Plan Overview

A Data Management Plan created using DMPTool

Title: BioNORAD: Fast Scalable Pandemic Risk Assessment of Influenza A Strains Circulating In Non-human Hosts

Creator: Ishanu Chattopadhyay

Affiliation: University of Chicago (uchicago.edu)

Principal Investigator: Ishanu Chattopadhyay

Contributor: Balaji Manicassamy

Funder: Congressionally Directed Medical Research Programs (cdmrp.army.mil)

Grant: https://cdmrp.health.mil/funding/pa/HT9425-23-PRMRP-DA-GG.pdf

Template: NIH-GEN: Generic (Current until 2023)

Project abstract:

Animal influenza viruses emerging into humans have triggered devastating pandemics in the past. Yet, our ability to evaluate the pandemic potential of individual strains that do not yet circulate in humans, remains limited. Here we propose to develop an experimentally validated platform called the
Emergenet (Enet), to predict in near-real-time where and when new variants of concern would emerge, using only observed sequences of key viral proteins, procured in ongoing global surveillance of Influenza A viruses. We bring together new machine learning algorithms customized to the problem at hand, key insights from information theory, evolutionary theory, epidemiology and precise statistical uncertainty quantification
to develop a rigorous framework, to track evolutionary trajectories of pathogens through a complex, poorly characterized, and dynamically changing fitness landscape. Our plan is to analyze close to 400K strains of curated Influenza A strains in public databases to learn actionable patterns of

mutational change that then enables predictive reasoning to forecast the likelihood of a species jump and attainment of human-to-human transmission capability for strains that do not yet circulate in humans. Our deliverable is best described as the foundations for creating a platform akin to bio-NORAD, identifying when and where an imminent zoonotic emergence event is likely, and if such novel strains are likely to achieve human-to-human transmission capability.

Start date: 10-01-2023

End date: 10-20-2025

Last modified: 05-10-2023

Copyright information:

BioNORAD: Fast Scalable Pandemic Risk Assessment of Influenza A Strains Circulating In Non-human Hosts

Data sharing plan

How do you plan to provide access to your data?

Computing Environment: The UChicago computing environment provided by the Center for Research Informatics (CRI) will be sandboxed from the internet as well as other servers and data sources at UChicago. It will accessible only to the PI, research assistant(s), software developer(s), and/or system administrator(s) who require access during the course of the project.

Box: Research data will be stored and preserved for the duration of the grant using Box, which uses AES 256-bit encryption and is also FedRamp authorized and HIPAA compliant (https://www.box.com/security). Box provides file versioning, which helps mitigate issues such as file corruption. UChicago is committed to using Box as the institutional cloud storage tool. If the university were to switch cloud storage solutions, we will meet the security needs of this and all other grant work supported by Box.

When will you make the data available?

Data and research resources generated in this project research will be made available to the research community, which includes both scientific and consumer advocacy communities, and to the public. This includes all data and research resources generated during the project's period of performance, including:

- **Unique Data**, defined as data that cannot be readily replicated. For this project, examples of unique data include curated models of genomic change for different sub-types of Influenza A, for different geographical locations.
- **Final Research Data** defined as recorded factual material commonly accepted in the scientific community as necessary to document and support research findings. In our context, examples are sequence ids of strains we use for our modeling, and the particulars of validations experiments, including the metadata needed to replicate those experiments in the laboratory.
- **Research Resources** include, but are not limited to, the full range of tools that we would develop and use in the laboratory. In this project, such resources include all developed software for modeling and prediction.

Which archive/repository/central database have you identified as a place to deposit data?

We will deposit software in Github repositories, allowing easy installation of such software in compatible systems. We will also deposit models, metadata and software copy at Zenodo for long-term

citable access to the research resources and products.

Will a data-sharing agreement be required?

No data sharing agreement is required for this project, since the underlying data on which we will learn our models are publicly accessible with minor restrictions.

What metadata/documentation will be submitted alongside the data?

Complete enumeration of sequence ids as obtained from NCBI and GISAID will be submitted, which is sufficient to replicate the results if using our developed software. Also descriptions of inferred and curated models will be made available. Example software programs based on our open-source library will be provided as well.

What file formats will you use for your data, and why?

No specialized file format is necessary for this project. All files will be shared as text files, csv files or compressed versions of those.

What transformations will be necessary to prepare data for preservation/data sharing?

No specialized transformation is necessary.

Do you need funding for the implementation of this data sharing plan?

The effort of the postdoctoral associate funded on this project will carry out the requirements of this plan, and his salary will be partially covered under the proposed budget.

Planned Research Outputs

Software - "Emergenet Python Library"

Software to learn mutational patterns from sequence databases, and make prediction on mutational change

---

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Emergenet Python Library	Software	2025-03-19	Open	None specified		GNU General Public License v3.0 or later	None specified	No	No