Article

# EMERGENT CHALLENGES FOR SCIENCE SUAS DATA MANGEMENT: FAIRNESS THROUGH COMMUNITY ENGAGEMENT AND BEST PRACTICES DEVELOPMENT

**Jane Wyngaard [1,†]*, Lindsay Barbieri [2,†], and Andrea Thomer [3,†]**

[1] University of Notre Dame; jwyngaar@nd.edu [2] University of Vermont; lkbar@uvm.edu [3] University of Michican; athomer@umich.edu

* Correspondence: jwyngaar@nd.edu; Tel.: +1-574-631-5163

† These authors contributed equally to this work.

Version April 10, 2019 submitted to Remote Sens.

**Abstract:** The use of small Unmanned Aircraft Systems (sUAS ) as platforms for data capture has rapidly increased in recent years. However, while there has been significant investment in improving the aircraft, sensors, operations, and legislation infrastructure for such, little attention has been paid to supporting the managment of the complex data capture pipeline sUAS involve. The following outlines a community engagement based investigation into what tools, practices, and challenges currently exist for particularly researchers using sUAS as data capture platforms. The key results of this effort are (1) a representation of the key characteristics sUAS captured data both share and have uniquely when compared to traditional remote sensing data, and (2) based on these characterists and community input we define 8 challenges that need to be addressed in order for the full value of sUAS captured data to be realised. We conclude that it is worth while for the community to address these challenges given current industry trends, the potential increase in value of sUAS captured data such would enable, and the anticipated immediate and future costs of not doing so.

**Keywords:** sUAS; drone; RPAS; UAV; Data; Management; FAIR; Community; standards; practices)

---

## 1. Introduction

Small Unmanned Aircraft Systems sUAS — also known as Remotely Piloted Aircraft Systems (RPAS), Unmanned Aerial Vehicles (UAV), or often colloquially as 'drones' — are rapidly becoming a ubiquitous tool for data collection across a wide range of private and public applications worldwide. This includes multiple academic fields (electrical, chamical, and civil engineering; multiple environmental sciences; and others) for which sUAS are changing how and which data are captured. While this new platform shares much with traditional remote sensing and a range of other sensor systems, the particular combination of spatiotemporal resolutions, operational practices, and wide spectrum of heterogeneous data being collected with sUAS has lead to a unique set of data management challenges. Additionally, various global efforts and technological advances in the sphere of data management are opening unique opportunities and potential for sUAS as a nascent technology for environmental sensing.

This paper compiles 4 years of extensive community engagement around the complexities, nuances, and importance of sUAS data management; and seeks to lay the motivations and foundations for future global sUAS user community engagements.We do so by: (1) outlining the potential value gains of normalising good data management practices for sUAS collected research data, (2) detailing the unique complexities of sUAS data while pointing to analogous sectors and existing resources that might be leveraged, and (3) map out the key challenges and needs - identified by the community - as necessary to realising the full value potential for sUAS data. Henceforth we will use "sUAS data" to refer to the primary research data captured on-board sUAS , rather than just data relating to the sUAS platform itself. In many cases the former requires and therefore includes the latter.

To provide context for later sections, the remainder of this section outlines the current state of sUAS use in academia and the corresponding state of sUAS data management. Following which, Section 2 details the authors' engagement with the global community on this topic. Section 2 aims to: summarise what methods of community engagement were undertaken including detailing which geographical regions and domains of expertise were included; and to highlight others working in this space and the resources that are currently available through such. Drawing on the outcomes of this engagement, Section 3 presents the core characteristics of sUAS captured data which are behind the need for sUAS specific data management practices and infrastructure. Finally, Section 4 discusses the community distilled key challenges arising from Section 2and 3, before Section 5 concludes.

### 1.1. Current use of sUAS in research

The rapid adoption of sUAS for scientific data collection has been driven largely by the flexible functionality now possible due to key technological advances: lowered hardware costs, increased battery energy density, wide spread sensor miniaturisation, and the availability of sophisticated autopilot hardware and software. Lagging but globally following these technological advances, have been new national and international aviation regulations [1]. Collectively, the above mean it is now possible and highly attractive for even small and modestly funded research teams to incorporate sUAS data into their investigations.

As platforms for scientific data collection, sUAS offer several functional advantages when compared with many traditional methods: (i) the ability to collect higher spatial and or temporal resolution data; (ii) a reduced impact on sensitive environments being monitored; (iii) lowered risks to workers and equipment involved in data collection in dangerous environments; (iv) a highly flexible platform from which an extremely wide range of parameters might be monitored simultaneously, and (v) access to many data that what would otherwise be practically inaccessible, all (vi) often at a significantly lower cost than traditional methods might incur [2–4]. sUAS data-sets are therefore generally parameter rich and uniquely high resolution data-sets, that consequently potentially offer unique and novel reuse value across multiple academic, commercial, governmental, and non profit use cases.

The value of these advantages to primary data users is clearly evident in the number and domain variety of recently published peer-reviewed articles that include various terms for sUAS (see Figure 1 and 2). And this growth is more than matched by the commercial sUAS sector, with some forecasts estimating a market value of USD 100 Billion in the next few years [5–8]. This non-research market is driving the rapid advancement of sUAS : flight platforms, sensor miniaturisation, wireless telemetry, sophisticated autonomous navigation, operations, and legislation; to meet the needs of commercial sUAS use in: Agriculture, Mining, Civil Engineering and Infrastructure, Search and Rescue and Disaster Responses, Cargo and Data Delivery, Conservation, Entertainment, and many more use cases.

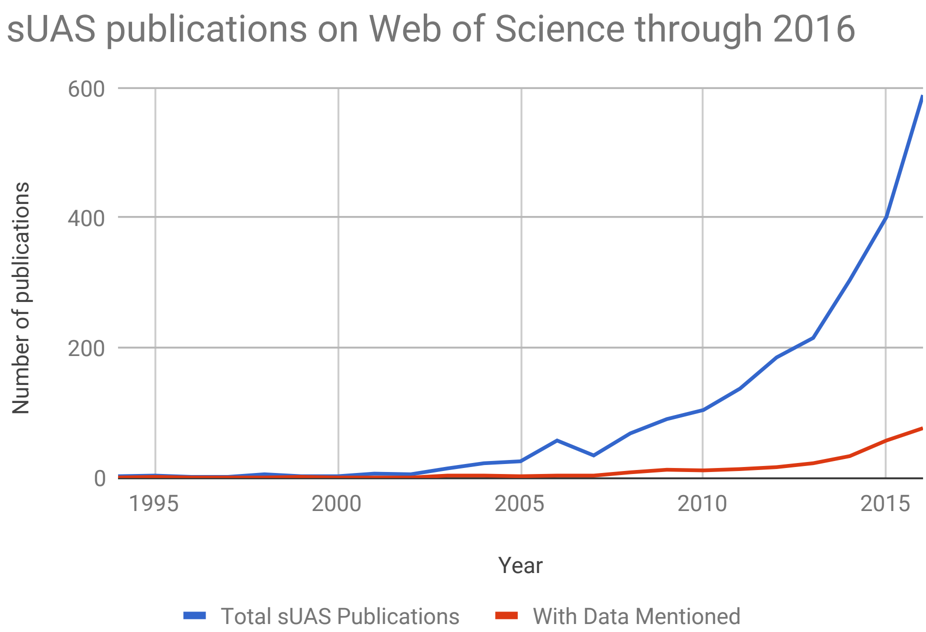


**Figure 1.** sUAS publications in web of science through 2018: As sUAS have become more prevalent as platforms for scientific data collection, there has been a corresponding increase in their prominence within the academic peer-reviewed literature. This chart shows this growth in blue, with the number of publications found in a Web of Science literature search on the topic of sUAS . By comparison, the number of sUAS publications that also referenced the management of data, is shown in orange at a much lower rate. Graph generation including the list of search terms are included in footnote [1]

