapriyan[1,2] and Kevin J. Murphy[3]

[1] Science Systems and Applications, Inc., US

[2] NASA Goddard Space Flight Center, US

[3] NASA Headquarters, US

Corresponding author: Hampapuram K. Ramapriyan (Hampapuram.Ramapriyan@ssaihq.com)

NASA has been collecting Earth observation data from spaceborne instruments since 1960. Today, there are tens of satellites orbiting the Earth and collecting frequent global observations for the benefit of mankind. Collaboration between NASA and organizations in the US and other countries has been extremely important in maintaining the Earth observation capabilities as well as collecting, organizing and managing the data. These collaborations have occurred in the form of: 1. NASA's developing and launching spacecraft and instruments for operation by other agencies; 2. Instruments from collaborating organizations being flown on NASA satellites; and 3. Instruments from NASA being flown on satellites from collaborating organizations. In addition, there are collaborations such as joint science teams, data exchanges, and participation in international organizations to promote interoperability of various data systems. The purpose of this paper is to describe some of the Earth science data-related collaborative efforts in which NASA participates, and highlight a few results relevant to Earth system science research obtained through such collaborations.

Keywords: Data systems; Earth science; Collaboration; Interoperability; Standards; Remote sensing

Introduction

NASA has been collecting Earth observation data from spaceborne instruments since 1960 when the first experimental weather satellite, called the Television Infrared Observation Satellite (TIROS-1), was launched in collaboration with the Department of Defense. Today, there are tens of satellites orbiting the Earth and collecting frequent global observations for the benefit of mankind. In addition to satellite missions, airborne as well as land- and ocean-based instruments are contributing to the wealth of measurements of geophysical properties contributing to the understanding of the Earth as a system. Collaboration between NASA and organizations in the US and other countries has been extremely important in maintaining the Earth observation capabilities as well as collecting, organizing and managing the data. The purpose of this paper is to describe some of the collaborative efforts in which NASA participates, and highlight a few results pertinent to Earth system science research obtained through such collaborations. It is not meant to cover all of NASA's Earth science activities, but to provide an illustrative set of examples supported by the Earth Science Data Systems Program benefitting from collaborations. While there are other significant activities such as NASA Earth Exchange (NEX 2017), NASA Center for Climate Simulation (NCCS 2017) and NASA Modeling Guru (NASA 2017) that benefit from and support collaborations, an exhaustive listing or discussion of all of NASA's Earth science related data and computational capabilities is beyond the scope of this paper.

Interagency and International Collaborations

We observe that of the over 120 missions listed on the NASA web site on Earth science missions (http://science.nasa.gov/earth-science/missions/), approximately 37% involve interagency collaborations within the US (DoD, DoE, NOAA, and USGS), and 27% involve international collaborations (Argentina, Belgium, Brazil, Canada, ESA, EUMETSAT, Finland, France, Germany, India, Italy, Japan, Netherlands, Russia,

Switzerland, and the United Kingdom). These collaborations have occurred in the form of: 1. NASA's developing and launching spacecraft and instruments for operation by other agencies; 2. Instruments from collaborating organizations being flown on NASA satellites; and 3. Instruments from NASA being flown on satellites from collaborating organizations. In addition, there are collaborations in the form of joint science teams, data exchanges, aircraft flights for collecting Earth science data, ground based field campaigns including those for validating satellite-derived data products, participation in international organizations such as the International Council for Science (ICSU) World Data System (WDS), Committee on Earth Observing Satellites (CEOS) and the International Standards Organization (ISO).

