Drone Data API

Enabling FAIR sUAS data to become the norm through community engagement and standards based tool

Interim Report 1 March 2020

PI: J Wyngaard
Co-PI: C Vardeman

Table of contents

1 Recap of Project Goals

1.0.1 Problem statement

In recent years many scientists have begun to use small Unmanned Aircraft Systems, [sUAS] (also known as drones) as standard tools for research data capture. As a nascent and emerging tool, the need for drone data management cyberinfrastructure has increased as adoption has grown. For the many reasons discussed in the proposal there is much to be gained by academia and society by meeting this need. This project therefore set out to build open source, community driven, foundational data management cyberinfrastructure so as to enable drone data to be easily made FAIR [Findable Accessible Interoperable Reusable]. Specifically we proposed to deliver:

1. A published and complete open source API and software tool stack for sUAS data
2. An RDA endorsed set of sUAS data best practice recommendations
3. At least three open source applications using the API to meet specific disciplinary needs

1.0.2 Project approach

Our target user community is scientific drone users (although we would welcome commercial adoption and input). In order to maximise adoption in a community that spans a majority of scientific disciplines, and to build on the existing potential for drones to democratise access to GIS data; all work is being done transparently in the open, all technical outputs are open source. Further, we have already and will continue to make a significant effort to carry out all work in a manner that grows a community around these tools and encourages outside comment and contribution. A key component of which has been to carry out all work in community with the Research Data Alliance (RDA), Earth Science Information Partners (ESIP), and the Open Geospatial Consortium (OGC) (amongst others).

This same approach is being applied to the technical design decisions made regarding the Application Programming Interfaces (APIs) and User Applications we are developing. Broadly this means: (a) whenever possible existing standards are being used or built upon, and (b) that all components are being built to exploit and inhabit the cross disciplinary trends towards: increasingly networked devices, API based tooling, containerised deployments, and the use of linked data.

To this end we proposed a 3 phase execution plan over 2 years: (1) Design, (2) Development, (3) Deployment. As of March 2020, we are currently half way through Phase 2: Development. The following therefore reports on: firstly our outcomes to date as they relate to the Design and Development phases thus far, and secondly outlines the detailed plans we now have for executing the rest of our development and the deployment and adoption phase. Finally, we conclude with some of the challenges that have been encountered, and how plans have been adjusted to account for such.

2 Outcomes To Date

At a high level, the acronym and logo shown in Figure 1 were created, and a basic github pages website created: www.landrs.org. Initially work began within the OGC Github space: https://github.com/opengeospatial/LANDRS. However, while this was done with a mutual desire to build a tight relationship with the OGC community, we quickly found we needed so many repositories that this was no longer tractable. All work has now been moved to a dedicated LANDRS space: https://github.com/landrs-toolkit

Linked-data API for Networked DRoneS

Figure1: Project Logo and acronym

2.1 Phase1: Design

As proposed in the first 6 months of the project we ran two hackathons with our existing known community. The primary purpose of these events was to gather input from the community as the design was developed, and grow awareness of the project. The first of these took place within the ESIP Drone Cluster as a USA focused event in July 2019. The second was hosted by the RDA sUAS data Interest Group and co-located at the Research Data Alliance's 14th Plenary in Finland in October 2019.