[1] Bar: add search information when updated to 2019

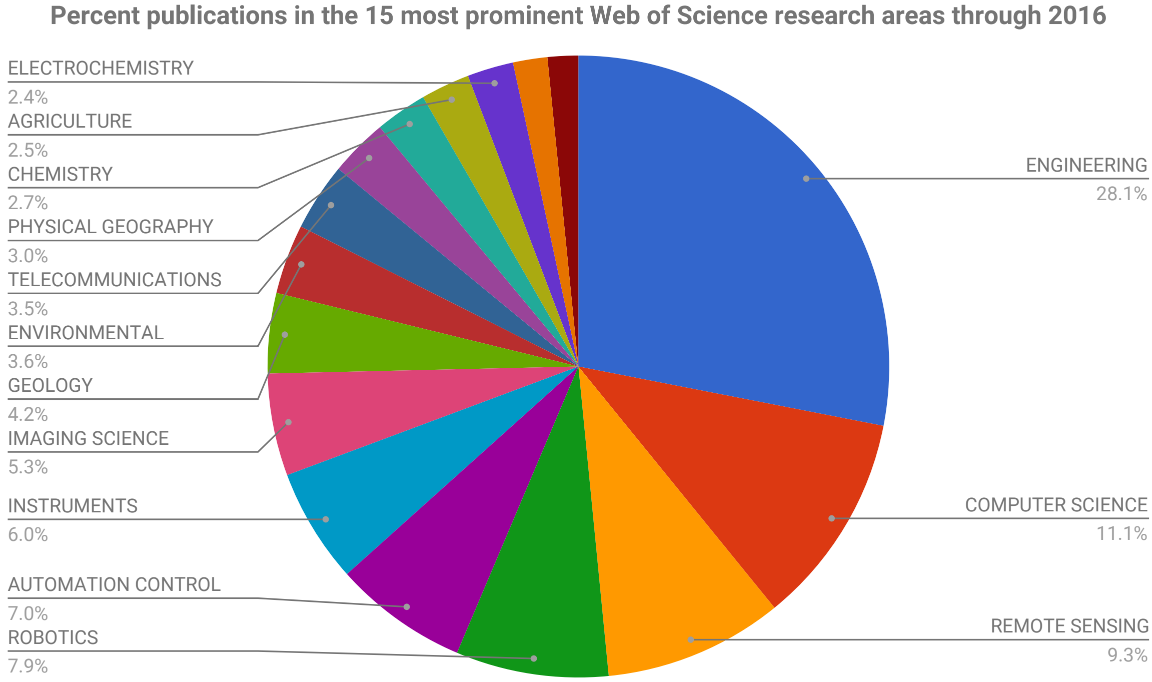


**Figure 2.** Percent publications in the 15 most prominent Web of Science research areas through 2016: Based on the same search terms as Figure 1, this categorises the total number of publications returned into the academic field under which they were published through 2016. Graph generation including the list of search terms are included in footnote [2]

These advances are being made across and through novel and developing commercialisation models as the industry evolves. As a result commercial models include the sale of both products and services, and both proprietary and well matured and diverse open source sUAS solutions. For researchers, each of these models offers a variety of value trade-offs; between fully customisable and scientific purpose built solutions with full access to all metadata at the cost of their own development effort and time, through to less configurable but ready to fly platforms, or even full data capture and analytics services. The latter generally also involves higher but arguably justifiable monetary costs, and provide less contextual information with their data but are easily and rapidly deployed.

### *1.2. Current sUAS data management*

Research data management infrastructure and procedures have always been important but have become more complex and costly as the quantity of available data has significantly increased [9,10]. Why and how sUAS data management is critical to realising its full value is an outcome of this engagment discussed in more detail in Section 3, however, sUAS users who have attempted to publish their data are familiar with why it is also particularly challenging. A typical sUAS -based project, for instance, will involve multiple stakeholders (e.g. scientists, engineer, pilot), technologies (e.g. sUAS , controllers, computers, software systems, sensors, paper notebooks), parameters (e.g. flight platform attitude, scientific sensor calibration date and processes, scientific parameters, comensual environmental conditions), and complex processes (e.g. data triage, data compression, data pre- and post- processing), many of which can impact a data-sets interpretation. Capturing multiple of these disparate components is commonly necessary for initial data product generation and interpretation,

[2] Bar: add search information when updated to 2019

many will be required for data publication, and an even larger superset would be required by a user seeking to reuse the data in a future investigation. Unfortunately, as indicated scholarly and scientific sUAS users represent a relatively small user market with niche needs. Consequently the challenge of sUAS data management have not yet been widely addressed either by industry stakeholders or researchers who are largely still exploring sUAS capabilities and potential value.

As a result, individual researchers out of necessity are developing their own ad hoc data management strategies. However, this is problematic in the long term for multiple reasons. First, this substantially adds to the learning curve of sUAS technologies: new-to-sUAS researchers must already navigate complex legal, technical, and institutional spaces, and developing a data management strategy from scratch further increases the required overhead. For researchers specifically seeking to take advantage of sUAS as a new and otherwise more affordable means of data capture, the economic and time costs of developing robust data management workflows can be prohibitive.

Second, the repeated reinvention of ad hoc data management workflows represents a significant amount of effort. Not only is this an inefficient use of finite research resources, but these idiosyncratic workflows pose a roadblock to the development of common tools and workflows. However, lacking a collective articulation of data management requirements, there is no alternative even for those motivated to collaborate on the development of common better commercial and open source software and tools meaning the roadblock will only continue to grow.

Third and finally, the lack of common data practices risks lowering the trustworthiness and reproducibility of Scientific research based on sUAS data. Without shared data practices and methods of documenting workflows, sUAS data based research is often plagued by poor or heterogeneous documentation, unknown or unstandardized quality control methods, and methodological uncertainty. The current opacity of sUAS data workflows makes thorough peer review extreemly difficult.