NASA's Earth Science Data Systems (ESDS) Program

NASA's ESDS Program supports several activities, most of which involve collaborations in some form or another. The key "core" component of this program is the Earth Observing System Data and Information System (EOSDIS). The Earth Science Data and Information System (ESDIS) Project at NASA's Goddard Space Flight Center is responsible for the development, maintenance and operation of EOSDIS. The EOSDIS consists of 12 Distributed Active Archive Centers (DAACs) spread across the United States, and is a collaborative effort involving four NASA Centers, the US Geological Survey, the US Department of Energy and four universities (Ramapriyan *et al.* 2010). Each DAAC archives and distributes data products in a specialized set of Earth science disciplines. EOSDIS has been in operation since 1994, and has been providing data to the global user community following the free and open data policy that NASA put in place in 1990 with the start of the EOS Program. The data available from the EOSDIS DAACs cover all the collaborative missions indicated above and are accessible through http://earthdata.nasa.gov. The Earth Data Search Capability (EDSC) available at this website provides access to over 6,000 datasets managed by EOSDIS and a total of over 30,000 datasets that include those from several other organizations around the world. During the year ending in September 2016, EOSDIS distributed over 1.5 billion files of data (over 14.6 petabytes) to users all over the world.

In addition to EOSDIS, the ESDS Program supports three competed data system programs – Advancing Collaborative Connections for Earth System Science (ACCESS), Citizen Science for Earth Sciences (CSES) and Making Earth System Data Records for Use in Research Environments (MEaSUREs). The projects under these programs are selected by peer-review of proposals received in response to periodic calls as a part of NASA's Research Opportunities in Space and Earth Sciences (ROSES).

The purpose of the ACCESS Program is to "enhance, extend, and improve existing components of NASA's distributed and heterogeneous data and information systems infrastructure" through research by a number of Principal Investigators. This program has been in place since 2005 and has supported 73 different investigations through calls for proposals approximately every two years. NASA ACCESS (2017) shows a complete list of these investigations with descriptions. Several of the technologies developed in the ACCESS Program may be deployed within EOSDIS as they are matured. An example of such a development is Geospatial Interactive Online Visualization ANd aNalysis Infrastructure (Giovanni), which is a web-based tool for accessing, visualizing and analyzing large quantities of Earth science remote sensing data without having to download them first (Berrick *et al.* 2009).

The purpose of the CSES Program is to develop and implement capabilities to harness voluntary contributions from members of the general public to advance understanding of the Earth as a system. This program complements NASA's capability of observing the Earth globally from space, air, land, and water by engaging the public in NASA's mission. This program was initiated in 2017 and currently supports 16 investigations prior to a selection of a subset of them for the implementation phase. This program is closely aligned with NASA's Global Learning and Observations to Benefit the Environment (GLOBE) Program which has been in place for over 20 years engaging students, scientists, and teachers in 114 countries in the scientific exploration of Earth's environments and climate (NASA GLOBE 2017).

The purpose of the MEaSUREs Program is to produce long-term, consistent "Earth System Data Records (ESDRs)" defined as "a unified and coherent set of observations of a given parameter of the Earth system, which is optimized to meet specific requirements in addressing science questions". Climate Data Records (CDRs) are a particular case of ESDRs. Following the first call for proposals from the MEaSUREs Program in 2006, 30 projects were selected. Currently there are 27 active projects under the MEaSUREs Program, which were selected following the second call in 2012. NASA MEaSUREs (2017) provides a complete list of the MEaSUREs projects and their descriptions. The input data used for generating the ESDRs in most of these projects are obtained from multiple satellites under international collaborative arrangements. **Table 1** shows the numbers of MEaSUREs projects using inputs from various countries. The designation "International" in

Table 1: Number of MEaSUREs projects using input data from various countries or international organizations.

Source of Input Data	Number of Projects
International	15
Australia	1
Brazil	1
Canada	4
China	1
France	4
Germany	4
India	1
Japan	11
Mexico	1
Netherlands	2
Russia	1
Switzerland	1
UK	2
USA	7

this figure indicates either activities involving multiple countries such as the International Satellite Cloud Climatology Project (ISCCP) or organizations such as the European Space Agency (ESA).

The ESDRs generated by the MEaSUREs Principal Investigators are archived and distributed by the DAACs in their respective disciplines. Many of the ESDRs have been included in the Essential Climate Variables Inventory (WGClimate 2017) maintained by the joint Committee on Earth Observation Satellites (CEOS), Coordination Group for Meteorological Satellites (CGMS) and Working Group on Climate (WGClimate).

