MEDINFO 2013
C.U. Lehmann et al. (Eds.)

doi:10.3233/978-1-61499-289-9-466

Design of Schistosomiasis Ontology (IDOSCHISTO) Extending the Infectious Disease Ontology

Gaoussou Camara[a,b], Sylvie Despres[a], Rim Djedidi[a] , Moussa Lo[b]

[a] *LIM&BIO, Université Paris 13, Bobigny, France*
[b] *LANI, Université Gaston Berger, Saint-Louis, Sénégal*

Abstract

Epidemiological monitoring of the schistosomiasis' spreading brings together many practitioners working at different levels of granularity (biology, host individual, host population), who have different perspectives (biology, clinic and epidemiology) on the same phenomenon. Biological perspective deals with pathogens (e.g. life cycle) or physiopathology while clinical perspective deals with hosts (e.g. healthy or infected host, diagnosis, treatment, etc.). In an epidemiological perspective corresponding to the host population level of granularity, the schistosomiasis disease is characterized according to the way (causes, risk factors, etc.) it spreads in this population over space and time. In this paper we provide an ontological analysis and design for the Schistosomiasis domain knowledge and spreading dynamics. IDOSCHISTO – the schistosomiasis ontology - is designed as an extension of the Infectious Disease Ontology (IDO). This ontology aims at supporting the schistosomiasis monitoring process during a spreading crisis by enabling data integration, semantic interoperability, for collaborative work on one hand and for risk analysis and decision making on the other hand.

Keywords:

Domain Ontology, Process Ontology, Infectious Diseases, Schistosomiasis, Epidemiological Monitoring.

Introduction

In this paper, we present the design of the shistosomiasis' ontology as an extension of the Infectious Disease Ontology (IDO[1]). The ontology is named IDOSHISTO for "Infectious Disease Ontology for SCHISTOsomiasis" and is organized in two modules: "domain ontology" and "process ontology". The building of IDOSCHISTO aims at supporting the monitoring of schistosomiasis. The schistosomiasis is an infectious disease caused by a parasite.

The monitoring of infectious diseases brings together many practitioners in the medical domain working at different levels of granularity (organism, host individual and host population), who have different perspectives (biology, clinic and epidemiology) on the same phenomenon. Beyond the medical domain, epidemiological monitoring involves several organizations (health control and prevention, pharmaceutical industries, socio-political organizations, etc.) all working towards the same target, but at different levels of decision (environmental, population, drug-manufacturing, etc.). This leads to a large number of actors (epidemiologist, biostatistician, pathologist, meteorologist, doctor, parasitologist, public health agent, political actor, etc.) participating in risk analysis and decision-making. These actors may have heterogeneous profiles and use different vocabularies for the same domain concepts which could lead to many misinterpretations. To settle these issues, IDOSCHISTO is designed to facilitate communication, interaction and collaborative work between these actors and organizations. For this purpose, a "domain ontology" module is built in IDOSCHISTO.

The IDOSCHISTO is also intended to be used for handling qualitative simulations. Indeed, modeling of complex systems such as the spreading of infectious diseases has been long focused on reproducing the dynamics of the phenomenon to understand its evolution and make predictions by numerical simulations. These numerical modeling approaches describe physical characteristics of processes. These models, built from descriptive surveys, help explain the dynamics of the spread of disease, and help to validate assumptions. However, the models are hardly usable for prediction and decision-support purposes in monitoring context. Indeed, simulation models are either designed for a very limited target, or require input data that are difficult to acquire in real-time simulations. Further, the diversity of formalisms used to represent these models (regular differential equations, agent-based models, etc.) restricts model composition and interoperability which are needed to answer complex queries. In order to provide an alternative and complementary solution to the monitoring system during the risk analysis and decision-making steps [1], we build a "process ontology" module for schistosomiasis' spreading processes. The ontology-based process modeling approach considers processes as concepts (abstraction). It allows reproducing possible behaviors of a system from the abstract description of its internal processes and its different possible states. Thus, a process ontology specifies classes of processes, relationships between processes, relationships between processes and objects, occurrence conditions of processes, process occurrence effects on states of other processes and objects, etc. Therefore, reasoning on the "process ontology" module of IDOSCHISTO will allow the prediction of possible states or process occurrences, or the explanation of the causes of process or state occurrences.

