



A Global Initiative On Sharing All Influenza Data

www.gisaid.org

Joachim Büch^{1*}, Kirsten Roomp^{1*}, Gunter Bach², Lars Steinbrück¹, Martin Beer⁹, Victoria Gregory¹⁰, Naomi Komadina⁴, Yu Lan⁷, Isabella Monne⁸, Catherine Smith³, Seiichiro Fujisaki⁶, Peter Bogner⁵ & Thomas Lengauer¹

¹ Max-Planck-Institute for Informatics, Saarbrücken, Germany, ² a3 Systems GmbH, Saarbrücken, Germany, ³ Centers for Disease Control and Prevention, Atlanta, GA, USA, ⁴ WHO Collaborating Centre for Reference & Research on Influenza, Melbourne, Australia, ⁵ GISAID Foundation, Washington, DC, USA, ⁶ National Institute of Infectious Diseases, Tokyo, Japan, ⁷ Chinese National Influenza Center, Beijing, Peoples Republic of China, ⁸ Istituto Zooprofilattico Sperimentale Delle Venezie, Padua, Italy, ⁹ Friedrich-Loeffler-Institute, Greifswald - Insel Riems, Germany, ¹⁰ National Institute for Medical Research, London, United Kingdom, *Joint first authors

BACKGROUND

The GISAID initiative offers an alternative to current public-domain database models in response to growing needs of the global influenza community for the sharing of genetic sequence and associated epidemiological and clinical data of all influenza strains. GISAID's publicly accessible EpiFlu™ database is governed by a unique sharing mechanism that protects the rights of the submitter, while permitting ongoing research as well as the development of medical interventions, such as drugs and vaccines.

OVERVIEW

For the GISAID Initiative, the Max-Planck-Institute for Informatics, Saarbrücken, Germany, has developed a web portal that is accessible at www.gisaid.org featuring the GISAID EpiFlu™ database that offers a unique collection of virus influenza sequence data. The database is based on software by Oracle and the dante® System by a3systems GmbH, Saarbrücken, Germany. Extensive metadata are also collected for most isolates. The database provides features for searching, filtering specific datasets for download and user friendly upload functionality. To uphold GISAID's unique sharing mechanism, all users must positively identify themselves. While access is free of charge, all users agree that they will not attach any restrictions on the data, but will acknowledge both the originator of the specimen and the submitter of the data, and seek to undertake to collaborate with the submitter.

All uploaded sequence data are submitted to rigorous curation.