### *1.3. Opportunities for sUAS data management*

The described landscape presents a problematic picture, yet the rapid growth of sUAS as a revolutionary sensor platform across multiple sectors has arrived at a highly opportune moment. Key developments and shifts in social, political, and particularly academic attitudes worldwide present a unique opportunity to the sUAS user community - one that has not been available for many other research technologies. Specifically, the coincidence of the following present an opportunity: (i) the push for open science and FAIR (Findable Accessible Interoperable Reusable) [9] data, (ii) the corresponding maturing of data technologies, and (iii) the lack of momentum behind any substandard normalised practices and the minimal amount of legacy sUAS data currently available that would otherwise require significant effort to migrate or reprocess. The following elaborates on each of these.

#### 1.3.1. The push for Open Science and FAIRness

At the same time as sUAS are emerging as a standard tool for researchers, the broader research community is building momentum in actively moving towards normalising open science and FAIR data practices. This is evidenced by the wealth of work calling for better research practices [11–14]; the numerous calls for improving reproducibility and cross disciplinary data use through better practices[15–18]; and the many non-academic calls for data sharing from a range of government bodies [19–21]. The significant traction that the FAIR nomenclature has gained - as a succinct framing of core good data management practices - demonstrates this momentum further [22–24].

#### 1.3.2. The Corresponding Maturation of Data Technologies

As industry has moved to extract economic advantages from Big Data, the technologies required to manage, manipulate, and mine value in large and heterogeneous, data-sets of mixed quality have significantly matured [25–27]. The breadth of associated tools available is extremely wide but a few high visibility relevant examples include; the growth in capabilities and use of cloud resources[28–31], Google's beta Data-set search engine [32] and the required enabling data-set schema, the international Earth and space science community's effort to develop standards that will connect researchers, publishers and data repositories[33,34], and the increase in efforts to utilise Machine Learning tools on classical Big Data for a multitude of applications including the Geo-sciences[35,36].

#### 1.3.3. The lack of norms or legacy sUAS data

The lack of community accepted best practices for sUAS data management is both a challenge and an opportunity. As a new technology, researchers are still grappling with how best to use sUAS . This provides a window of opportunity within which: (a) with little past effort to be discarded on previously used methods, the cost of adopting new practices is minimal, and (b) with the net quantity of sUAS captured scientific data still relatively small, the cost of adopting new formats, metadata standards, calibration methods — and all of the other crucial components of data archive — will not be significantly added to by the need for backwards compatibility or mass re-ingestion and processing of previously captured data. This window, however, is closing rapidly, as researchers globally — out of necessity — are creating all of these components for themselves in ad hoc and isolated manners, and rapidly accumulating data.

## 2. Materials and Methods

In light of the above, over the past 4 years (2015 - 2019), the authors have pursued a wide-ranging, largely volunteer based, effort, to engage with the nascent community of researchers using sUAS for data collection on challenges of data management. To begin with this involved looking to both the emerging sUAS science community and to the many mature analogous domains for applicable best practices, and included considering standards and conventions used by large scale government and research institutions using both sUAS and more traditional remote sensing technologies. This was followed by running multiple workshops and conference sessions with the aim of identifying key needs and available resources for sUAS data management. The progression of core engagement meetings involved are shown in 3, at each of which we sought input from both academic and commercial sUAS users, suppliers, and developers and data management professionals. This process was not planned out originally, or for the most part directly funded, but driven by researcher needs, the perceived value opportunity, community request, and by which doors opened when knocked on. In the following sections the key threads of this engagement are summarised and groups are highlighted for the purposes of directing interested parties to possible resources or potential starting points for future efforts.

## 2.1. Community engagement

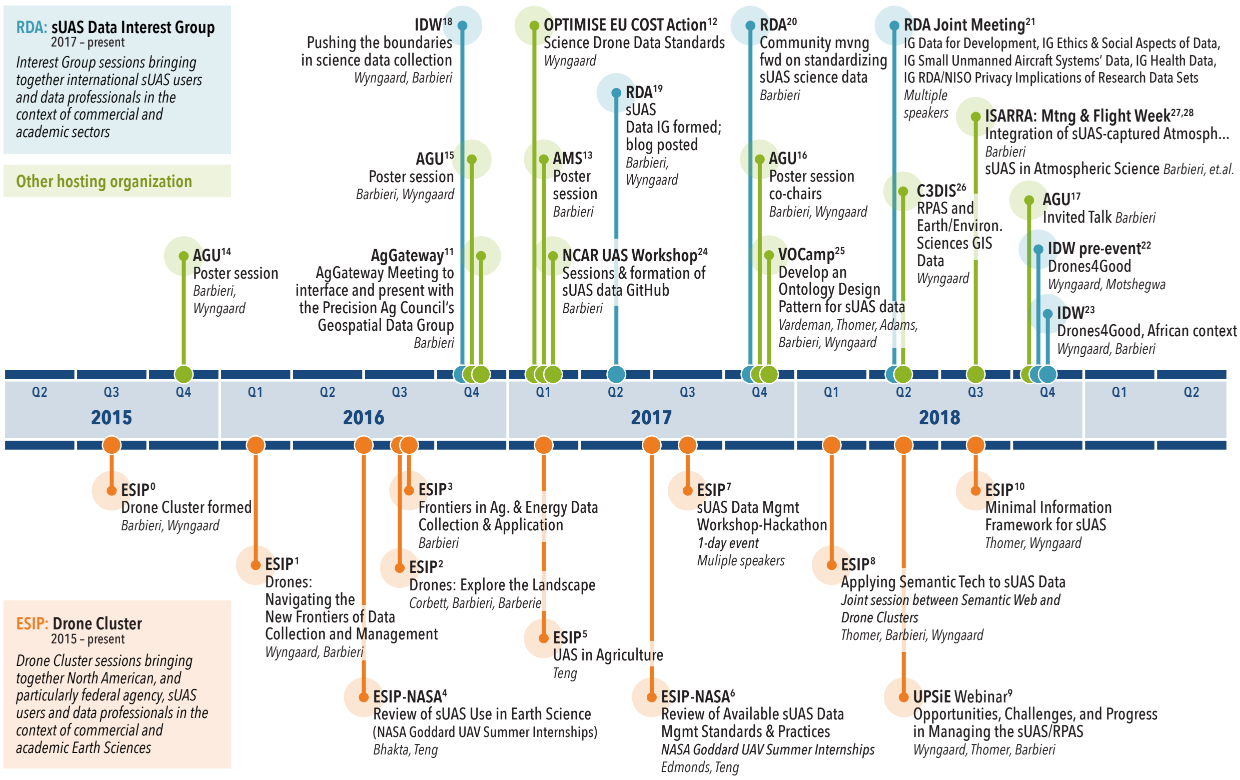


**Figure 3.** This timeline summarise the events the authors have used to engage with governmental organisations, commercial sUAS platform and tool providers, academic scientists, and both commercial and academic data professionals.How do we want to do citations for this - could list urls right here in caption like footnotes/or add them to the full reference list. The latter, however, will mean significant edits to the diagram are requiredany time we change any citation in the whole doc....