The paper is organized as follows: in the first section we present the materials used, namely the schistosomiasis knowledge base and the IDO-Core ontology. The second section specifies the methodology for building the IDOSCHISTO. The third section presents the ontological analysis results and models. We present related works before the discussion and conclusion.

[1] http://infectiousdiseaseontology.org

Materials

Domain experts provided the expert knowledge for our model via interviews and scientific papers on the schistosomiasis disease.

Schistosomiasis disease

The schistosomiasis (or bilharzia) is a parasitic disease that constitutes an important public health problem[2] around the world, particularly in tropical areas. The parasite causing the infection grows both in water and in the human organism at different stage of its life cycle. The parasite life cycle follows several stages: egg, miracidium, sporocyst, cercaria, schistosomula, and adult. Each of these stages – requiring an interval of time and a specific environment – represents a state of the pathogen in its life cycle. Transition from one state to another is subject to certain conditions (time or event). Schistosomiasis contamination mode is based on indirect transmission.

Schistosomiasis spreading process

The schistosomiasis spreading process is based on the standard SIR (Susceptible, Infected, and Recovered) model [2]. Furthermore, an Exposure (E) state is included between Susceptible and Infected states (SEIR). It is also possible for a recovered person to lose his immunity against the disease and become susceptible; thus, the complete spreading model for schistosomiasis is SEIRS.

Infectious disease ontology - core

The IDO's core ontology (IDO-Core) contains common entities for all infectious diseases and relevant to biological and clinical perspectives [3]. Several ontologies (for Brucellosis, Dengue fever, infective endocarditis, influenza, malaria and other vector-borne diseases, Staphylococcus aureus, tuberculosis) have been already designed as extensions of the Infectious Disease Ontology (IDO). They provide a formal representation of specific disease domain knowledge to support interoperability and reasoning capabilities. These ontologies, usually called "domain ontologies", focus on describing domain entities and their relationships regardless of the way they unfold specifically for the occurrent [4] entities.

Methods

IDOSCHISTO design approach

IDOSCHISTO is designed to take into account the schistosomiasis domain knowledge and spreading process dynamics. Therefore, these two aspects are respectively represented in the domain and process ontologies modules. Domain ontology modeling focuses on specifying knowledge about domain entities and their relationships regardless of their evolution in time. It provides a common vocabulary and an explicit specification of the domain underlying axioms to facilitate communication, enable semantic data integration, ensure interoperability between applications, and support semantic reasoning [5]. In order to represent the dynamic aspect of the infectious diseases in the "process ontology", we start by analyzing the macro-process, meaning the spreading process occurring at the population scale. We also study its causal relationship with the underlying processes at the individual and biological scales. We then propose a global model of the spreading process based on the interactions between the entities that compose the infectious diseases domain, such as host, pathogen, vector, etc.

IDOSCHISTO design framework

The diverse perspectives on the disease and the multi-scale structure of the spreading process are taken into account respectively in the domain and process modules of IDOSCHISTO. The epidemiological, clinical and biological perspectives in the domain ontology module correspond to the population, individual and biological scales in the spreading process analysis (cf. Schistosomias process ontology analysis section). In what follows, for a purpose of simplification, we keep talking about perspectives except for the spreading process analysis. Therefore, the domain and process ontologies modules are organized following these three perspectives: Epidemiological-Perspective Module (EPM), Individual-Perspective Module (IPM) and Biological-Perspective Module (BPM). Furthermore, relations between entities of different perspectives and entities of the same perspective have to be considered. Thus, we consider two types of relations: Inter-Perspective Relations (Inter-PR) and Intra-Perspective Relations (Intra-PR).

For the domain ontology, each of the perspectives on the schistosomiasis disease is built following a conceptual framework that reuses a foundational ontology, a core ontology and some relevant domain ontologies as recommended in [6]. The organization of the framework is showed in Figure 1. The core level is the IDO-Core ontology of infectious diseases domain which is already reusing the BFO (Basic Formal Ontology) foundational ontology. BFO is narrowly focused on the task of providing a genuine upper ontology [7] which can be used in support of domain ontologies developed for scientific research, as for example in biomedicine within the framework of the OBO (Open Biomedical Ontology) Foundry [3]. In the domain specific level we will reuse existing ontologies from the OBO project and others that fit to the criteria of the epidemiological (spreading process and risk factors, prevention and control), clinical (diagnosis and treatment) and biological (pathology and pathogen life cycle) perspectives.