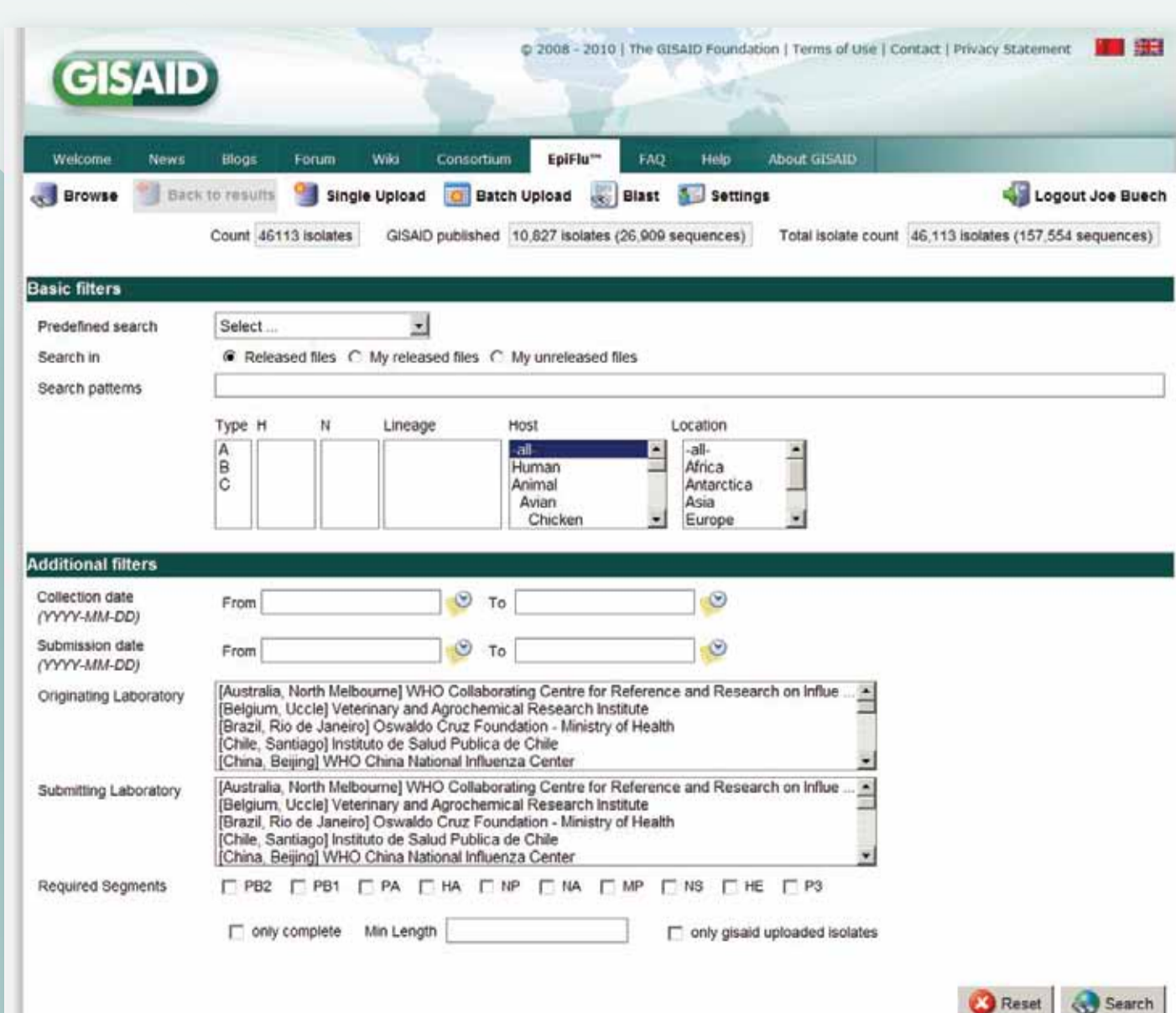


Figure 1. The browser menu

FEATURES IN DETAIL

- Contains influenza sequences and associated meta data with each isolate
- Genetic, clinical, epidemiological & geographical data for human isolates plus species specific data associated with non-human isolates
- Batch and single upload functions
- Ability of submitter to edit data submitted to the platform
- Each isolate accompanied by an audit trail with a history of edits
- Submitted data available for view to other users immediately
- Download facility of meta data associated with isolates in an Excel format
- Sequence download in fasta format with user defined headers