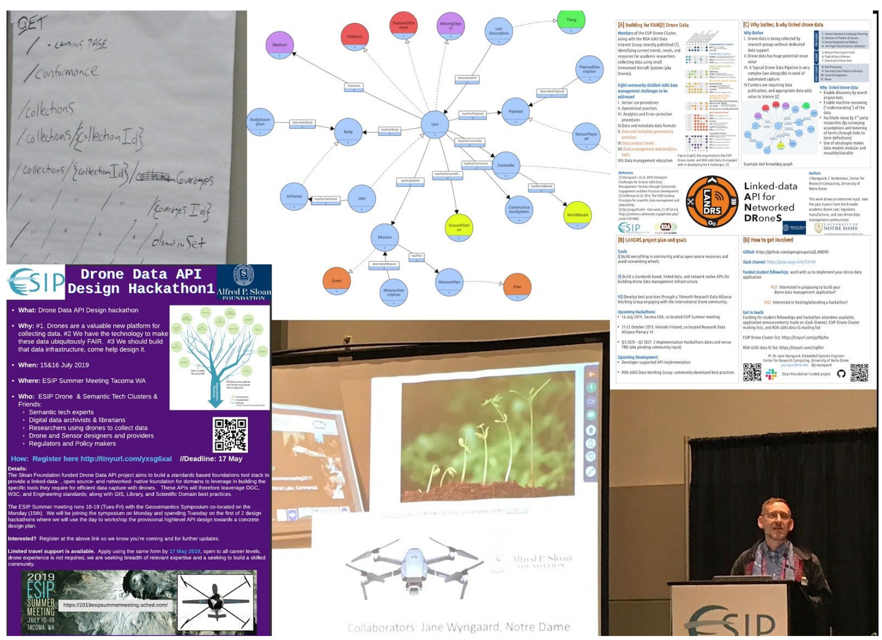

Figure 2: Images from the ESIP LANDRS hackathon. Circularly from top left: API Design brainstorming notes, Draft Ontology, Poster, event flyer, Dr Jens Klump from CISRO speaking about Drone data work in Australia.

2.1.1 ESIP Tacoma Hackathon

The ESIP Summer 2019 meeting took place in Tacoma Seattle . Using LANDRS funding we directly supported 8 out of a total of 15 attendees to a 1 day (limited by venue) event. Seven of those supported by LANDRS were from 5 USA academic institutions, with 1 attendee from Australia.   Other self supporting attendees joined from various ESIP member organisations.

The primary outcome of this event was a refinement of the LANDRS ontology (discussed further in Section 2.2).

2.1.2 RDA Helsinki Hackathon and sUAS Data IG session

Collocated with and preceding the RDA 14th plenary in Helsinki, we held the 2nd hackathon at the University of Helsinki, where we were hosted by Professor L Ruotsalainen (see Figure 3). LANDRS sponsored 3 students from the University of Botswana, and Barbara Magagna who co-chairs the RDA InteroperAble Descriptions of Observable Property Terminology WG (I-ADOPT WG) and is from the Environmental Agency Austria.

While smaller in numbers than ESIP this 2 day event was spent working on aligning the LANDRS ontology with the OGC's Sensors, Observations, Actuation, and Sampling ontology and starting to build the Python code required to use this (PySOSA discussed further later in Section 2.1.3).

Figure 3: Images from the RDA Helsinki LANDRS hackathon. Circularly from top left: PySOSA design, Hackathon attendees (Left to Right:B Mosesane, L Ruotsalainen, B Magagna, M Taupe, C Vardeman, W Raseonyana, J Wyngaard, and the Unconference postits grouped and ranked by votes.

RDA Session

During the following official RDA conference and the sUAS data IG session, we presented the LANDRS project and the concept of starting a RDA working group to work with LANDRS to develop RDA endorsed drone data best practices. As the PI chairs the IG, an unconference style mechanic was used to engage the community about this concept (see results in Figure 3), while the audience was also given space to discuss what else the IG next wanted to work on.

Unfortunately (for the project), the LANDRS proposal was not what the community in attendance voted for. Rather those in the session indicated that they saw (1) Development of recommendations regarding cyberinfrastructure for drone data (such as compute and storage capacity needs), and (2) Recommendations for Legal and Ethical considerations regarding drone data, as the areas they wished to focus on.

While this was not the desired outcome, RDA is a primarily volunteer organisation and for the most part can therefore only make progress where volunteers are willing to donate time. The PIs, however, believe that the above results are more a reflection of the particular attendees present at P14 than the global community. Evidence for this is in both the lack of volunteers to work on these after the plenary, and the many other communities that are very supportive and interested in where LANDRS is going (see planned adopters and community reviews in Section 3.13).

Consequently, we therefore intend to re-propose the Best Practices WG again once LANDRS has made further progress and can then use demonstrations to make the concept more tangible. Additionally, an effort that has developed with the USA Long Term Ecological Research Network (LTER) and which is discussed in Section 2.1.3 is leading to us creating best practices specifically to meet their organisational needs. This will further demonstrate the practicality of the concept, and hopefully make the WG more attractive.

2.1.3 Project Fellows

In the original proposal it was planned that two groups of students would apply to spend a period of weeks at the University of Notre Dame to work on the LANDRS tools. However, after our initial hackathons, and based on community response we determined that it was going to be difficult to attract the right skill sets for the given period of time and reward we were able to offer.

As an alternative, we therefore have advertised for what we termed remote fellows. In short we are continuously looking for interested students who we hire as fellows to work part time with us remotely. This has proved more successful and offers a cost efficiency gain as there are less travel expenses incurred. Students are being paid according to their University local rates equivalent to their expertise. In addition to the remote work payment, todate and going forward when possible we also bring these fellows to events as this builds our relationship with both them and their home institution supervisors, affords efficient periods of face to face work, and provides them greater experience and professional gains through attendance at co-located conferences. To date the following components of LANDRS are being worked on by the following fellows. All of these components are for the most part stand alone components, the core complexity of LANDRS APIs are being created by progrogrammers at the University of Notre Dame for now.

PySOSA

At the RDA hackathon we worked to design an alignment of OGC SOSA to the LANDRS ontology. In order to build tools on top of this alignment we need a code library to build linked drone metadata RDF graphs. PySOSA is should be released as beta in the next few weeks (source: https://github.com/landrs-toolkit/PySOSA) and will be published on PyPi for use both anyone else as well as LANDRS.

Following the release of PySOSA, these fellows will use their now acquired understanding of semantics, to create Python modules to accomplish the same with: the LANDRS ontology, an image annotation graph (http://slod.fiz-karlsruhe.de/), and OGC's Semantic Trajectory ontology.

PySOSA and community outreach fellow

- Badisa Mosesane
- 4th year Computer Science student
- University of Botswana
- Attended RDA P15
- Working with LANDRS: September 2019 - Present
- Roles: Python developer, social media manager, and website maintainer

PySOSA developer and African drone community fellow

- Wame Raseonyana
- Master in Computer Science student
- University of Botswana
- Attended RDA P15
- Working with LANDRS: September 2019 - Present
- Roles: Python developer, and co-ordinator with African drone events

Babelfish

The tool stack for onboard the drone is currently very early in the development stage, but it was always known that we would need to communicate with multiple autopilots and to annotate their data. The primary open source drone communication protocol is Mavlink so we began on a C library to serve as the start of this babelfish application. Unfortunately the student was not able to complete this work before her studies resumed. Either someone else will need to complete

this or she will be brought on again if her schedule allows, which happens will depend on the timing of other components.
https://github.com/landrs-toolkit/LANDRS_Babelfish

Babblefish fellow

- Johanne Jansen Van Vuuren
- 4th year Electrical Engineering student
- University of Cape Town
- Working with LANDRS: November - December 2019. Interested to be involved again during semester breaks
- Role: Develop the C code to annotate Mavlink data onboard the drone.

LTER Information Managers Drone Metadata Best Practices

The PI was approached by the LTER Information Managers for help as they are receiving drone data at their archives and the PI (within ESIP) has developed a draft set of minimal metadata a drone dataset should be annotated with. Given the RDA sUAS data IGs reluctance to develop best practices just now, this appeared to be a convenient opportunity to develop something that is beyond theory, will already have an immediate adopter, and which will serve as a reference as we build out tha APIs for the core LANDRS tooling.

Practical Drone MetaData best practices for Data Archives fellow

- Lindsay Barbieri
- Doctoral candidate University of Vermont
- Working with LANDRS: February 2020 - March 2020 (anticipated)
- Role: Integrate the ESIP theoretical drone minimal information framework, with the latest USGS data package metadata, and example LTER drone data metadata, to create a practical list of metadata for LTER information managers to require their users submit with drone data. These metadata will initially pertain to visible and thermal imagery, and will list each as required for either discovery (Finadable), applicability (Accessibility and Interoperability) or Reuse. This categorisation is necessary for the LTER managers to map this information to their data management model.

2.2 Phase2: Development

Towards the end of 2019 we moved into the Development phase of the project. In addition to the already discussed outputs of the hackathons, at each progress was also made in thinking about how the technical tool stack should be created. At the ESIP hackathon the concept of a "Drone Data Buddy" application was proposed by Josip Adams of USGS, and attracted enthusiastic support. In short, the proposal was for a service that runs on a drone, captures all required metadata, and on walking into your home lab it would automatically upload and synchronise all data with a cloud or local server.   We have used this concept as a scaffold around which to build what is emerging as the core components of a LANDRS Toolkit. Figure 4 shows a high level diagram of the proposed LANDRS toolkit core components.