### 2.1.1. Earth Science Information Partners Federation

This effort was originally born out of a perceived need within the Earth Science Information Partners (ESIP) Federation that resulted in the creation of the Drone Cluster[37] in 2015. ESIP is *"an open, networked community that brings together science, data and information technology practitioners. ESIP is supported by NASA, NOAA, USGS, OGC, and 110+ member organizations"* [38]. Since then this cluster has run multiple sessions at ESIP meetings, hosted interns, and produced prototyping projects[39,40]. At the 2017 Summer ESIP meeting, the cluster held a 1-day workshop on sUAS data where individual researchers and representatives from multiple commercial (ESRI, DJI, SenseFly, OGC), and federal (NASA, NIST, NOAA, USGS) organisations attended and presented on their perspectives on sUAS data management approaches[41].

### 2.1.2. Rsearch Data Alliance

To engage a more global community (ESIP is a largely North American based organization), in late 2016 a sUAS Data Interest Group (IG) [42] was chartered within the Research Data Alliance (RDA). RDA is a multinational organisation funded to *"...build the social and technical bridges that enable open sharing of data."* [43]. Since review and endorsement, the IG has held sessions at each of the biannual RDA plenaries; through these efforts, it has been possible to initiated working relationships with multiple other RDA groups pioneering technological, legal, political, and ethical efforts in the global push for better open data practices and tooling. Further, as an international organisation with annual meetings in both North America and the European Union, it has been possible to engage with a geographically far larger distribution of researchers.

### *2.2. Additional key events and communities:*

Through and beyond the RDA and ESIP, this effort has been privileged to engage constructively with multiple groups who are examining issues related to sUAS data. In many cases these groups are creating resources of value to the broader community, while others are exemplars for the sUAS community to look to for guidance and foundations. The following seeks to highlight some of these for two reasons: (1) to facilitate greater collaboration within and across domains where groups have developed a resources others might reuse and build on, and (2) to propose possible foundational building blocks from existing analogous efforts. It should be noted, however, that this list is not a complete set of all relevant parties, and is biased by (a) the practical limitations of who the volunteer based ESIP and RDA efforts were able to reach, and (b) by the fact that in many cases those doing notable work simply do not currently have any public facing instance of such . Regardless of these limitations, Figure 4 summarises which organisations and community groups have been key in identifying particular challenges to sUAS data management, and the following sections briefly highlight some of the key community groupswe suggest are worth future reference and further engagement.

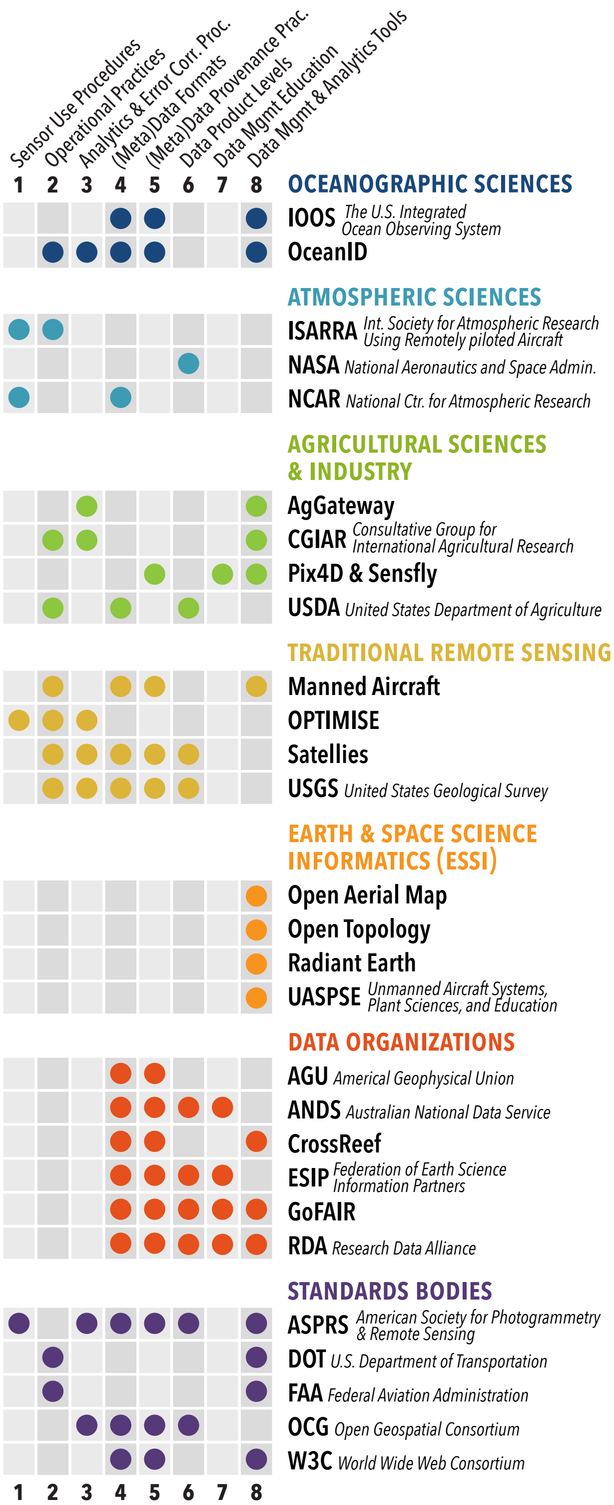


**Figure 4.** The above diagram is a summarised view of the many different communities the authors have interacted with in considering sUAS data management. Additionally, it calls out the eight key challenges to sUAS data management that this paper seeks to highlight. These eight challenges are discussed in detail in Section 3, but are listed here in order to point to the primary sources for such. Each community, organisation, and field listed here has served to call out these needs through reports, papers, posters, conference sessions, community calls, and a multitude of informal conversations at various meetings, flying fields, and hallways. The grouping by common discipline, mandate, or role is an indication of the context within which this effort has engaged with each; but in many cases these communities and organisations are contributing to multiple fields and act in multiple roles.

#### 2.2.1. Atmospheric Science

In February 2017 NCAR's Earth Observing Laboratory (EOL) hosted a workshop "...to collect information about the needs of the NSF funded community in using sUAS for atmospheric research...". While the workshop was focused on key issues other than data management the final report [44] emphasises the need for formal sensor qualification research and the creation of standardised use procedures, an issue the International Society for Atmospheric Research using Remotely piloted Aircraft (ISARRA)[45] is also discussing. For instance, the impact of placement of common atmospheric sensors on multirotors on data quality has now been the subject of multiple studies[46,47]).

#### 2.2.2. Sciences and Industry

The commercial field of precision Agriculture has been evaluating solutions particularly for data analytics and integration from not just drones but also the many diverse sensor streams feeding commercial precision farming now. An AgGateway 2017 meeting highlighted these challenges in a panel, and commercial sUAS providers (Sensefly) have worked with ESRI and Pix4D analytics tool providers to develop defacto standards for image data [48]. The United States Department of Agriculture - in addition to hosting a key controlled vocabulary Bar can you add the citation for this, also edit this paragraph as you see fit! - additionally is exploring standardised protocols for sUAS data capture[49].