The collaboration between the core and competed components of the ESDS Program occurs through the Earth Science Data System Working Groups (ESDSWG). Individual working groups within the umbrella of ESDSWG are formed to address key technology and information system issues. They provide recommendations to the ESDIS Project regarding specific capabilities of data and information systems for the benefit of the community. They address various challenges, issues, and opportunities in response to data system priorities and community-identified needs. They also develop guidelines and best practices that address the practical implementation of standards and technologies, enhance data interoperability, and improve software development and software architecture practices. Examples of such working groups are: metadata for airborne investigations, atmospheric science users' forum, cloud computing, dataset interoperability, data preservation, data quality, data visualization, geospatial web services, search relevance, data citations, software citations and time series.

Members of the ESDIS Project and the DAACs take an active role in the Earth Science Information Partners (ESIP), which is an organization with over 180 members from several government agencies, universities, non-profit organizations and commercial entities. The mission of ESIP is "to support the networking and data dissemination needs of its members and the global Earth science data community by linking the functional sectors of observation, research, application, education and use of Earth science". Within ESIP, collaboration areas are formed to address technical and other issues of common interest. There is some commonality and overlapped membership between NASA's ESDSWG and the ESIP collaboration areas, even though their functions and scopes are different. There is a two-way exchange of information on best practices in various technical areas for the benefit of all the participating organizations.

NASA's ESDS Program takes an active role in interagency activities such as the US Global Change Research Program, US Group on Earth Observations, Big Earth Data Initiative, and the Networking and Information Technology Research and Development (NITRD) Program.

Also, internationally, the ESDS Program is active in the Committee on Earth Observing Satellites (CEOS) Working Group on Information Systems and Services (WGISS), the Group on Earth Observations (GEO) Data Sharing Working Group, the International Standards Organization (ISO), and the Open Geospatial Consortium (OGC). The ESDIS Project is a Network Member of the ICSU/WDS, and ten of the 12 DAACs are Regular Members of the ICSU/WDS. All these collaborations promote increased interoperability of various data systems, increased utility of data held by many organizations, and adherence to standards.

Examples of Datasets Resulting from Collaborations

Among the many benefits of collaborations are datasets useful for scientific research and applications. As indicated above there are several thousand datasets openly accessible through http://earthdata.nasa.gov. In this section we briefly describe a few of the datasets, which have resulted from interagency/international collaborations, and which cover several Earth science disciplines. A few examples of MEaSUREs ESDRs are included, as well as one from the EOS Terra mission, resulting from a collaboration between Japan and US.

MEaSUREs ESDRs

Rignot *et al.* (2011) have developed an ESDR called "MEaSUREs Antarctic Grounding Line from Differential Satellite Radar Interferometry". A grounding line is "the transition boundary where ice detaches from the bed and becomes afloat in the ocean". It is "critical to ice sheet mass budget calculations, numerical modeling of ice sheet dynamics, ice-ocean interactions, oceanic tides, and subglacial environments." This ESDR uses data from the Earth Remote Sensing Satellites 1 and 2 (ERS-1 and ERS-2), RADARSAT-1, RADARSAT-2, and the Advanced Land Observing System Phased Array type L-band Synthetic Aperture Radar (ALOS PALSAR) for years 1994 to 2009.

Rignot *et al.* (2012) have developed an ESDR called "MEaSUREs InSAR-Based Ice Velocity Maps of Central Antarctica: 1997 and 2009". Information about ice velocity is essential for estimating mass balance of glaciers and ice sheets and in studies of ice dynamics. This ESDR consists of two high-resolution digital mosaics of ice motion in Central Antarctica, assembled from satellite interferometric synthetic-aperture radar (InSAR) data acquired by RADARSAT-1 in 1997 and by RADARSAT-2 in 2009.

Joughin *et al.* (2015) have developed an ESDR called "MEaSUREs Greenland Ice Sheet Velocity Map from InSAR Data". It contains ice-sheet-wide velocity maps covering much of the Greenland ice sheet for winters of 2000/2001 and 2005/2006. The ESDR is derived from InSAR data obtained by RADARSAT-1, the Advanced Land Observation Satellite (ALOS), and the TerraSAR-X satellite. According to Joughin *et al.* (2010), "Comprehensive mappings such as these, at regular intervals, provide an important new observational capability for understanding ice-sheet variability."