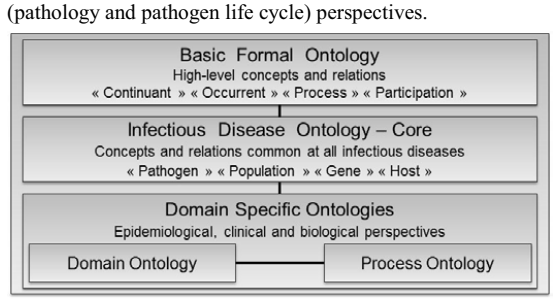

Figure 1- IDOSCHISTO design framework

The process ontology provides the unfolding specification of the complex processes of each perspective by combining the atomic (or less complex) sub-processes, for instance the spreading, contamination and physiopathology processes respectively in the epidemiological, clinical and biological perspectives. At the present time, there is a lack of existing process ontologies to be reused for this module. However, we present the related existing works in the related works section.

[2] http://www.who.int/mediacentre/factsheets/fs115/en/index.html

Results

The results are divided in two parts: schistosomiasis domain and process ontologies models. In each of these sections the perspectives module and the relations modeling are presented.

Schistosomias domain ontology

Schistosomias domain ontology analysis

The IDOSCHISTO domain module is built following the same schema as the existing extensions of IDO such as [8], namely extending the biological and clinical perspectives of IDO for schistosomiasis disease. Moreover, we also add the epidemiological aspect to cover all perspectives on infectious diseases within the medical domain. The epidemiological perspective studies the causes and the spreading dynamics of infectious diseases in a population of host over space and time. The IDO-Core already contains many concepts relevant to epidemiological perspective. We only highlight them and add the missing ones such as the "*spreading*" concept. Indeed, even if concepts such as "*infectious disease epidemic*" or "*infectious disease pandemic*" already exist in IDO, their descriptions correspond to a specific way of disease spreading according to the covered area and the temporal frequency. From our ontological analysis of the epidemiological perspective, we have proposed a general model (Figure 2) of the key concepts that we consider common for the spreading of all infectious diseases and their relationships. The concept *Spreading* is directed modeled here as a subclass of the "*processual entity*" concept in BFO because we consider it more general than the "*infectious disease epidemic*" and "*infectious disease pandemic*". The *Risk Factor* concept here is at the epidemiological level and represents the events that cause the spreading of the disease in a population. Notice that risk factor may be also a relevant concept at the clinical and biological perspectives. A fully formalized ontology is probably needed for risk factors. All the other concepts (*Population*, *Host*, *Contamination*, and *Pathogen*) already exist in IDO or in the imported ontologies. The *participation* relation is here used to link the objects *Host* and *Pathogen* to the *Contamination* process, and the *Population* collection (that is a continuant [4]) to the *Spreading* process. And finally, one can distinguish (i) the probabilistic causality between an event risk factor and the spreading process, and (ii) a contamination occurrence that automatically increases the number of infected persons in the population.

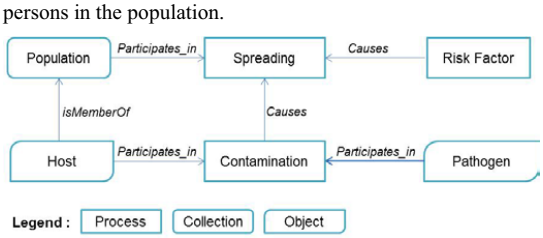

Figure 2 - General spreading model for infectious diseases

Domain ontologies reused

In Table 1, in addition to the BFO foundational and the IDO-Core ontologies, we provide the selected domain ontologies that we aim at reusing for each of the three perspectives. Moreover, there exist other ontologies that are relevant for more than one perspective. The ✓ sign means that this domain ontology is relevant for this perspective and ✗ means that it is not relevant.