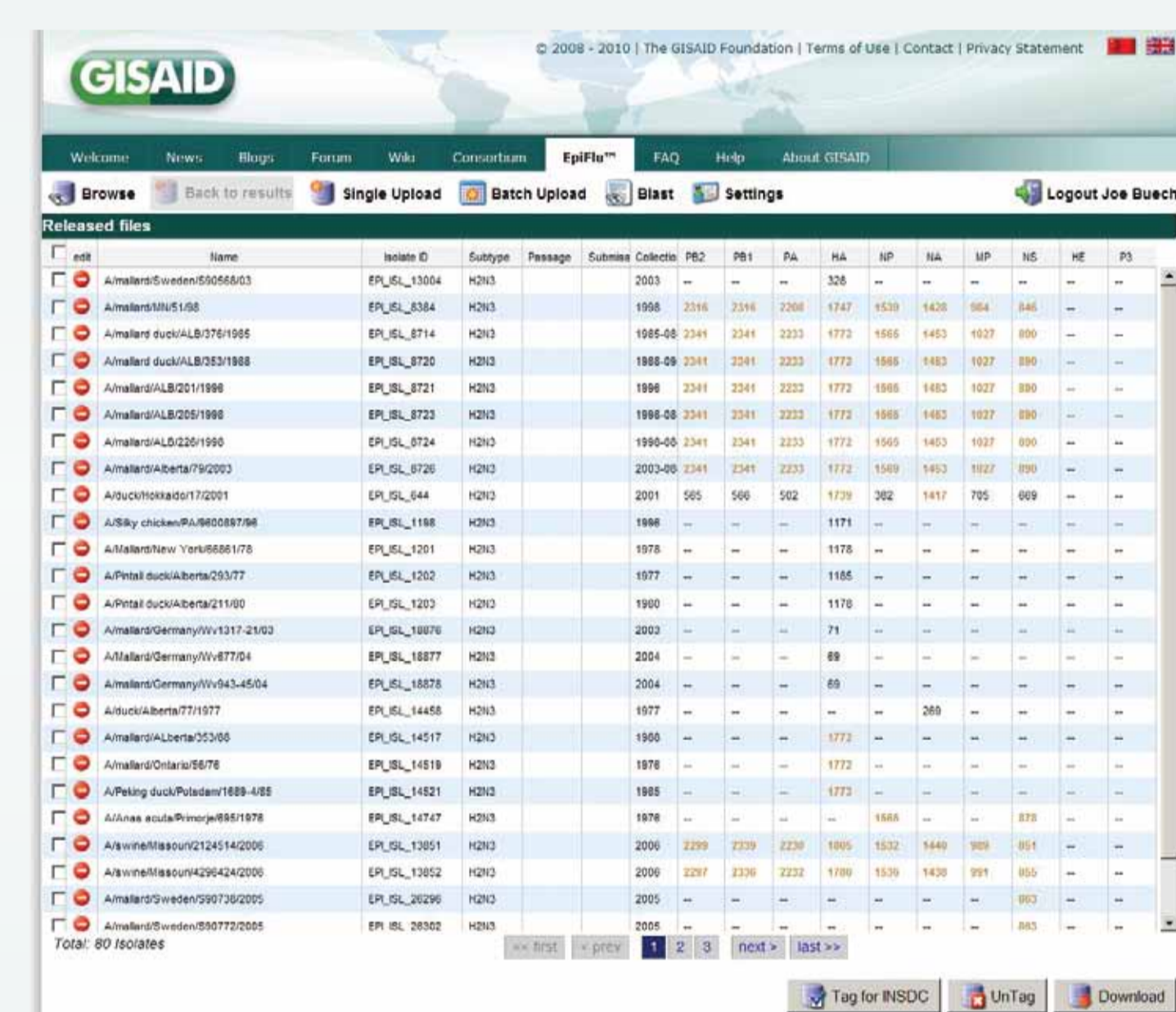


Figure 2. The search results page

Browse Functions

- Can choose to browse all isolates in the database, user submitted isolates, or isolates not available on other databases (Figure 1)
- Can browse using one or more filters for:
 - Type, subtype, isolate name, genes, species, region, country of origin
 - Lineage swl or seasonal for H1N1, Yamagata or Victoria for Type B influenza viruses
- Isolate originating laboratory
- Isolate sequence submitting laboratory
- Isolate submission date
- Isolate submission date
- Can define fields to be displayed in search results (Figure 2)