#### 2.2.3. Oceanographic Sciences

Underwater gliders are an obvious analogous system to sUAS and the oceanographic research community has put significant effort into standardising their data management procedures. It may consequently serve the sUAS community well to adapt some of their tools and practices to its purposes. Key members of this effort include the US Integrated Ocean Observing System (IOOS) who have a glider Data Assembly Center (DAC)[50] and have therefore defined a NetCDF standard to which glider data submitted to their data archive must adhere. The UK Oceanids command and control data system[51,52], alternatively, have a real time web portal interface to deployed science gliders. The tool stack created to support this iterface was built to enable the automation of both operations and science data analytics (including data quality control and assurance procesesses) and is built largely on standards by the Open Geospatial Consortium (OGC) and World Wide Web Consortium (W3C).

#### 2.2.4. Traditional Remote Sensing

Large scale data management infrastructure built to manage both large scale and small format satellites and manned aircraft are not entirely portable to sUAS applications for various reasons discussed in Section 3. However, while many studies continue to explore where sUAS are and are not the optimal remote sensing solution; a great deal of traditional remote sensing expertise, knowledge, and infrastructure can be drawn on in building infrastructure for sUAS . The clearest instance of this is perhaps the use of Photogrammetric techniques in stitching sUAS imagery and the use of standardised spectral band processing algorithms and indices for sUAS data interpretation. An example of specific knowledge transfer from remote sensing to sUAS is the work by the EU based OPTIMISE[53] who have been working on standardised spectral information systems for many years, and who have most recently expanded to naturally including practices for sUAS mounted spectral sensors. Their engagement with the spectral sensing community, including a in-depth survey of optical sUAS practices and community knowledge is ongoing, with initial survey results are available in[54]. Similarly the United States Geological Survey, who have extensive experience using manned aircraft, have one of the few publicly accessible sUAS data management plans[55] based largely on their historical experience and domain knowledge.

Finally, the ESA and NASA's culture of making data appropriately open [56,57] and of using or publishing open source software [58,59] are arguably models for sUAS to follow. While manned aircraft engineering standards are not commonly directly applicable to sUAS , the development processes, decision metrics, and operational practices used are increasingly applicable as sUAS are integrated into controlled airspaces. Furthermore, manned aircraft data processing tool stacks are often build on widely used standards such as those from the OGC's Aviation Domain Working Group[60], again sUAS would likely do well to follow this example.

## 3. Results

Emmerging from the above described engagement efforts have been two key results: (1) how sUAS data are unique and consequently in part require custom management solutions, and (2) eight challenges to be addressed in order to access the full potential value of sUAS data.

### *3.1. sUAS data are unique and in need of unique managment infrastructure*

**sUAS data are uniquely 4+ dimensional**
All sUAS data is associated with a location in both time and 3-dimensional space. While location- and time- stamped data are not unusual, multiple streams of simultaneously captured values captured from a moving 3-dimensional trajectory at sporadic temporal intervals such as sUAS enable are uncommon. Furthermore, to correctly interpert many sUAS data requires additional metadata streams, such as the time-series stream of the sUAS attitude, or an instantaneous measure of local luminosity. sUAS data is therefore unique for its mandatory 4+ dimensionality: multiple co-captured geospatially-tagged measurements of varing precisions, taken within multiple discretised time periods, along a 3-dimensional trajectory.

**sUAS data provide uniquely high spatiotemporal resolutions**
sUAS are being used in the sciences largely as they are a low-cost way of quickly capturing high spatial and temporal resolution data. For instance, spatially, even low cost sUAS can achieve <5cm/pixel horizontal ground resolution imagery, and they have the entirely unique ability to sample at similar resolutions in fully customised vertical profiles. Further, temporally, sUAS systems may be deployed both repeatedly, and dynamically in response to real time changing circumstances, with periodocties ranging from minutes to years. This high temporal resolution is most visibly advantageous in the use of sUAS in disaster response (e.g wildfires, flooding, or earthquakes), but it is equally useful in scientific research that can be subject to both unforeseen changes in long planned observations (e.g unpredictable wildlife activity, or unforeseen operational restrictions) and spontaneous opportunities (e.g. an unanticipated flooding event of an area of interest). sUAS consequently are providing a uniquely high resolution low cost offering that neither manned aircraft systems or satellites — both of which require months of planning and very large budgets – nor ground based sensors or other low altitude platforms (e.g. kites, balloons) cannot readily offer.

**sUAS data are classically *Big***
sUAS data are **Big** in all four of the classic *Big Data* characteristic 'Vs' [61]. The *variety* in form, function and *veracity* of sUAS data is only limited by current sensor miniaturising technology and regulations, but currently commonly includes both low cost and professional grade: multi- and hyper- spectral imagery; multiple LIDAR and RADAR sensing technologies; a wide range of gas and particulate matter sensors; mechansims for water, genomics, and other physical sample capture; and common time series parameters such as temperature, pressure, humidity, and the local characteristics of radio frequency signals. The *volume* of data that spectral sensors particularily can quickly capture is nontrivial, with a single flight able to return tens of GB of raw data. And finally the faster sUAS mounted sensors can capture quality data, the faster the sUAS can fly, and therefore the larger the area and amount of data that can be covered in a single flight, all of which mean sUAS data capture rates are most likely to continue increasing in *velocities* going forwards as technology improves.

**sUAS data are increasingly created by small science**

Large unmanned system technologies such as unmanned planes or underwater gliders have historically been accessible only to researchers working at large scale and often government based research institutions with the resources to build and maintain large scale research facilities. However, small sUAS have made it possible for small and modestly funded teams of researchers to use unmanned technologies. The adoption of sUAS technology by these smaller and more ad hoc teams has consequences for the management of these data both as it increases the quantity of data being captured by researchers overall, and because it increases the need for common practices that cross discipline boundaries. Whereas large scale research endeavors (sometimes called *big science* often have correspondingly robust plans and infrastructures for data archiving and management, smaller scale teams (sometimes called *small science* or *little science* have correspondingly ad hoc and idiosyncratic data management practice [62,63].