Using data from TOPEX/Poseidon, Jason-1 and OSTM/Jason-2, Beckley *et al.* (2015) have developed global mean sea-level trends for the period covering September 1992 to present. These data are updated with a time lag of up to 4 months. Accurate computation of mean sea-level is critical to understand climate change effects, and determination of the orbital positions of the satellites precisely is essential to achieve the necessary accuracy. Precision orbit determination in turn requires many globally distributed, international resources including the Global Navigation Satellite Systems (GNSS).

Kwok has generated ESDRs with measurements of ice motion, deformation, ice age and thickness, and backscatter histogram. The ESDRs cover the period of fall 1997 to May 2008 using RADARSAT-1 data from the Canadian Space Agency and are available from the Alaska Satellite Facility DAAC. The products cover the Arctic basin, with the most consistent coverage over the Western Arctic Basin. (Kwok 2016).

The JPL Multiscale Ultrahigh Resolution (MUR) MEaSUREs Project, led by T. Chin generates level-4 sea surface temperature analysis with four-day latency and also a near-real-time dataset with one-day latency. The current version (version 4) of the MUR analysis is based on nighttime Group for High-Resolution Sea Surface Temperature (GHRSST) L2P skin and sub-skin SST observations from several instruments (AMSR-E from Japan on NASA's EOS Aqua satellite, Moderate Resolution Imaging Spectroradiometer (MODIS) on NASA's EOS Terra and Aqua satellites, US Navy microwave WindSat radiometer, Advanced Very High Resolution Radiometer (AVHRR) on several NOAA satellites, and in situ SST observations from the NOAA iQuam project. (JPL MUR MEaSUREs Project, 2015).

Kim *et al.* (2013) provide a global record of the daily freeze/thaw (FT) status of land areas derived from satellite observations of radiometric brightness temperatures. Two records are provided: (1) Scanning Multichannel Microwave Radiometer (SMMR) and Special Sensor Microwave/Imager (SSM/I) record for the years 1979 to 2010 and (2) Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) record for the years 2002 to 2011. "The landscape freeze–thaw (FT) signal from satellite microwave remote

sensing is closely linked to vegetation phenology and land–atmosphere trace gas exchange where seasonal frozen temperatures are a major constraint to plant growth." (Kim *et al.* 2012).

In a project called the "Global Ozone Chemistry and Related trace gas Data records for the Stratosphere (GOZCARDS)", Froidevaux *et al.* (2015) and Wang *et al.* (2013) have focused on hydrogen chloride (HCl), water vapor (H_2O), and ozone (O_3). The resulting dataset is a long-term global ESDR of monthly zonal mean time series starting in 1979. This dataset is of high relevance to studies of ozone decline and recovery, for understanding changes in atmospheric composition and for constraining model representations of atmospheric dynamics and photochemistry. Data from instruments on-board several NASA satellites and the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) on the Canadian satellite SCISAT are used as inputs.

ASTER GDEM

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) is another example of a very useful dataset resulting from international collaboration. The release of Version 2 of this dataset (US/Japan ASTER Science Team 2011) was announced jointly by the Ministry of Economy, Trade, and Industry (METI) of Japan and NASA in October 2011. This dataset is a contribution from METI and NASA to the Global Earth Observation System of Systems (GEOSS).

Summary

NASA has been collecting Earth science data using observations from spaceborne and airborne instruments as well as through field campaigns for several decades. The observing missions as well as the data system activities needed for gathering, processing, archiving and distributing the data and derived products require collaborations. NASA collaborates with universities, commercial entities, agencies within the US as well as other countries and international organizations. This paper has shown several activities benefiting from such collaborations and illustrated some scientific products useful for Earth science research that have resulted from collaborative sharing of data.

Acknowledgements

Hampapuram Ramapriyan was supported by NASA contract NNG15HQ01C with Science Systems and Applications, Inc. Kevin Murphy contributed to this paper as a part of his duties as an employee of NASA.

