AUTOIMMUNE RELAPSE PREDICTION USING MULTIPLE PARALLEL DATA SOURCES

Data Sharing Plan

Mark Little

Contents

Preamble

By the nature of this project, we will be engaging with a number of other research bodies, and intend to publicly share results and processes as AVERT develops in the literature and at relevant conferences. Research groups that wish to have access to the data will be eligible to apply to do so; IGB clearance is required all such groups. All data shared outside the AVERT research group will be irreversibly anonymised. Because of the challenges of sharing the data, some groups may be eligible only for access to aggregated data or a limited dataset. Special circumstances would be necessary for access to the raw dataset in its entirety, explained below. Data will only be made available when the IGB are satisfied that any academic or commercial interests of our own study would not be unreasonably negatively impacted by the sharing of AVERT's database.

Open Data

The partners of AVERT are strongly supportive of open access and the FAIR data principles (see below). AVERT is committed to the principles of open data, data sharing, reusing data resources and research transparency. However, the partners cannot currently commit to an unconditional open access policy, but agree on a policy of "as open as possible, as restricted as necessary".

Wherever possible AVERT will create data sets that are sufficiently de-identified and/or aggregated to enable them to be published and reused. However, two important constraints need to be highlighted for transparency here. Firstly, some of the data accessed and used will be pre-existing data sources, such as the RKD dataset, that have been curated and are being held by other research projects, which will continue to have ownership of those data. Secondly, much of the novel research will be undertaken using subject level data, which will be inevitably difficult to de-identify. In compliance with national and European laws, AVERT will primarily respect conformance to data protection legislation above any wish to make research data openly accessible.

In situations where potentially useful data sets are under the custodianship of AVERT and which cannot be made openly accessible, AVERT will publish a data sharing specification which makes clear how other external research teams may formulate a data sharing request in order to be able to undertake specified research using the data sets. The members of the group are acutely aware of the sensitivity of personal data and the need for its protection, and will implement an appropriate Data Management Plan given this context. AVERT partners have kept track of the openAIRE pilot, and will align with the policies and metadata specifications of its successor open data service.

Standard metadata describing the data set

We intend to publish according to "minimal metadata standards" ensuring that – at the very least – the list of fields used as part of analyses (and other relevant metadata) are made available at a FAIR data point to ensure its findability. We currently intend that this will be available on the MOLGENIS catalogue and prepared according to the Data Documentation Initiative standard[1].

[1] http://www.ddialliance.org/Specification/RDF

The metadata tables will be established once the dataset is defined in detail

FAIR data principles: Findable, Accessible, Interoperable, Reusable

TO BE FINDABLE:

F1. (meta)data will be assigned a globally unique and eternally persistent identifier.

F2. data will be described with rich metadata.

F3. (meta)data will be registered or indexed in a searchable resource.

F4. metadata will specify the data identifier.

TO BE ACCESSIBLE:

A1 (meta)data will be retrievable by their identifier using a standardised communications protocol.

A1.1 the protocol will be open, free, and universally implementable.

A1.2 the protocol will allow for an authentication and authorisation procedure, where necessary.

A2 metadata will be accessible, even when the data are no longer available.

TO BE INTEROPERABLE:

I1. (meta)data will use a formal, accessible, shared, and broadly applicable language for knowledge representation. In the case of AVERT, most of the study data will be uplifted into the data model RDF, which is considered the standard by the GO-FAIR group.

I2. (meta)data will use vocabularies and ontologies that follow FAIR principles.

I3. (meta)data will include qualified references to other (meta)data.

TO BE RE-USABLE:

R1. meta(data) will have a plurality of accurate and relevant attributes.

R1.1. (meta)data will be released with a clear and accessible data usage license.

R1.2. (meta)data will be associated with their provenance.

R1.3. (meta)data will meet domain-relevant community standards.

External data access mechanisms

To gain access to the data itself, researchers should in the first instance contact the head of the information governance board (IGB, Prof Lucy Hederman (hederman@tcd.ie)). They should include their name, position, institution, a brief description of the project, the intended data use, the list of fields required and whether the project is intended to be commercialised. The field of email should be entitled

"AVERT data request –" as well as the name of the researcher's project. Under normal circumstances the email will be responded to within 14 days, although this may take longer under certain circumstances.