The reused domain specific ontologies are briefly described[3] below to show their relevancy to the perspectives module:

- The Information Artifact Ontology (IAO) of information entities is originally driven by work by the OBI digital entity and realizable information entity branch.
- Pathogen transmission (PT) is a vocabulary describing the process of how a pathogen is transmitted from one host or reservoir to another host.
- The Human Disease Ontology (DOID) aims at creating a comprehensive hierarchical controlled vocabulary for human disease representation. Its mission is to provide an open source ontology for the integration of biomedical data that is associated with human disease.
- The Ontology for General Medical Science (OGMS) is an ontology of entities involved in a clinical encounter. OGMS includes entities across medical disciplines such as 'disease', 'disorder', 'disease course', 'diagnosis', 'patient', and 'healthcare provider'. The scope of OGMS is restricted to humans, but many terms can be applied to a variety of organisms.
- The Symptom Ontology (SO) is organized primarily by body regions with a branch for general symptoms.
- The Vaccine Ontology (VO) aims at facilitating data standardization, integration, and analysis in this domain. Current focuses in the VO development are on vaccine categorization, vaccine components, vaccine quality, and vaccine-induced host responses.
- The Ontology for Biomedical Investigations (OBI) provides integrated ontology for the description of biological and clinical investigations. The ontology will represent the design of an investigation, the protocols and instrumentation used, the material used, the data generated and the type of analysis performed on it.
- Chemical Entities of Biological Interest (ChEBI) is a database and ontology of molecular entities focused on 'small' chemical compounds. These molecular entities are either products of nature or synthetic products used to intervene in the processes of living organisms.
- The Protein Ontology (PO) provides an ontological representation of protein-related entities by explicitly defining them and showing their relationships.
- The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. The NCBI houses genome sequencing data in GenBank and an index of biomedical research articles in PubMed Central and PubMed, as well as other information relevant to biotechnology.

Intra-PR and Inter-PR

The intra-perspective relations (Intra-PR) and inter-perspective relations (Inter-PR) of the IDOSCHISTO domain ontology module reuse the existing relation ontologies and specifically those which are semantically relevant to the medical domain. For defining inter and intra perspective relations we are reusing the Relation Ontology (RO) and the RO-Bridge ontologies that are the only ones we have found in the medical domain. RO provides relevant and well-founded relations in biomedical ontologies. This guarantees the interoperability

[3] The ontologies were manually checked but their descriptions are taken from their respective web pages.

purpose of ontologies and supports automated reasoning about the spatial and temporal dimensions of biological and medical phenomena [9]. RO-Bridge provides domain and range constraints using BFO entities. For example the *participates_in* relation is used for Inter-PR between a *Host* and the *Spreading* process (cf. Figure 3). This same relation is also used for the participation of a *Host* and a *Pathogen* in the *Contamination* process. The *transformation_of* relation is used to model the life cycle of the different species (e.g. here for the parasite).

Table 1 - Relevant domain ontologies to be reused for perspectives module

Ontologies	ESO	ISO	BSO
Basic Formal Ontology (BFO)	✓	✓	✓
Infectious Disease Ontology (IDO)	✓	✓	✓
Information Artifact Ontology (IAO)	✓	✓	✓
Pathogen transmission (PT)	✓	×	×
Human Disease Ontology (DOID)	×	✓	×
Ontology for General Medical Science (OGMS)	×	✓	×
Symptom Ontology (SO)	×	✓	×
Vaccine Ontology (VO)	×	✓	×
Ontology of Biomedical Investigation (OBI)	×	✓	✓
Chemical Entities of Biological Interest (ChEBI)	×	×	✓
Protein Ontology (PO)	×	×	✓
NCBI Taxonomy Database	×	×	✓

Schistosomias process ontology

Schistosomias process ontology analysis

The analysis provided here can be adapted for any infectious disease and more specifically a parasitic disease. The spreading of infectious diseases in an epidemiological perspective describes the mechanism of disease spreading in a population. The spreading itself is a consequence of the increasing number of infected hosts (human) by a pathogen (parasite) through the contamination process. The contamination process depends on the pathogen life cycle and on the physiopathological development because we are here (schistosomiasis spreading) in the case of contamination by transmission between infected and healthy hosts. Thus, the spreading of infectious disease can be modeled with three scales:

- The epidemiological scale concerns the population of living beings. A restriction is made at this scale by considering only the human population (Susceptible, Exposed, Infected and Recovered) in which the disease spreads. We will therefore have at this scale the process of spreading of the disease.
- The individual scale is for individuals among which we can distinguish human, pathogen, intermediary host, vector, etc. Processes resulting from the interaction between two individuals of different populations such as the contamination process are modeled at this level.
- The biological scale concerns living beings with processes such as the life cycle of the pathogen, the pathological response of the infected person, etc.