Competing Interests

The authors have no competing interests to declare.

Author Information

Hampapuram Ramapriyan is a Research Scientist/Subject Matter Expert at Science Systems and Applications, Incorporated (SSAI). He supports the Earth Science Data and Information System (ESDIS) Project at NASA GSFC through its contract with SSAI. This project is responsible for archiving and distributing most of NASA's Earth science data using the Earth Observing System Data and Information System (EOSDIS), which is a component of NASA's Earth Science Data System Program. Ramapriyan's primary focus is on data stewardship and preservation. Prior to his employment with SSAI, he was the Assistant Project Manager of the ESDIS Project. His responsibilities included management of Science Investigator-led Processing Systems that processed and delivered data to the EOSDIS Distributed Active Archive Centers (DAACs) as well as being the data liaison between the ESDIS Project and the Principal Investigators of the projects under the MEaSUREs (Making Earth System Data records for Use in Research Environments) Program. He is the co-chair of the Data Quality Working Group, which is one of NASA's Earth Science Data System Working Groups. As an active member of the Federation of Earth Science Information Partners (ESIP) since 1998, he is currently a member of its Data Stewardship Committee and chairs the Information Quality Cluster.

Kevin Murphy is the Program Executive for Earth Science Data Systems at NASA HQ. In this capacity Mr. Murphy manages a portfolio of programs encompassing the Distributed Active Archive Centers, Science Investigator-led Processing Systems and a number of competitively funded programs. Prior to assuming his current role Mr. Murphy served as System Architect for EOSDIS, conceived, developed and managed major system development projects including near real-time systems, search engines, a large scale visualization system and Earthdata.nasa.gov. Murphy has received numerous awards during his NASA career, including the NASA Exceptional Achievement Medal, Robert H. Goddard Exceptional Achievement for Engineering, and Charles S. Falkenberg Award, among others.

References

Beckley, B, Zelensky, N P, Holmes, S A, Lemoine, F G, Ray, R D, Mitchum, G T, Desai, S and **Brown, S T** 2015 *Global Mean Sea Level Trend from Integrated Multi-Mission Ocean Altimeters TOPEX/Poseidon Jason-1 and OSTM/Jason-2 Version 3*. PO.DAAC, CA, USA. http://podaac.jpl.nasa.gov/dataset/MERGED_TP_J1_OSTM_OST_GMSL_ASCII_V3.

Berrick, S, Leptoukh, G, Farley, J and **Rui, H** 2009 Giovanni: A Web services workflow-based data visualization and analysis system. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1): 106–113. DOI: https://doi.org/10.1109/TGRS.2008.2003183

Froidevaux, L, Anderson, J, Wang, H-J, Fuller, R A, Schwartz, M J, Santee, M L, Livesey, N J, Pumphrey, H C, Bernath, P F, Russell, J M, III and **McCormick, M P** 2015 Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS): methodology and sample results with a focus on HCl, H_2O, and O_3. *Atmospheric Chemistry and Physics*, 15(18): 10471–10507. DOI: https://doi.org/10.5194/acp-15-10471-2015

Joughin, I, Smith, B, Howat, I and **Scambos, T** 2015 *MEaSUREs Greenland Ice Sheet Velocity Map from InSAR Data, Version 2*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. DOI: https://doi.org/10.5067/OC7B04ZM9G6Q

Joughin, I, Smith, B, Howat, I, Scambos, T and **Moon, T** 2010 Greenland Flow Variability from Ice-Sheet-Wide Velocity Mapping. *Journal of Glaciology*, 56(197): 415–430. DOI: https://doi.org/10.3189/002214310792447734

JPL MUR MEaSUREs Project 2015 GHRSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis (v4.1). Ver. 4.1. PO.DAAC, CA, USA. DOI: https://doi.org/10.5067/GHGMR-4FJ04

Kim, Y, Kimball, J S, Glassy, J and **McDonald, K C** 2013 *MEaSUREs Global Record of Daily Landscape Freeze/Thaw Status, Version 2*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. DOI: https://doi.org/10.5067/MEASURES/CRYOSPHERE/nsidc-0477.002

