Open software and open science in air transportation

Open tools for Air Traffic Management modelling and research workshop

Xavier Olive

7 June 2024, University of Westminster

- **Crowdsourcing**: it is possible to collaborate and enrich large-scale databases
- Research and industry leverage capabilities to harness **larger amounts of data**
- Shift from model-informed science to data-informed science
- In academic papers:
    - you expect proofs for theorems
    - you expect description of frameworks and models
    - for data-based science, you expect …?

Things are moving fast, but some institutions are careful:

- there are still valid legacy practices
- intellectual property

Is it enough to submit a research paper based on closed data and say "I swear it works"?

Smells of things that have gone wrong:

- closed **by default**: *"Let's sign an NDA"* (to not disclose what?)
- employees in some institutions don't have access to their own data
  (unless they go to sport, music, etc. with the right colleagues)
- public institutions (+ academia) pushing to monetize proof of concepts

Open {software, science, data}

Academia, public institutions are paid with public, taxpayer money

It is fair that we produce knowledge and tools accessible to the public

We apply science on sensitive data, but should **communicate with open data**

Controversial take: there is no such thing as free software
There is always someone paying somewhere:
taxpayer money, developer's private time, ads, "you are the product", etc.

Open software represents a mindset and a matter of best practices which suit very well the conducting of proper science

ReScience (Reproducible Science)

ReScience (Reproducible Science)

Publish independent reproductions of previously published computational studies

ReScience (Reproducible Science)

Publish independent reproductions of previously published computational studies

Focus on reproducibility in scientific research:

validate, verify, build upon existing computational findings

crucial for the credibility and advancement of science

ReScience (Reproducible Science)

Publish independent reproductions of previously published computational studies

Focus on reproducibility in scientific research:

validate, verify, build upon existing computational findings

crucial for the credibility and advancement of science

- **open source software**:
  good baseline in research, provide the scripts to reproduce your figures and tables
- **open data**: provide your datasets to reproduce your research
  if complicated, show how your code works on open datasets, then illustrate on closed ones
- **open practice**:
  - don't be shy: poor code is better than no code
  - sign your contributions and reviews (it helps to behave)

- **Ownership**: the original creator, the maintainer, the employer
- **Authorship**: whoever contributes to the project holds authorship
  Must be acknowledged: respect authors and contributors
- **Licence**: what others can do with the software.
  For example, whether they can use it for commercial purposes, modify it, or redistribute it.
  By default, nothing. **Always add a licence file**: MIT, GPL, Apache, or custom

I would love my code to be open, but my employer does not let me

- **Ownership**: the original creator, the maintainer, the employer
- **Authorship**: whoever contributes to the project holds authorship
  Must be acknowledged: respect authors and contributors
- **Licence**: what others can do with the software.
  For example, whether they can use it for commercial purposes, modify it, or redistribute it.
  By default, nothing. **Always add a licence file**: MIT, GPL, Apache, or custom

I would love my code to be open, but my employer does not let me

1. Listen to your employer
2. Consider writing in your proposals that the output of the project will be open
3. You can open part of the project, and keep sensitive parts as extensions

Open sources of data

Some public datasets

https://atmdata.github.io/sources/

- Flight schedule data, by ANSP, airports, airlines
  through developers API
- Aircraft related data
- Airports, airspaces, ATS routes and navigational points
  FAA Data Delivery Service, Wikidata, OpenStreetMap
- Weather (meteorological agencies)
- Noise and air quality
- Data papers by researchers (Zenodo, figshare, Data in brief, etc.)

- EUROCONTROL Demand Data Repository (discontinued for academics)
- EUROCONTROL Aviation Data for Research
  https://www.eurocontrol.int/dashboard/rnd-data-archive
- FAA Data Portal https://www.faa.gov/data
  https://adds-faa.opendata.arcgis.com/datasets

Specific license terms needed: we are researchers, we want to publish reproducible results!
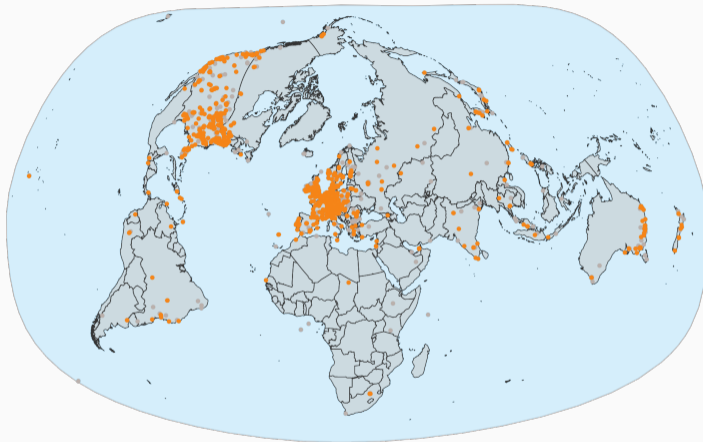
- 4855 registered receivers worldwide (1500+ always online)
- Most coverage in Europe and the US; 900 major airports with ground coverage

The Open Aviation Data Initiative

- Making extracted information accessible to a non-specialist audience (general public, data enthusiasts, students, newcomers in the field)
- A(n open) team of open data evangelists
- A collection of **Observable** interactive browser-based notebooks https://observablehq.com/@openaviation
- A set of open-software tools (mostly Python, Rust and Javascript)
- An open-access APC-free academic journal https://journals.open.tudelft.nl/joas/
- An open-access book project (*hatching*) https://aviationbook.netlify.app