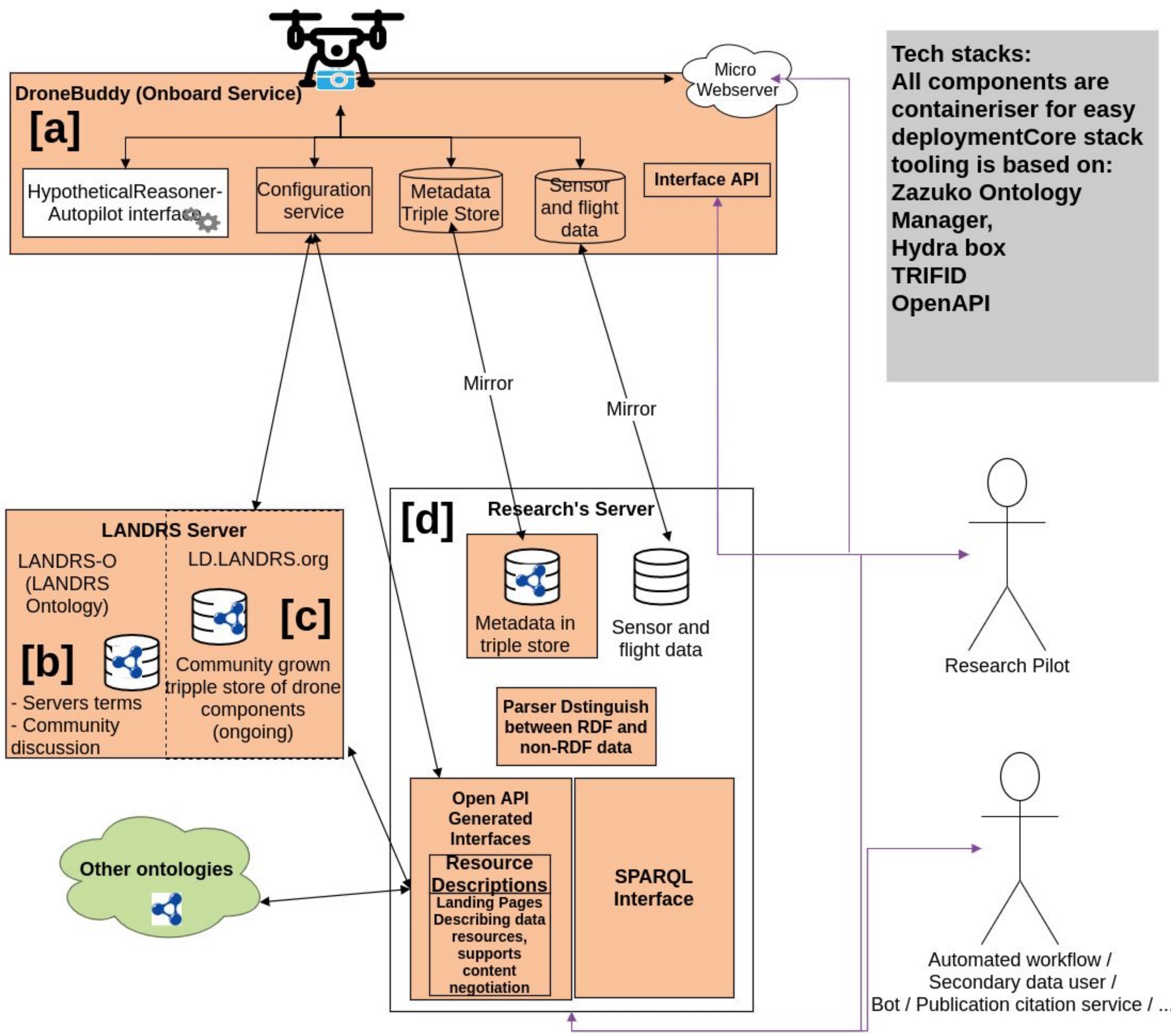

Figure 4: LANDRS-Toolkit-Core component diagram. Components shown as shaded orange are being build by the LANDRS project.

In Figure 4, four high level components are labeled a-d, and briefly described below.   In each case the application and services are containerized for easy deployment by anyone else who wishes to do so. All work is available in the associated github repository under the LANDRS Toolkit organisation https://github.com/landrs-toolkit, each repository has been given a permissible Apache 2.0 license and a basic README. However, as none of these components have yet achieved their first release in many cases documentation or contribution guides are not yet available.

Figure 4 [a] Onboard Drone Data Buddy: The Drone data buddy application runs on a companion computer (eg a Raspberry Pi) onboard the drone. This application provides the basic functionally as first described; it will see data onboard the drone is annotated with the required metadata, and when provided with a network connection these data will synchronise with services on a Researcher's Server Figure 4 [d]). As of the end of February 2020, beyond the Babelfish work no code has been written for this is the component, which exists as a whiteboard diagram. However, this development work will begin in March. The aim is for this toolkit to integrate with the existing open source drone software with Ardupilot and the open Mavlink messaging protocol being the target for this first prototype.

Figure 4 [b] LANDRS ontology: A LANDRS server (a cloud machine currently hosted by Notre Dame) hosts and serves the LANDRS ontology. This is will primarily serve requests from APIs for terms, but is available in a human readable form at http://www.landrs.org/ontology A screenshot is given in Figure 5.

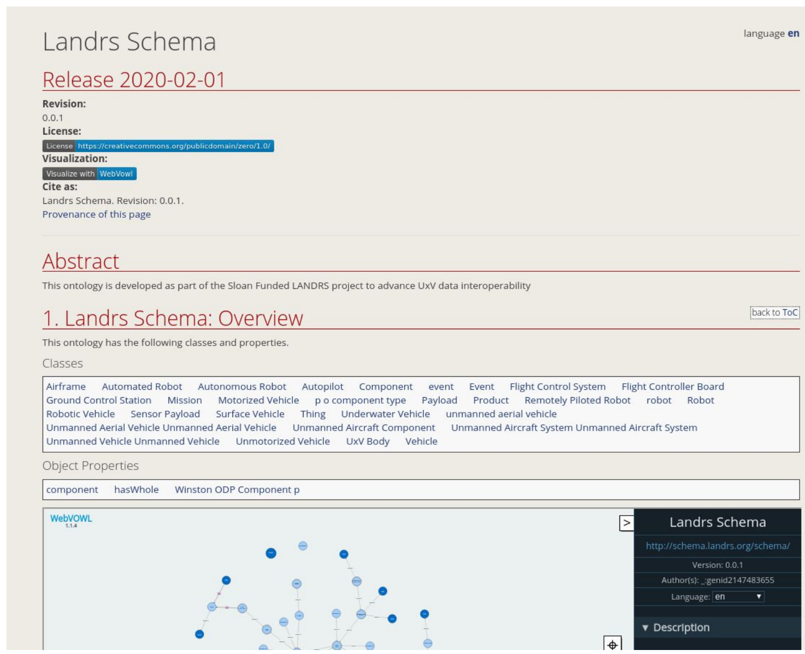

Figure 5: Screenshot of the LANDRS Ontology documentation.

Figure 4 [c] LANDRS ontology development server: The LANDRS server also hosts an instance of the open source Zazuko Fuseki Ontology Manager https://github.com/zazuko/ontology-manager. This is an open source project that provides a means for general communities to discuss and develop ontologies in a collaborative manner. Users can sign in with github credentials, propose new vocabulary terms, definitions and links to other vocabularies. Users can then discuss proposals and vote on their inclusion in the core ontology. We link and reuse terms from existing ontology standards such as schema.org, OGC/W3C Sensor, Observation, Sample, and Actuator (SOSA) ontology, W3C Annotation Vocabulary for images and video observations, W3C Web of Things Description[1]. While we have not yet released (and started advertising for community input) the first version of the

[1] https://schema.org/
https://www.w3.org/TR/vocab-ssn/
https://www.w3.org/TR/annotation-vocab/
https://www.w3.org/TR/wot-thing-description/

ontology, this engagement service is currently available at https://schema.landrs.org/ A screenshot is shown in Figure 6.

As an extension to this project, if resources and time allow we also plan to explore integration of W3C Verifiable Credentials and Decentralized Identifiers[2] that provide verifiable decentralized identity validation.

Figure 4 [d] Drone Data Buddy Server components: The most complex components of the LANDRS toolkit will be run on any researchers' local (or cloud based) server who wishes to utilise the toolkit. These could potentially be operated by a third party (such as a cloud provider or a University computing resources department) on behalf of drone users. Containerisation will make deployment and configuration of these components relatively trivial.