Kim, Y, Kimball, J S, Zhang, K and **McDonald, K C** 2012 Satellite detection of increasing Northern Hemisphere non-frozen seasons from 1979 to 2008: Implications for regional vegetation growth. *Remote Sensing of Environment*, 121: 472–487. DOI: https://doi.org/10.1016/j.rse.2012.02.014

Kwok, R 2016 *Lagrangian Sea-Ice Kinematics*. Available from Alaska Satellite Facilityt DAAC. DOI: https://doi.org/10.5067/SSMPINYI15UU

NASA 2017 *NASA Modeling Guru*. https://modelingguru.nasa.gov/index.jspa (Last accessed 8 August 2017).

NASA ACCESS 2017 *Advancing Collaborative Connections for Earth System Science (ACCESS) Projects*. https://earthdata.nasa.gov/community/community-data-system-programs/access-projects (Last accessed 14 August 2017).

NASA GLOBE 2017 *The Global Learning and Observations to Benefit the Environment (GLOBE) Program*. https://www.globe.gov/about/overview (Last accessed 14 August 2017).

NASA MEaSUREs 2017 *Making Earth System Data Records for Use in Research Environments (MEaSUREs) Projects*. https://earthdata.nasa.gov/community/community-data-system-programs/measures-projects (Last accessed 14 August 2017).

NCCS 2017 NASA Center for Climate Simulation. NASA Goddard Space Flight Center. https://www.nccs.nasa.gov/ (Last accessed 8 August 2017).

NEX 2017 NASA Earth Exchange. NASA Ames Research Center. https://nex.nasa.gov/nex/ (Last accessed 8 August 2017).

Ramapriyan, H K, Behnke, J, Sofinowski, E, Lowe, D and **Esfandiari, M** 2010 "Evolution of the Earth Observing System (EOS) Data and Information System (EOSDIS)". Chapter 5, *Standard-Based Data and Information Systems for Earth Observation*, *Springer Series: Lecture Notes in Geo-Information and Cartography*, Di, L and Ramapriyan, H K (eds.). DOI: https://doi.org/10.1007/978-3-540-88264-0

Rignot, E, Mouginot, J and **Scheuchl, B** 2011 *MEaSUREs Antarctic Grounding Line from Differential Satellite Radar Interferometry, Version 2*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. DOI: https://doi.org/10.5067/MEASURES/CRYOSPHERE/nsidc-0498.001.

Rignot, E, Mouginot, J and **Scheuchl, B** 2012 *MEaSUREs InSAR-Based Ice Velocity Maps of Central Antarctica: 1997 and 2009, Version 1*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. DOI: https://doi.org/10.5067/MEASURES/CRYOSPHERE/nsidc-0525.001.

U.S./Japan ASTER Science Team 2011 *ASTGTM: ASTER Global Digital Elevation Model V002*. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South

Dakota (https://lpdaac.usgs.gov). (ASTER GDEM is a product of NASA and METI). DOI: https://doi.org/10.5067/ASTER/ASTGTM.002

Wang, R, Froidevaux, L, Anderson, J, Fuller, R A, Bernath, P F, McCormick, M P, Livesey, N J, Russell, J M, III, Walker, K A and **Zawodny, J M** 2013 *GOZCARDS Merged Ozone 1 month L3 10 degree Zonal Means on a Vertical Pressure Grid V1*. Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). DOI: https://doi.org/10.5067/MEASURES/GOZCARDS/DATA3006.

WGClimate 2017 *Climate Monitoring from Space*. http://climatemonitoring.info/ (Last accessed 14 August 2017).

How to cite this article: Ramapriyan, H K and Murphy, K J 2017 Collaborations and Partnerships in NASA's Earth Science Data Systems. *Data Science Journal,* 16: 51, pp. 1–7, DOI: https://doi.org/10.5334/dsj-2017-051

Submitted: 31 October 2016 **Accepted:** 04 September 2017 **Published:** 13 November 2017

]u[*Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS