

# Complexities of Data Exchange in the World of Open (Aviation) Data

Kim Miller

INTERNATIONAL AVIATION SAFETY FOUNDATION

iSTARS USER GROUP (iUG/o1) December 17-19, 2018

# Complexities of Data Exchange in the World of Open (Aviation) Data

- Significance of Open Aviation Data
- Open Data by Description
- Limitations and Liabilities
- Management of Expectations

# Significance of Open Aviation Data

Data relating to aviation activities can carry several significant responsibilities and/or liabilities not present in typical open data. In varying conditions as “open,” such aviation data may also include information with specific legal standing for regulators, industry and the traveling public such as:

- Certifications
- Records
- Instructions (advisory/mandatory)
- Approvals
- Reference materials/maps/navigation
- Licenses
- Registrations
- Permits
- Authorizations
- Programs

# Significance of Open Aviation Data

## Considerations:

There is a wide range of aviation data that may or may not be available within each International Civil Aviation Organization member State.

The national legislation and regulation for data of each State will establish which data elements are considered open and eligible to be shared.

Those sharing data are obligated to determine what restrictions may be in place.

# Open data by description

For the purposes of today's discussion, the term "open data" will refer to publicly available data structured in a way that enables the data to be fully discoverable and useful by end users. In general, the data should be:

- *Accessible*
- *Descriptive*
- *Distributable*
- *Complete*
- *Timely*
- *and With a Responsible Contact*

# Data Accessibility

To the extent permitted by law, the data formats should be non-proprietary, publicly available, and have no restrictions placed upon their use.

The data should be available in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched.

Ideally, the format should be machine-readable (i.e., structured to allow automated processing, or scraping).

The data structure used should not discriminate against any person or group being made available to the widest range of users for the broadest range of purposes.

# Data Accessibility

## Considerations:

In project design, consider how the data will be accessed by both provider and recipient.

Will the connection be over an “open” web connection, secured under a VPN or maybe using a new technology platform, such as Blockchain (various national requirements for connectivity)?

Consider what equipment the end user will have to process the data (some may still be on a 10 year old computer)

What will their connection method be, i.e. network, wi-fi, satellite or dial-up? (Not everyone has a good web connection.)

# Data Accessibility

## Considerations:

To the extent permitted by law, these formats should be non-proprietary, publicly available, and their use be unrestricted.

The development of data sharing tools would require an awareness and accommodation of any conditions or restrictions by each of the governments that are hosting the data user.

As discussed, ideally formats should allow access and be machine-readable (i.e., automated processing/scraping). However, recognize that national requirements for readability by assistive devices do vary from country to country that could impact collection.

# Data Description

It is critical that the data be described accurately so data users have sufficient information to understand its strengths, weaknesses, analytical limitations, legal standing, and security requirements.

The description should include the use of robust, granular metadata (i.e., fields or elements that describe data), including comprehensive documentation of data elements and dictionaries.

Where appropriate, there should be descriptions of the purpose of the collection, the population of interest, the characteristics of the sample, the method of data collection, as well as how to process it.

# Data Description

## Considerations:

It is critical that the data be described in formal documentation providing sufficient information to understand the strengths, weaknesses, analytical limitations, security requirements and any legal considerations. Formal documentation encourages appropriate use of the data.

The data description should involve the use of robust, granular metadata, clear delineation of data elements, provide data dictionaries and, as a matter of practice, descriptions of the purpose of the collection, the population of interest, the characteristics of the sample, and the method of data collection.

# Data Distribution

Open data should be made available under an open license that places no restrictions on their use.

Data should be provided in a form and manner that permits redistribution, including the mixing with other datasets.

It is critical that the data be established in a manner to make it interoperable with a wide range of diverse data datasets.

# Data Distribution

## Considerations:

Interoperability of datasets is critical in providing the ability of different datasets to work together.

The ability to freely combine datasets will enable the establishment of a larger more effective data resource.

Practicing true interoperability in design also prevents building silos, which typically results in the inability to combine datasets into larger systems of value.

# Data Complete

Open data should be published in primary forms (i.e., as collected at the source), with the finest possible level of granularity that is practicable and permitted by law and other requirements.

Derived or aggregate open data should also be published, but must reference the primary data.

## Caution!

The best derived or aggregate data will not necessarily over-rule the primary source data (i.e. legislation, regulatory certification, approval, license, registration, etc.).

# Data is Timely

Data should be made available as quickly as practical to preserve the value of the data to other users and systems.

Frequency of release should ideally account for key audiences and downstream needs when possible.

# Data is Timely

## Considerations:

Establish and publish a policy on method and the frequency of data updates so other users can establish a level of confidence as to the currency of the data.

Data that is updated only as conditions change needs to be identified as such to other users.

Users should be notified of significant delays in publication of scheduled updates.

# Limitations and Liabilities

Aviation data could carry significant limitations and liabilities.

Considerations:

Differences in content between primary data and actions taken using aggregate data may have significant legal considerations.

Who is responsible for the unauthorized release of protected or restricted data?

Who will be the party responsible for instances of noncompliance resulting from the utilization of erroneous secondary data?

Who will be the party responsible for the use of secondary data that results in the loss of life or property?

# Responsible Contact

It is important to designate a point of contact (POC) to assist with data use and to respond to complaints about adherence to the “open” data requirements.

Consideration:

The POC should have a resource in regards to questions regarding any direct or indirect legal issues in regards to utilization of the subject data, i.e. governmental law or regulations, licensing, copyright, security, or theft of data

# Management of Expectations

## Considerations:

The activity or result from the data needs to be clearly defined as a part of the initial project development process.

The utilization of aviation data requires the active management of realistic expectations with leadership in what the data can provide or represent.

It is important for leadership to accept that some things will be limited by fundamental differences beyond data.

Thank you