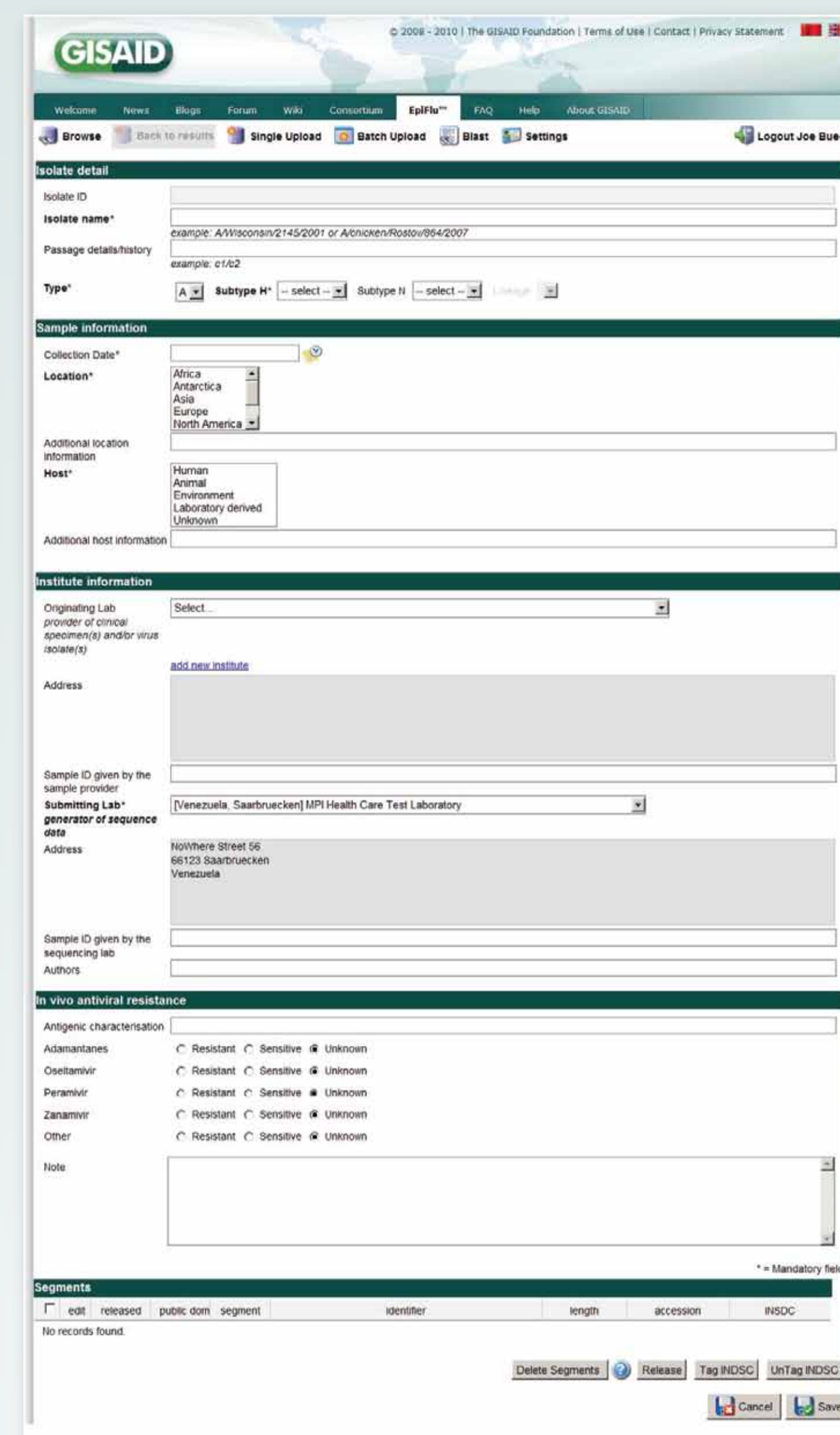


Figure 3. The single upload function

Upload Functions

- Can upload data via single upload function (Figure 3) or batch upload sheet (Figure 4) for multiple isolates
- Meta data only needs to be entered once for each isolate
- Once isolates uploaded, extra gene sequences can be added via the edit function
- For isolates added via the batch upload sheet multiple extra gene sequences can be added by re-using the initial batch upload sheet
- Batch upload occurs in real time, with an immediate response
- Error messages are displayed for isolates & sequences not uploaded
- Duplicate sequences & isolates are flagged and not added
- Successful uploads are flagged and isolate accession and segment accession numbers are added to the return batch upload sheet