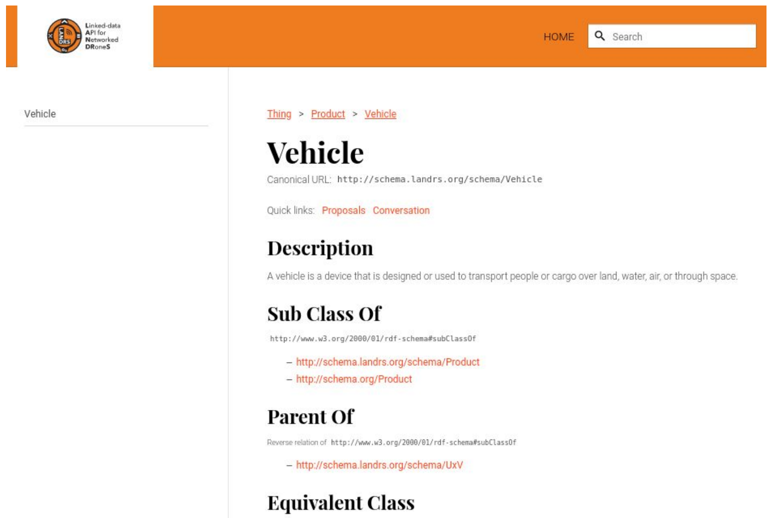

Figure 6: Screenshot of Zazuko Ontology manager as configured for the LANDRS ontology, On release this tool will offer a means of the community commenting on our proposed ontology and knowledge graph as it evolves.

Component Metadata Forms

- To capture information regarding a drone/autopilot/sensor/pilot/mission/project as linked data, human readable forms are machine generated using the latest instance of the LANDRS ontology and multiple levels of data model verification. These provide a way for a user to create reusable linked data metadata components or snippets. For instance, a user will once off create a linked data snippet that describes a drone model they have, and there after reuse that snippet (with a unique identifier) for each instance of that model they have. This development is using World Wide Web Consortium (W3C) published standards called recommendations for the Shapes Constraint Language (SHACL) for validation and the W3C Hydra Community group vocabulary that describes API definitions.
- If this is a popular drone model, they may create it and submit it to a public service (not shown) that will then allow others to reuse it.
- Ultimately, all data captured using that drone, will then be annotated with a uniquely identifiable instance of that metadata snippet. Provided a server is serving the LANDRS

[2] https://www.w3.org/TR/vc-data-model/
https://www.w3.org/TR/did-core/

and other ontologies used, all information related to the drone will therefore be machine (or human) dereferencable.

- As a similar form exists for creating linked data snippets for each sensor/pilot/project/mission all associated important metadata will be automatically attached to the scientific data captured using it. And similarly, all information should remain dereferencable as long as the metadata and ontologies remain being served.

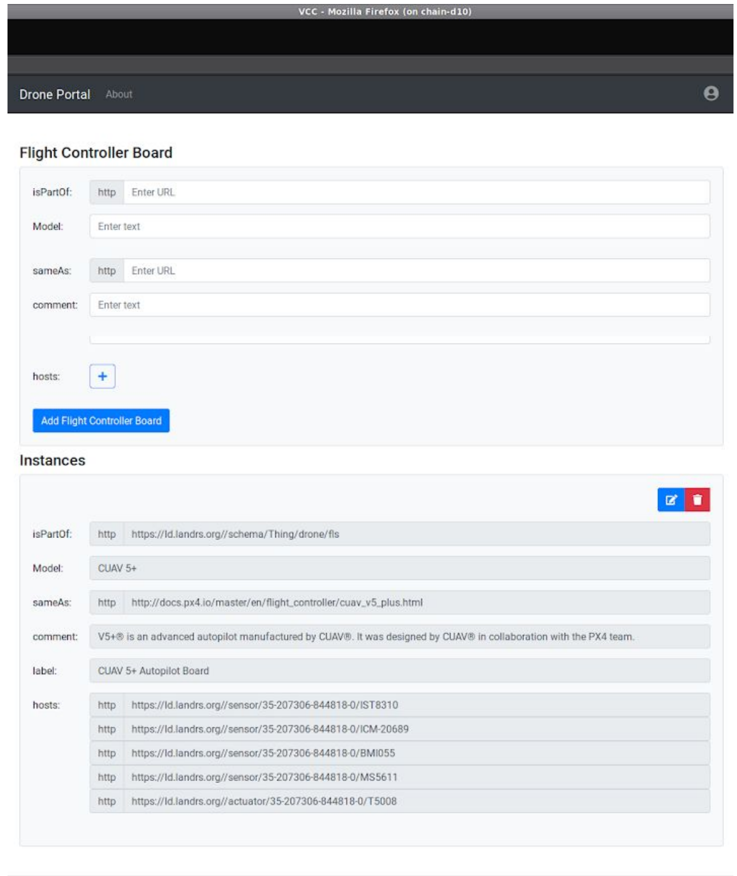

Figure 7 is a screenshot of a generated form for creating a metadata snippet for a flight controller board. This service is not yet live, but is being developed in: https://github.com/landrs-toolkit/landrs-stack-schema