- Big publishers
- Wide range of journals
- mostly closed journals
- some paid open-access, but costly

- Domain specific publishers
- Closed journals with paid open access
- US centric

- Full open access publishers
- Lack of aviation specific journal
- Costly publication fees
- [MDPI] many low quality journals

Journal of
Open
Aviation
Science

About the journal

Journal of Open Aviation Science is an open-access peer-reviewed journal, established by researchers for researchers. It promotes high quality open science, open data and reproducible scientific research for aviation, as such practice is largely missing in current aviation research. Our primary objectives are:

- to promote reproducible research in the aviation sector using open source tools and open data; and
- to encourage the reproduction of existing research in aviation research domain.

In addition to original scientific research, we welcome replication of existing research publications using open source implementations and open data. Particularly, we welcome contributions comparing the performance of reference models with their open source counterparts.

Each submission is carefully and openly reviewed, based on the quality of the manuscript and on the reproducibility of research. Tables and figures must be submitted with accompanying publically available code, explanations, tests and data. Code shall be tested in order to guarantee that any researcher can re-use it.

We welcome the following types of papers

⚖ Research article	🗃 Research article - data	🕸 Research article - software	↻ Re-Science article
General aviation research article with open data and open code.	Research articles focus on a significant open dataset in the aviation domain, including processes and applications.	Research articles focus on open-source tools for aviation studies.	Research articles focus on reproducing results from previous aviation-related research and sharing open code and data.
Submit	Submit	Submit	Submit

Some stats (as of June 3rd)

- **Research articles**: 2 published, 2 declined, 1 under review
- **Re-science article**: 1 draft (not submitted)
- **Data paper**: 1 under review
- **Software paper**: 2 published
- **OpenSky Symposium**: 9 poster papers, 14 full papers, 1 declined, 1 under review
- approx. 2 months between submission and publication
- 2 or 3 reviewers per paper, 39 unique reviewers from 7 countries
  (BE, CH, DE, FR, GB, JP, NL, US)
- all reviews are signed and published

A set of open-source tools

aviationbook
"A journey through aviation data", an open-access book
- TeX ☆ 6

pyopensky
The Python interface for OpenSky database
- Python ☆ 45 ⑂ 10

tangram
An open platform for modular, real-time aviation research
- Python ☆ 7 ⑂ 1

pitot
A toolbox for aeronautic calculations
- Python ☆ 7 ⑂ 1

TUDelft-CNS-ATM/bluesky
The open source air traffic simulator
- Python ☆ 335 ⑂ 236

UoW-ATM/Mercury
No description provided
- Python ☆ 3

AeroMAPS/AeroMAPS
AeroMAPS: Multidisciplinary Assessment of Prospective Scenarios for air transport
- Jupyter Notebook ☆ 15 ⑂ 3

fast-aircraft-design/FAST-OAD
FAST-OAD: An open source framework for rapid Overall Aircraft Design
- Python ☆ 44 ⑂ 23

Open software contributions

xoolive/traffic

A toolbox for processing and analysing air traffic data

● Python ☆ 350 ⑂ 82

junzis/openap

Open Aircraft Performance Model and Python Toolkit

● Python ☆ 80 ⑂ 24

xoolive/cartes

Create great maps in Python 🐍 with openstreetmap 🌍

● Python ☆ 43

junzis/openap-top

Open flight trajectory optimizer built with non-linear optimal control method

● Python ☆ 15 ⑂ 6

junzis/pyModeS

Python decoder for Mode S and ADS-B signals

● Python ☆ 517 ⑂ 151

xoolive/rs1090

Decode Mode S, ADS-B and FLARM signals in Rust/Python

● Rust ☆ 8 ⑂ 2

achevrot/impunity

A Python library to check physical units

● Python ☆ 9

ramondalmau/metafora

No description provided

● Python ☆ 8

Difficult questions:

- **Who will maintain your project?**
  What happens after you have new interests, a new position, a new life?
- It is very easy to let such projects take over your personal life
- Academia should value open software/open dataset as much as they value papers
- **Any good project will be superseded by a better one**

Some hints about how alive is a project:

- Stars on GitHub
- Date of the latest commit

Those are only indications, such metrics are **broken by nature** anyway

Conclusion

- Open {software, science, data} is good practice in academia

- Open science is compatible with industrial constraints

- The Open Aviation Data Initiative is not a closed club!
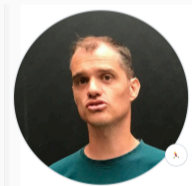
Xavier Olive
xoolive

I like maps 🗺️, code 💻 and data visualisation 😎. I keep an eye on anything that can fly ✈️.

Junzi Sun
junzis

Assistant Professor at TU Delft, conducts research in the areas of air traffic management, aircraft surveillance, performance modeling, and data science.

Enrico Spinielli
espinielli

I am interested in dataviz and am looking forward to a world where human brains are freed from boring, repetitive, manual, click-copy-paste tasks.

Benoit Figuet
figuetbe

Co-founder of SkAI Data Services and researcher at ZHAW Centre for Aviation. Passionate about data science, machine learning, and risk modeling.

You!
aeronerd