### *3.2. Eight community distilled sUAS data management challenges to be addressed*

1. **Sensor use procedures:** Sensor specific, tested and qualified use procedural best practices and standards are urgently needed in a common language. These best practice methodology and procedural guidelines should be developed and provided either by the manufacturer or the research community and include: mounting requirements on various platforms, calibration, ground truthing, and maintenance procedures, sample rates, flight patterns, and required metadata for data use and publication. The need for these is both for user ease and so as to enable greater automation in the capture of data provenance. As mentioned existing initial work on this issue has already appeared within the atmospheric community [46,47] and the Agricultural Sciences[48]. While these procedures, are largely currently not instantiated in open machine readable forms, they represent a direction for others to follow and contribute further to.
2. **Operational practices:** Having best practices regarding operational protocols for scientific research will lower the barrier to entry for new users, allow training materials to be standardised for the many new training courses being created, and reduce the burden on operators which can only lead to safer operations. Further, while many countries have now begun to settle on regulations, many research organisations are still grappling with their own internal policies and protocols. Researcher operational best practices, created based on the experience of those who have been operating for longer, could serve to accelerate organisational protocol deployment in a country agnostic manner. One examples of such that is readily accessible comes from University of Exeter 'sRemote Sensing Laboratory [64].
3. **Analytics and Error correction procedures:** Best practices and acceptable error tolerances for primary sensor taxonomy branches and the associated processes need to be defined so as to avoid unintentional — but easy to introduce — errors [65]. These are needed equally by tool providers (commercial and open source) so as to allow them to build to a standard, and by user community so as ensure correct data interpretation. Defining such will additionally contribute to efforts to define sensor use best practices and metadata creation, capture, and archive tooling.
4. **Data and metadata data formats:** Guidelines regarding best practice metadata and data formats would serve the community, not as any form of restriction, but rather as a simple means of reducing workloads for both research sUAS operators and technical developers of: sensors, sUAS platforms, and the many components necessary in a data management tool stack. Having published recommended open formats based on community experience would similarily lower the barrier to novel experiments and enable both open source and commercial developers to create reusable tools.
5. **Data and metadata provenance practices:** Definitions of what parameters are required to make a data value, set, or product reusable – in potentially other scenarios than that for which it was originally captured or created – is necessary as both a practical guideline for operations and to facilitate the creation of tools to support the automated capture of this provenance.

6. **Data product levels:** Defining suggested data product levels for various data types would facilitate both data archives and single researchers in determining what data should be archived, at what quality levels, at what resolutions, and with what associated metadata as required for likely reuse. This could be done for various primary parameter taxonomy branches, such for spectral data captured for Agricultural Sciences, and for atmospheric time series for Atmospheric Sciences.
   A crucial and complex sub-component to data product level definitions is the potential ethics driven policies that will govern sharing sUAS data. FAIR does not require open access, and others are exploring the ethical implications of both FAIR and open data in general [66,67]. Not least because of their historical military associations of sUAS but also due to the potential to easily violate important privacy restrictions with sUAS mounted sensors, the community needs to discuss both locally and internationally, what best practices might be for governing sUAS data's desireable degree and form of openness.
7. **Data management and analytics tools:** As shown in 4, many of the relevant organisations already have some portion of a sUAS data analytics and management tool stack. However, the tools these bodies offer are only sUAS specific in a minority of cases. Rather, the majority were developed for other data types and are now being adapted for sUAS . More resources and effort are therefore necessary to accelerate these adaptations; and it is note worthy that by addressing the above challenges, it would becomes significantly easier for resource pooling across development efforts.
8. **Data management education:** As the domain grows there is an increasing demand for introductory information that properly addresses the multitude of new expertise needed to effectively use sUAS . In response many universities and other institutions are beginning to formally train research sUAS operators. An acknowledged but core missing component of these training curricula is any information on comprehensive consideration for science data good practices. Bringing together data management training and sUAS training offers a convenient opportunity, but one that depends heavily on investment being made first in the above challenges
   .

## 4. Discussion

As detailed in Section 3, the primary outcomes of engaging with the nascent sUAS community are: (a) the identification of how sUAS data are unique and where there are shared characteristics with more traditional data capture platforms, and (b) that as a result there are eight community identified challenges to improving sUAS data management.

Regarding how sUAS data are unique, the following should be noted. (1) Though there are many geospatial data formats that capture vector and raster data, stationary time series data, and high dimensionality data, and while tool stacks exist for processing and managing these data, these tools do not currently readily support the particular combination of metadata streams and multiple parameter capture sUAS data often consist of or require for correct interpretation. Similarly (2), the high spatial and temporal resolutions sUAS data are capable of capturing presents a new complicating factor for data management infrastructure. These resolutions require both potentially new multidimensional formats, schemas, and ontologies (or at least new workflow tools for handling the novel combination of such sUAS data involved), and also demand high processing times, more automated quality control, and new data product distribution tools. (3), considering that the majority of tool stacks build for *BigData* assume operation on cloud or at least mains powered computing resources, while the veracity, variability, velocity, and volume of sUAS data might be equivalent or even lower than common *BigData* tasks, in many cases the need is for processing of such on low power or low bandwidth edge compute devices. And finally, (4) research has shown Andrea - citation?, that the range of data practices utilised by smaller teams should be considered a feature rather than a *bug*; this is because the data workflows and practices must be customized to the unique contexts and goals of a given group, project, and organizational structure. Standardized workflows across all smaller research teams are neither achievable nor desirable. Consequently, sUAS data management solutions need to be created with the necessarily diverse data practices of a small lab researcher specifically in mind, and this is all the more so true given the wide spectrum of disciplines sUAS users include.

Regarding the challenges outlined. As new sensors, sUAS platforms, and analytics techniques develop, it is clear that addressing solutions to these challenges will require updates and extensions. However, initial efforts on each are the only way to ensure such periodic updates, extensions, and community driven maintenance will be plausibly practical, sustainable, and backwards compatible to any degree in the long term. Further, by initiating the development of solutions to any of the following in a collaborative manner with a view to long term sustainability, partial solutions will be both immediately accessible for use by others and accessible for extension, iteration, and improvement such that gradually more complete solutions naturally arise. That is, provided long term maintainability and extensiability are considered in initial work.

## 5. Conclusions

The use of sUAS for data capture is increasing rapidly, both for commercial applications and as a new platform for data capture for a wide and diverse spectrum of research fields. As a nascent field with many avenues of development underway to increase both operational and scientific platform maturity, the issue of managing and optimising the data flow from sample to knowledge product has not been extensively explored. This paper describes an effort to explore what resources are currently available for handling sUAS data, what approaches are currently being used, and where there are challenges to fully realising sUAS data's value. As a largley unfunded effort subject largely to the authors abilities to take advantage of opportunities that either arose organically or were comensually available, this exploration was not comprehensive. It has, however, engaged with a significant breadth of domain users, developers, commercial participants, and analogous mature fields from which sUAS might learn. In addition to finite scope, a key limitation in this engagement is that the majority of work was done in North America, however, this was not exclusive with 6 out of the 28 formal engagements listed occuring elsewhere in the world. To the best of our knowledge this is the only effort to achieve the above at any international scale. There are two significant novel outcomes of this work.

(1) The identification of the combination of characterists that sUAS data commonly exhibit, shows that while it shares many trates with more traditional methods of data capture, the combined differences mean existing infrastructure as currently developed and deployed is not capable of enabling users fully realise the potential value of sUAS data. These primary characteristcs were: (i) sUAS data are uniquely 4+ dimensional, (ii) sUAS data provide uniquely high spatiotemporal resolutions, (iii) sUAS data are classically Big, and (iv) sUAS data are increasingly created by small science.

(2) The detailing of eight specific challenges that must be addressed in order for sUAS to become a trusted, reliable, and optimally useful data capture platform: (i) Sensor use procedures, (ii) Operational practices, (iii) Analytics and Error correction procedures, (iv) Data and metadata data formats, (v) Data and metadata provenance practices, (vi) Data product levels, (vii) Data management and analytics tools, (viii) Data management education.