Isolate ID	Seq ID	Isolate Name	Subtype	Lineage	Passage	History	Country	Province	Sub Province	Location	Additional Info	Host	Host Additional Info	Seq ID (HA)	Seq ID (NA)
66761	HA: EPI233047 / NA: EPI233046	ANICTORIA/0204/2009	H1N1	swl	Clinical Specimen		Australia	Victoria		Devonport		Human	deceased	0954280	0954281
66762	HA: EPI233048 / NA: EPI233048	ANICTORIA/0209/2009	H1N1	swl	Clinical Specimen		Australia	Victoria		Lancefield		Human	deceased	0954282	0954282
66763	HA: EPI233051 / NA: EPI233050	ANICTORIA/0206/2009	H1N1	swl	Clinical Specimen		Australia	Victoria		Richmond		Human	deceased	0954283	0954283
66764	HA: EPI233052 / NA: EPI233052	ANICTORIA/0205/2009	H1N1	swl	Clinical Specimen		Australia	Victoria		Frankston		Human	deceased	0954284	0954284
66765	HA: EPI233055 / NA: EPI233054	ATAGMANA/0204/2009	H1N1	swl	Clinical Specimen		Australia	Tasmania		Hobart		Human	deceased	0954285	0954285
66766	HA: EPI233057 / NA: EPI233056	ANICTORIA/0206/2009	H1N1	swl	Clinical Specimen		Australia	Victoria		Hobart		Human	deceased	0954286	0954287
66767	HA: EPI233059 / NA: EPI233058	APERITH/0204/2009	H1N1	swl	MDCK		Australia	Western Australia		Busselton		Human	deceased	0954287	0954288
66768	HA: EPI233061 / NA: EPI233060	APERITH/0205/2009	H1N1	swl	MDCK		Australia	Western Australia		Ashfield		Human	deceased	0954288	0954289
66769	HA: EPI233062	ATIAHTI/0204/2009	H1N1	swl	Clinical Specimen		French Polynesia	Tahiti				Human	deceased	0954289	
66769	HA: EPI233041 / NA: EPI233040 / MP: EPI233039	APERITH/01/2009	H1N1	seasonal	E2		Australia	Western Australia				Human		0954294	0954298
66769	HA: EPI233044 / NA: EPI233043 / MP: EPI233042	ANICTORIA/04/2009	H1N1	seasonal	E2		Australia	Victoria		Melbourne		Human		0954295	0954298
66770	HA: EPI233064 / NA: EPI233063	ANICTORIA/0205/2009	H1N1	swl	Clinical Specimen		Australia	Victoria				Human		0954479	0954506
66771	HA: EPI233066 / NA: EPI233065	ANVELLINGTON/0205/2009	H1N1	swl	mdck-284e-sw2		New Zealand	North Island		Waikato		Human		0954511	0954510
66772	HA: EPI233068 / NA: EPI233067	ANAKATO/01/2009	H1N1	swl	mdck-sw2-sw2		New Zealand	North Island		Waikato		Human		0954512	0954511
66773	HA: EPI233070 / NA: EPI233069	ANVELLINGTON/02/2009	H1N1	swl	mdck-sw2-sw2		New Zealand	South Island		Diago		Human		0954513	0954520
66774	HA: EPI233072 / NA: EPI233071	ASAMDA/02/2009	H1N1	swl	mdck-sw2-sw2		Senegal					Human		0954514	0954521
66775	HA: EPI233074 / NA: EPI233073	ANICTORIA/0201/2009	H1N1	swl	MDCK1		Australia	Victoria		Frankston		Human		0954515	0954522
66776	HA: EPI233076 / NA: EPI233075	ANICTORIA/02/25/2009	H1N1	swl	Clinical Specimen		Australia	Victoria		Melbourne		Human		0954522	0954540
66777	HA: EPI233078 / NA: EPI233077	ANICTORIA/02/25/2009	H1N1	swl	Clinical Specimen		Australia	Victoria		Melbourne		Human		0954523	0954541
66778	HA: EPI233079	ANICTORIA/02/26/2009	H1N1	swl	Clinical Specimen		Australia	Victoria		Melbourne		Human		0954526	
66779	HA: EPI233081 / NA: EPI233080	ANICTORIA/02/29/2009	H1N1	swl	Clinical Specimen		Australia	Victoria		Melbourne		Human		0954526	0954544
66780	HA: EPI233083 / NA: EPI233082	ANICTORIA/02/30/2009	H1N1	swl	Clinical Specimen		Australia	Victoria		Melbourne		Human		0954527	0954545
66781	HA: EPI233085 / NA: EPI233084	ANICTORIA/02/31/2009	H1N1	swl	Clinical Specimen		Australia	Victoria		Melbourne		Human		0954528	0954546
66780	HA: EPI233045	ANICTORIA/04/03/2009	H3N2	swl	Clinical Specimen		Australia	Victoria		Lysterfield		Human		0954529	
66782	HA: EPI233091 / NA: EPI233088 / MP: EPI233086	ANMALAYSIA/22/2009	H1N1	swl	MDCK/MDCK1		Malaysia					Human		0954535	0954535
66783	HA: EPI233091 / NA: EPI233090 / MP: EPI233089	ANUOLAND/02/2009	H1N1	swl	swl		New Zealand	North Island				Human		0954536	0954536

Figure 4: The batch upload facility, successfully uploaded isolates

RESULTS

As of October 31, 2010, the rapidly growing GISAID dataset comprises 170,478 nucleotide sequences (from 49,801 isolates) with 29,878 (from 12,285 isolates) uniquely submitted to this database. Among its contributors are OIE and National Reference Laboratories for AIV and all of the WHO Collaborating Centers for Surveillance, Epidemiology and Control of Influenza who routinely contribute data in addition to using the EpiFlu™ database for their semi-annual vaccine strain selection. To provide a complete picture of data, all data available in other public databases is routinely imported.

OUTLOOK & CONCLUSION

Starting in 2011, Germany's Federal Ministry of Food, Agriculture & Consumer Protection will be the long-term host of the GISAID platform. Curation and quality control of data will be ensured by the Friedrich-Loeffler-Institute to meet highest quality standards. The Max-Planck-Institute for Informatics will continue to develop the portal and database software and enable GISAID to act as a catalyst for the development of advanced bioinformatics software connected directly to the database. Software development is underway to extend the spectrum of data analysis tools. The functionality of the database will also be expanded to include more datatypes.

GISAID has become an indispensable resource for the international scientific community on influenza. The consortium will expand its activities and offers to catalyze research and development on a wide variety of issues pertaining to risk analysis, drug development and therapy of influenza.