3 Challenges and Evolved Plans

Undeniably the hardest part of this project so far has been convincing people to put time into working with us. While there appears to be no shortage of groups wanting what we are building (again see section 3.1.3). We believe this to be primarily because we are working at the bleeding edge of semantic technologies (a very short skilled domain), and at the intersection between domain scientist, computer scientist, drone or sensor technician, and data expert, this work is not core enough to anyone's respective discipline to be attractive to a researcher needing publications. Although we could most likely have done a better job at advertising ourselves too. It is hoped that engagement at the upcoming RDA Melbourne plenary, visits to Australian drone researchers, employment of the community fellow, the July ESIP hackathon where all components will be showcased in beta release, and the above discussed impact of having tangible products for people to interact with, will improve our community building efforts.

The following sections discuss the concrete steps now envisioned for the latter half of the project.

3.1 Project Timeline

The Gantt chart in Figure 8 shows an updated plan. The milestones listed are discussed below under the headings of: Phase2 Development Part2, Hackathons, and Phase3 Deployment.

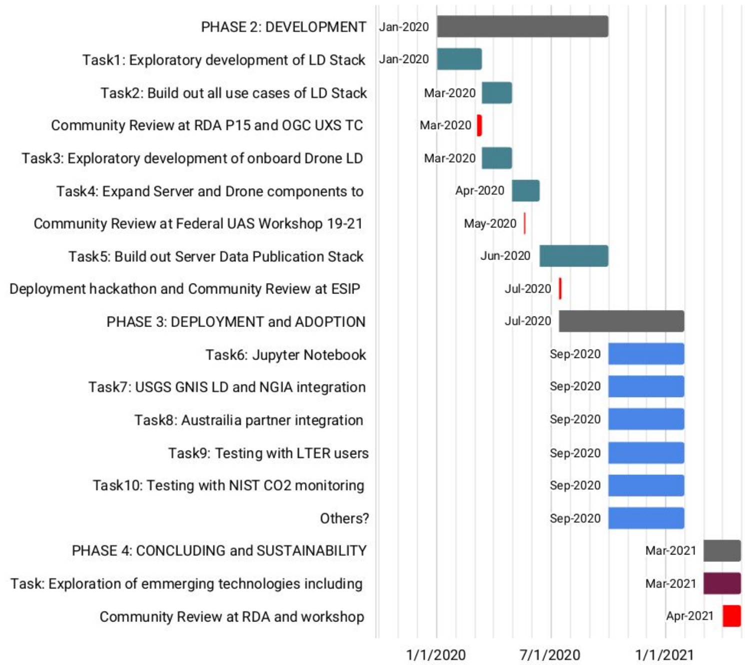

Figure 8: March 2020 - April 2021 gantt chart

3.1.1 Phase 2 Development Part 2

The technical milestones planned for the LANDRS toolkit are outlines as tasks in the gantt chart.

The July ESIP hackathon is intended to serve as the first major in person public release of many of the LANDRS toolkit components. In the lead up to this, however, community lists and other forums will be used to publicise parts as they achieve beta status. Additionally, 3 opportunities for community review are already planned (more will likely emerge in the near future):

- March 2020: Invited to present (remotely) at the March 2020 OGC technical Committee meeting joint session by the UxS and Aviation domain working groups.
- May 2020: Invited to present (in person) at the Federal UAS Workshop.
- July 2020: Invited to present (in person) at the July OGC technical Committee meeting joint session by the UxS and Semantic Data on the Web domain working groupsHackathons
- March 2021: RDA P17

3.1.2 Hackathons

RDA P15

Originally it was proposed to host a hackathon at the RDA P15 in Melbourne. However, based on the above difficulties in attracting the needed skills, and the expense and distance of Australia, we have decided not to do this. The PI will present LANDRS at both the RDA meeting and at the co-located C3DIS meeting, and additionally to researchers at Western Sydney University's Hawkesbury Institute for the Environment. We have repeatedly attempted to recruit an Australian based research fellow, it is hoped that this trip will finally achieve this goal.