Based on these, we conclude that a conscience and determined effort by a global selection of researchers, to openly draft community driven data management best practices for the capture and managment of sUAS data, would likely realise many gains and be an important step towards supporting the reproducibility and reliablity of drone data research, as well as increasing the reuse of sUAS data. In the immediate, it would cost time and effort, but in the very near future it would achieve: (i) significantly reduce the total quantity of poorly curated sUAS data likely to otherwise be lost in the near future, (ii) minimise the length of what will otherwise be an extended period of partial and inadequate data management tooling for sUAS users making operations inefficient over a longer period of time than necessar, (iii) allow the community to circumvent the familiar larger and more expensive challenges of legacy data rescue and community wide retooling, and retraining, (iv) lower the barrier to entry for researchers entering the field and seeking to produce robust and reusable data. (v) enable collaborative rather than disparate and ad hoc building of common sUAS data infrastructure, and finally (vi) increase the transparency of sUAS data processing workflows. However, the window of opportunity within which to craft such is finite and closing, given the immediate need for data tooling and practices and already growing set of sUAS data.

**Author Contributions:** For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y.", please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

**Funding:** Please add: "This research was partially funded by Earth Science Information Partners." and "The Research Data Alliance US".

**Acknowledgments:** In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

**Conflicts of Interest:** "The authors declare no conflict of interest."

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| sUAS | Small Unmanned Aircraft Systems |
| IG | Interest Group |

## References

1. Stöcker, C.; Bennett, R.; Nex, F.; Gerke, M.; Zevenbergen, J. Review of the current state of UAV regulations. *Remote Sensing* **2017**, *9*, 33–35.
2. Anweiler, S.; Piwowarski, D. Multicopter platform prototype for environmental monitoring. *Journal of Cleaner Production* **2017**, *155*, 204 – 211. Sustainable Development of Energy, Water and Environmental Systems, doi:https://doi.org/10.1016/j.jclepro.2016.10.132.
3. Cho, R. How Drones are Advancing Scientific Research | Lamont-Doherty Earth Observatory. Technical report, Columbia University, 2017. Accessed: 2017-11-14.
4. Colin, C. The Potential of Drones in Research. Technical report, Environmental Biophysics, 2015. Accessed: 2017-11-14.
5. DRONES Reporting for Work. www.goldmansachs.com/insights/technology-driving-innovation/drones/. Accessed: 2019-03-19.
6. Levick, R. Drone Industry Just Beginning To Take Off. *Forbes Magazine* **2018**.
7. Commercial drones are here: The future of unmanned aerial systems. https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/commercial-drones-are-here-the-future-of-unmanned-aerial-systems. Accessed: 2018-11-25.
8. Shacklett, M.E. How the Drone Market is Picking Up Speed. https://www.pobonline.com/blogs/23-geodatapoint-blog/post/101409-how-the-drone-market-is-picking-up-speed, 2018. Accessed: 2018-11-25.
9. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; Bouwman, J.; Brookes, A.J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C.T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A.J.G.; Groth, P.; Goble, C.; Grethe, J.S.; Heringa, J.; Hoen, P.A.C.t.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S.J.; Martone, M.E.; Mons, A.; Packer, A.L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M.a.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **2016**, *3*, 160018.
10. Realizing the Power of Enterprise Data. https://www.veritas.com/form/whitepaper/realizing-the-power-of-enterprise-data. Accessed: 2019-03-19.

11. Cragin, M.H.; Palmer, C.L.; Carlson, J.R.; Witt, M. Data sharing, small science and institutional repositories. *Philos. Trans. A Math. Phys. Eng. Sci.* **2010**, *368*, 4023–4038.
12. Kowalczyk, S.; Shankar, K. Data sharing in the sciences. *Ann. Rev. Info. Sci. Tech.* **2011**, *45*, 247–294.
13. Wallis, J.C.; Rolando, E.; Borgman, C.L. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS One* **2013**, *8*, e67332.
14. Borgman, C.L. The conundrum of sharing research data. *Journal of the American Society for Information* **2012**, pp. 1–40.
15. Nosek, B.A.; Spies, J.R.; Motyl, M. Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspect. Psychol. Sci.* **2012**, *7*, 615–631.
16. Peng, R.D. Reproducible research in computational science. *Science* **2011**, *334*, 1226–1227.
17. Martone, M.E. FORCE11: Building the Future for Research Communications and e-Scholarship. *Bioscience* **2015**, *65*, 635–635.
18. Lecarpentier, D.; Wittenburg, P.; Elbers, W.; Michelini, A.; Kanso, R.; Coveney, P.; Baxter, R. EUDAT: A New Cross-Disciplinary Data Infrastructure for Science. *International Journal of Digital Curation* **2013**, *8*, 279–287.
19. Orszag, P. Open government directive. *Memorandum for the Heads of Executive Departments and Agencies* **2009**, *2*.
20. Commission Decision of 12 December 2011 on the reuse of Commission documents. *OJ* **14-12-2011**, *3*, 229 – 232. Accessed: 2018-6-11.
21. Open Data Gaining Momentum in Africa **29-09-2015**. Accessed: 2018-6-11.
22. Stall, S.; Parsons, M.A.; Lehnert, K.; Cutcher-Gershenfeld, J.; Hanson, B. Enabling FAIR Data Across the Earth and Space Sciences - Eos. https://eos.org/agu-news/enabling-fair-data-across-the-earth-and-space-sciences, 2017. Accessed: 2018-12-11.
23. Wilkinson, M.D.; Verborgh, R.; da Silva Santos, L.O.B.; Clark, T.; Swertz, M.A.; Kelpin, F.D.L.; Gray, A.J.G.; Schultes, E.A.; van Mulligen, E.M.; Ciccarese, P.; Kuzniar, A.; Gavai, A.; Thompson, M.; Kaliyaperumal, R.; Bolleman, J.T.; Dumontier, M. Interoperability and FAIRness through a novel combination of Web technologies. *PeerJ Comput. Sci.* **2017**, *3*, e110.
24. Mons, B.; Neylon, C.; Velterop, J.; Dumontier, M.; Da Silva Santos, L.O.B.; Wilkinson, M.D. Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Inf. Serv. Use* **2017**, *37*, 49–56.
25. Elankavi, R.; Kalaiprasath, R.; Udayakumar, R. DATA MINING WITH BIG DATA REVOLUTION HYBRID. *International Journal on Smart Sensing & Intelligent Systems* **2017**, *10*.
26. Lv, Z.; Song, H.; Basanta-Val, P.; Steed, A.; Jo, M. Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics. *IEEE Trans. Ind. Inf.* **2017**, *13*, 1891–1899.
27. Gökalp, M.O.; Kayabay, K.; Zaki, M.; Koçyiğit, A.; Eren, P.E.; Neely, A. Big-Data Analytics Architecture for Businesses: a comprehensive review on new open-source big-data tools **2017**.
28. Stein, L.D.; Knoppers, B.M.; Campbell, P.; Getz, G.; Korbel, J.O. Data analysis: Create a cloud commons. *Nature* **2015**, *523*, 149–151.
29. Hashem, I.A.T.; Yaqoob, I.; Anuar, N.B.; Mokhtar, S.; Gani, A.; Ullah Khan, S. The rise of "big data" on cloud computing: Review and open research issues. *Inf. Syst.* **2015**, *47*, 98–115.
30. Chaudhary, S.; Somani, G.; Buyya, R. *Research Advances in Cloud Computing*; Springer, 2017.
31. von Oehsen.; Fox, G.; Ziolkowski, M.; Bedrossian, A.; Fay, D.; Bottum, J.; Atkins, D.; Blatecky, A.; Mcmullen, R.; Tannenbaum, T.; Cheetham, J.; Wilgenbusch, J.; Bhatia, K.; Deumens, E.; Barr. The Future of Cloud for Academic Research Computing. Technical report, National Science Foundation, 2017.
32. Dataset Search. https://toolbox.google.com/datasetsearch. Accessed: 2018-11-25.
33. Stall, S. Enabling Findable, Accessible, Interoperable, and Reusable Data - Eos. https://eos.org/editors-vox/enabling-findable-accessible-interoperable-and-reusable-data, 2017. Accessed: 2018-11-25.
34. Enabling FAIR Data Project. http://www.copdess.org/enabling-fair-data-project/. Accessed: 2018-11-25.
35. L'heureux, A.; Grolinger, K.; Elyamany, H.F.; Capretz, M.A.M. Machine learning with big data: Challenges and approaches. *IEEE Access* **2017**, *5*, 777–797.
36. Karpatne, A.; Ebert-Uphoff, I.; Ravela, S.; Babaie, H.A.; Kumar, V. Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Trans. Knowl. Data Eng.* **2018**, pp. 1–1.
37. J Wyngaard, L.B. Drone Cluster - Federation of Earth Science Information Partners. http://wiki.esipfed.org/index.php/Drone_Cluster. Accessed: 2018-11-14.