This analysis reveals that the spreading process, as the macro-process of the infectious disease, emerges from the interactions between entities at the individual scale and process occurrences at the biological scale. Applying this to the schistosomiasis, we have therefore proposed a multi-scale model of its spreading process by integrating epidemiological, individual and biological scales in a single model (Figure 3).

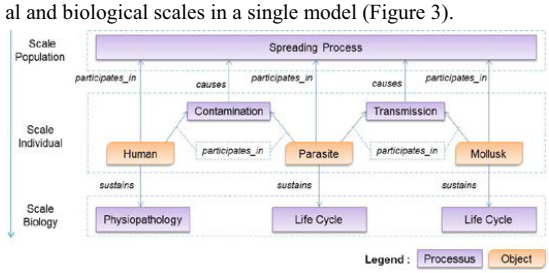

Figure 3 - Multi-level analysis of ID spreading process

Schistosomias processes modeling

There exists no process ontology in the medical domain for our needed purpose of running qualitative simulations. We have provided here process ontologies for each perspective regardless of time, duration, and space. We only consider for this first version the precedence relationship between processes. Instead of using the precedence relation in RO, we formalized the processes in Process Specification Language – PSL[4] [10], which was originally designed to model processes. Even if PSL was designed originally for manufacturing processes, it provides a domain-independent core module that is complete enough for our actual purpose and generic enough to allow extensions if needed. As an example, we provide below the relation between the *Exposition* and the *Contamination* processes (individual scale) based on the SEIRS models already presented in the materials.

```
;; The occurrence of a contamination is preceded by an exposition
(forall (?occContamination)
  (implies (and (occurrence_of ?occContamination Contamination)
                (subactivity_occurrence ?occContamination ?occSpread))
           (exists (?occExposition)
             (and (occurrence_of ?occExposition Exposition)
                  (subactivity_occurrence ?occExposition ?occSpread)
                  (earlier ?occExposition ?occContamination)))))
```

Reasoning on process ontologies can help interpret the origin of an observation using preconditions and precedence relationships. It could be useful in identifying causes of infectious disease emergence or spreading. For instance, the occurrence of a new infected host is the result of the occurrence of a contamination process. Since every contamination is preceded by an exposition according to the model above, this means that there exists at least a water point (as it is the source of contamination) in the area of the infected host that is infested by the schistosomiasis parasite. Therefore, several decisions can be made depending on whether the host lives in the same area where he was diagnosed (block access to water points and carry out disinfestation) or comes from outside (trigger an alert to the public health agent of the area where he comes from).

[4] http://www.mel.nist.gov/psl/

Related works

Many specific disease domain ontologies are extending the IDO-Core. However, the dynamic aspect of the epidemiological perspective is not usually represented in these disease ontologies because of their target limited to domain knowledge representation and to clinical and biological interests. Therefore, these existing ontologies could not be used for the qualitative simulation purpose. However, there are some works on the topic. Biocaster [11] Ontology aims at supporting monitoring of infectious diseases by detecting and assessing public health risk events from online media reports. The Network of Epidemiology-Related Ontologies (NERO) [12] is designed for epidemiological resource annotation and can serve in a monitoring process by helping in simulation model discovery and composition. The Epidemiological Ontology [13] - modeling terms that are mostly related to statistical studies - and the Dictionary of Epidemiology [14] provide respectively guidelines on how the epidemiological concepts should be organized and the concepts that are relevant to this perspective.

Discussion and Conclusion

In this paper we presented the ontological analysis and design of IDOSCHISTO ontology as an extension of the IDO-Core for schistosomiasis' disease. IDOSCHISTO is designed to support the monitoring of the spread of schistosomiasis. On one hand, it is aimed at being used as "domain ontology" for facilitating semantic data integration, semantic interoperability of applications of involved organizations, and collaborative work between domain experts. On the other hand, its "process ontology" module is designed to perform qualitative simulation for risk prediction, explanation of observed phenomena, and decision making. IDOSCHISTO is also designed as a multi-perspective model - with a sub-module for each perspective - and follows the systemic modeling approach [15]. Each of these perspective modules in the domain ontology reuses the IDO-Core ontology, the BFO foundational ontology and relevant domain specific ontologies.

Although the paper focused on the design approach for the IDOSCHISTO, its novelty lies in constructing a process ontology of the spread of infectious diseases for monitoring purposes. Although the IDOSCHISTO ontology is not fully available, we have successfully studied and imported all the relevant domain ontologies that we will reuse for each perspective and the relations ontologies for the inter-PR and intra-PR.

As future work, we will initially undertake enriching the relational ontologies, and resolve any redundancies that may occur during the importation process of the different domain ontologies we are reusing. Next, the remaining domain specific concepts and relations to schistosomiasis will be added to finalize the extension. The final step will be the evaluation of the ontology itself and the evaluation of its contribution to the schistosomiasis monitoring effort in Senegal.

Acknowledgments

We want to thank the embassy of France in Senegal through the SCAC scholarship program that supports this work and the schistosomiasis domain experts for their collaboration. Many thanks to the anonymous reviewers and J. B. Lamy of LIM&BIO lab for their comments and suggestions for improving this paper.

References

[1] Camara G, Despres S, Djedidi R, and Lo M. Towards an ontology for an epidemiological monitoring system. In: Rothkrantz L, Ristvej J, Franco Z, eds. Proc. 9th International ISCRAM Conf. : Simon Fraser Univ.; 2012.

[2] Kermack WO, and McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond 1927 Aug: 115(772): 700-721.

[3] Grenon P, Smith B, and Goldberg L. Biodynamic ontology: applying BFO in the biomedical domain. Stud Health Technol Inform 2004;102: 20-38.

[4] Simons P, and Melia J. Continuants and occurrents. Proceedings of the Aristotelian Society: Supplementary Volumes 2000: 74: 59-75+77-92.

[5] Studer R, Richard Benjamins V, and Fensel D. Knowledge engineering: principles and methods. IEEE Transactions on Data and Knowledge Engineering 1998: 25(1-2): 161-197.

[6] Rector AL, and Rogers J. Patterns, properties and minimizing commitment: Reconstruction of the galen upper ontology in OWL. In: Gangemi A, and Borgo S, eds. EKAW'04 Workshop on Core Ontologies in Ontology Engineering. CEUR, 2004.

[7] Grenon P, and Smith B. SNAP and SPAN: Towards dynamic spatial ontology. Spatial Cognition and Computation. 2004: 4(1): 69-103.

[8] Lin Y, Xiang Z, and He Y. Brucellosis Ontology (IDOBRU) as an extension of the Infectious Disease Ontology. J Biomed Semantics 2011: 2: 9.

[9] Smith B, Ceusters W, Klagges B, KöhlerJ, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, and Rosse C. Relations in biomedical ontologies. Genome Biol 2005: 6(5): R46.

[10] Gruninger M. Using the PSL Ontology. In: Staab S, and Studer R, editors. Handbook on Ontologies. Berlin: Springer-Verlag; 2009. pp. 419-431.

[11] Collier N, Matsuda Goodwin R, McCrae J, Doan S, Kawazoe A, Conway M, Kawtrakul A, Takeuchi K, and Dien D. An ontology-driven system for detecting global health events. In: Proc. 23rd International COLING Conference; 2010 Aug; Beijing. Stroudsburg, PA: Association for Computational Linguistics; 2010. pp. 215-222.

[12] Ferreira JD, Pesquita C, Couto F, and Silva M. Bringing epidemiology into the Semantic Web. In: Cornet R, Stevens R, editors. Proc. 3rd International Conference on Biomedical Ontology (ICBO 2012), KR-MED Series; 2012 Jul 21-25; Graz, Austria.

[13] HuGE Net. Guidelines for the epidemiological ontology. 2007. Available from: http://www.hugenet.org.uk/resources/informatics/Ontology_Version_1.pdf

[14] Porta M, editor. A dictionary of epidemiology. 5th ed. USA: Oxford University Press; 2008.

[15] Le Moigne J-L. La modélisation des systèmes complexes. Paris: Dunod; 1990.

Address for correspondence

Gaoussou Camara, gaoussoucamara@gmail.com