As an alternative, it is now our intention to identify a suitable co-located meeting outside of the USA during our deployment phase (ideally within the period November 2020 - February 2021). At that point we will have a full toolkit to demo, and a growing number of adoption demonstrators to compliment it. Two possibilities currently in view are either the Ardupilot Developers workshop which historically takes place in Canberra Australia in March, or The African Drone Forum which historically takes place in February. The primary goal of this event would be to increase the number of adopters and thereby attract developers interested in ongoing maintenance development and feature growth beyond the end of the project.

ESIP July 2020
We are in the early stages of planning and discussion. This meeting will be in Burlington where the University of Vermont's Spatial Analysis Lab (SAL) is a mature drone user group. We intend to both host training in using the LANDRS tools, and to collaborate with SAL to perform demonstration flights at this event. Once minimum details are in place we will again advertise widely for appropriate attendees.

RDA P17 March 2021
While this will be the concluding event of the project, growing collaboration (currently informal) with various EU based groups suggests we will be well positioned to have a larger EU contingent than was possible in Helsinki.

3.1.3 Phase 3: Deployment

Currently 3 adopter-user-groups are earmarked as likely first adopters with whom we will work to develop demonstrator applications.

1. In 2019 collaborators at the University of Botswana received drone equipment from the University of Manchester to explore use in partnership with rangers at a local Rhino Sanctuary. At a partner engagement event (as part of the RDA sUAS Data IG at IDW in 2018), detailed discussions with park rangers over two days lead to a clear needs assessment and evaluation of what role drones might best play in their work. Common to most needs was the use of image streaming from drones. We are discussing how we can create a demonstration of LANDRS support for such with current LANDRS fellows.
2. The USGS has recently developed a linked data instance of their geospatial names information service. We are discussing how we might demonstrate incorporating their name service into the ontologies accessible to a Drone Data Buddy for USGS data collection.
3. The University of Notre Dame currently has NSF funding to work with NIST researchers to improve their CO_2 on drone monitoring systems. We will apply LANDRS toolkits to their application in the improved system we develop for them.

In addition to above groups, based on the work underway with the LTER information modelers, we have requested to demonstrate the LANDRS toolkit to their drone users with the intention of working with 1 or more of them to use LANDRS to automate the data pipeline between their drones the the LTER data archives. Additionally, at the ESIP 2020 winter meeting, the concept of developing direct integration between LANDRS services and the newly developed Jupyterlab metadata service and data explorer[3] arose. This would allow users to create drone data processing pipelines in a Jupyter notebook and both access LANDRS annotation on drone data and add workflow provenance annotations to their drone data processing pipeline.

4 Conclusions

In summary, the originally proposed project outcomes are being addressed as follows:

1. A published and complete open source API for sUAS data
    - Has been designed through community engagement and prototyping
    - Is well under way in development by both professional programmers at Notre Dame and external student fellows
    - A plan is in place for the completion of its current design after which extensions (including potentially the use decentralised identifiers and 5G) will be evaluated according to the responses we receive through the planned iterative community reviews.
2. An RDA endorsed set of sUAS data best practice recommendations
    - This has proved very difficult as the original plan was not embraced by the community.
    - However, the opportunity to do a component of this work with the LTER is allowing progress in-spite of the RDA setback
    - As discussed we do not believe the RDA community, if presented with LANDRS developments at this point, would continue to be uninterested.
    - The prospect of a RDA working group to develop RDA endorsed best practices will continue to be raised. It is hoped that either new community members will provide the support needed in the next 12 months, or perhaps another body will work with us to accomplish this goal (the OGC UXS group perhaps for instance), or potentially once LANDRS is completed and multiple groups have begun using it, when presented at the final RDA event in March 2021 there will then be interest.
    - Regardless, we will have produced the practices adopted by the LTER in their data archives.
3. At least three open source applications using the API to meet specific disciplinary needs
    - As discussed there are already five applications waiting to be developed, 4 of which have immediate user communities.

[3] https://github.com/jupyterlab/jupyterlab-metadata-service
https://github.com/jupyterlab/jupyterlab-data-explorer

- The challenge in completing this task lies primarily in sourcing the needed skill. Preferably we will source fellows from adopting organisations. However, if this proves to not be possible, we will re-evaluate which applications should be prioritised and select the most critical three to develop with professional programmers.