38. Burgess, A. ESIP | Connecting Science, Data and Users. http://www.esipfed.org/. Accessed: 2018-4-14.
39. Barbieri, L.; Wyngaard, J.; Bhakta, K.; Teng, B. Resource Catalogs **2016**.
40. Edmonds, K. UAS Spectral Sensing Data Management, 2018.
41. ESIP Summer Meeting 2017 | July 25-28, 2017 – Bloomington, IN | ESIP. http://www.esipfed.org/meetings/upcoming-meetings/esip-summer-meeting-2017. Accessed: 2018-4-14.
42. J Wyngaard, L.B. Small Unmanned Aircraft Systems' Data IG. https://www.rd-alliance.org/groups/small-unmanned-aircraft-systems%E2%80%99-data-ig, 2016. Accessed: 2019-2-18.
43. Rda. Research data alliance, http://rd-alliance.org/ **2013**.
44. Oncley, S.; Roberts, G.; Schoenung, S.; Vömel, H.; Wolff, C. The NCAR / EOL Community Workshop on Unmanned Aircraft Systems for Atmospheric Research Authors **2017**.
45. International Society for Atmospheric Research using Remotely piloted Aircraft. http://www.isarra.org/. Accessed: 2019-03-19.
46. Rossi, M.; Brunelli, D. Gas Sensing on Unmanned Vehicles: Challenges and Opportunities. *2017 New Generation of CAS (NGCAS)* **2017**, pp. 117–120.
47. Villa, T.; Gonzalez, F.; Miljievic, B.; Ristovski, Z.; Morawska, L. An Overview of Small Unmanned Aerial Vehicles for Air Quality Measurements: Present Applications and Future Prospectives. *Sensors* **2016**, *16*, 1072.
48. EXIF and XMP tag information read by Pix4D Desktop. https://support.pix4d.com/hc/en-us/articles/205732309-EXIF-and-XMP-tag-information-read-by-Pix4D-Desktop. Accessed: 2018-11-17.
49. LTAR-NEON Collaboration to Quantify Rangeland Vegetation Production. https://www.ars.usda.gov/research/project/?accnNo=429391. Accessed: 2019-03-19.
50. ioosngdac.
51. NMF Vehicles. https://mars.noc.ac.uk/. Accessed: 2018-11-26.
52. Buck, J.; Phillips, A.; Lorenzo, A.; others. Oceanids command and control (C2) data system-Marine autonomous systems data for vehicle piloting, scientific data users, operational data assimilation, and big .... *AGU Fall Meeting* **2017**.
53. ESSEM COST Action ES1309. http://optimise.dcs.aber.ac.uk/. Accessed: 2019-03-19.
54. Aasen, H. State-of-the-art in UAV remote sensing survey-First insights into applications of UAV sensing systems. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **2017**, *42*, 1–4.
55. Cynthia, F.; Benjamin, T.; Kim, R. USGS UNMANNED AIRCRAFT SYSTEMS DATA DELIVERY SPECIFICATION. Technical report, United States Geological Survey, 2018. Accessed: 2019-03-19.
56. OPEN ACCESS AT THE EUROPEAN SPACE AGENCY. http://open.esa.int/open-access-at-esa/. Accessed: 2019-03-19.
57. Data and Information Policy. https://science.nasa.gov/earth-science/earth-science-data/data-information-policy. Accessed: 2019-03-19.
58. GSFC Open Source Software. https://opensource.gsfc.nasa.gov/. Accessed: 2019-03-19.
59. Mattmann, C.A. Computing: A vision for data science. *Nature* **2013**, *493*, 473.
60. Aviation Domain Working Group. https://www.opengeospatial.org/projects/groups/aviationdwg. Accessed: 2019-03-19.
61. Hitzler, P.; Janowicz, K. Linked Data, Big Data, and the 4th Paradigm. *Semantic Web* **2013**, *4*, 233–235.
62. Wallis, J.C.; Rolando, E.; Borgman, C.L. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PloS one* **2013**, *8*, e67332.
63. Cragin, M.H.; Palmer, C.L.; Carlson, J.R.; Witt, M. Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2010**, *368*, 4023–4038.
64. Cunliffe, A.M.; Anderson, K.; DeBell, L.; Duffy, J. A UK Civil Aviation Authority (CAA)-approved operations manual for safe deployment of lightweight drones in research. *Int. J. Remote Sens.* **2017**, *38*, 1–8.
65. Congalton, R.G. Remote sensing and geographic information system data integration: error sources and. *Photogrammetric Engineering & Remote Sensing* **1991**, *57*, 677–687.
66. Boeckhout, M.; Zielhuis, G.A.; Bredenoord, A.L. The FAIR guiding principles for data stewardship: fair enough? *European Journal of Human Genetics* **2018**, p. 1.

67. Serwadda, D.; Ndebele, P.; Grabowski, M.K.; Bajunirwe, F.; Wanyenze, R.K. Open data sharing and the Global South—Who benefits? *Science* **2018**, *359*, 642–643.

© 2019 by the authors. Submitted to *Remote Sens.* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).