Prof Hederman's response may:

- request further clarifications,
- grant access to the relevant data where no sensitive data is requested and no other special considerations are required, or
- otherwise will provide a formal Data Request Form (DRF). This will require that the researcher describes how the data will be used, provides justification for sharing of sensitive fields, and commits to managing the data in an appropriate way on behalf of all project members.

When a DRF is required, the IGB will be responsible for confirming the appropriateness of sharing the requested data, and reserve the right to provide only the data they themselves consider justifiable. This decision will normally be made within 30 days of receipt of the form.

We expect to share the data in RDF format. Sharing of data in other formats such as CSV should be stated in the DRF, will require special circumstances and may not always be possible.

Intellectual Property and Authorship

The primary remit of the AVERT study is to develop algorithms that predict flare risk. However, we also welcome the development of additional related projects which have only been possible as a result of the enormous efforts of many of the lead investigators themselves and of many other enthusiastic researchers, at both senior and junior levels, who have taken advantage of the opportunities offered by the AVERT project to undertake further studies. We have therefore developed two separate policies to address the issues of authorship: one for the primary papers and a separate one for the associated papers.

The authorship committee consists of: Mark Little (ML), Declan O'Sullivan (DO'S), Brett Houlding (BH), John Kelleher (JK) and Lucy Hederman (LH). For this policy and for all decisions on authorship and acknowledgements, the decision of the authorship committee is final and binding.

1. Principles for authorship

These principles are based on consideration of the following:

1.1. Compliance with the most current version of the International Committee of Medical Journal Editors (ICMJE) "Uniform requirements for manuscripts submitted to Biomedical Journals: Writing and Editing for Biomedical Journals" (www.icmje.org) requires named author to meet all 3 of the following:

- "substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data"

- Involvement with drafting of the article or “revising it critically for important intellectual content”
- final approval of version to be published

1.2. To be as inclusive as feasible. However, complying with Authorship guidelines “drafting the article” or “revising it critically for important intellectual content” becomes increasingly impractical with large numbers.

2. Primary Papers

We will need to recognise any restrictions/requirements required by funding bodies regarding authorship rules, which may affect the subsequent points in this section. The primary papers are considered to be all those publications which define algorithms for prediction of vasculitis flare or onset. This would include all the methodological papers which led to the development of the algorithms because these methods and papers are necessary to underpin the algorithm papers themselves. These papers will conform to the original requirements defined in our funding application for the project. These projects have been largely driven by the efforts of the authorship committee, and we therefore propose the following:

Each of the main investigators (ML, DO’S, BH, JK, LH) will take senior authorship role in rotation, for all papers. For each paper, the other chief investigators will take 2nd, 3rd etc place in authorship list, unless there is no readily identifiable primary author amongst the group, in which case one of the main investigators will take primary authorship for the work. We will attempt to encourage all papers to include a “junior” primary author wherever possible and appropriate, to recognise their contribution to the project. This is coupled with an expectation that the primary author will undertake the main task of preparing the first draft of each paper and help to oversee the editing changes with assistance from the main investigators, primarily the nominated senior author for that paper. If the primary author cannot undertake this task, or there is no suitable primary author, one of the main investigators can propose themselves to this task and this role and this will need to be agreed by the main investigators. Members of the core AVERT team, including actively participating co-applicants as well as staff who have been essential to the running of AVERT and who constitute part of the core team, as defined by the main investigators, will be listed as authors, with the proviso that they fulfil the role of an author in being involved in reviewing and revising the manuscript in a timely fashion.

3. Associated Papers

The authorship guidelines for associated projects will reflect the nature of the work done. As the authorship committee, we retain the right to be authors on all of these associated projects. However, it is up to the main investigators to decide whether or not to enforce this for each paper on a case by case basis; in some cases there will be a requirement for all main investigators to be included, or only 1 or 2 of them, or if there are projects where none of the main investigators have made a significant contribution, then we would acknowledge this and none of them would be authors. We think it is likely that all the associated papers will, however, be based on projects for which the 3 chief investigators have provided substantial input because all sub-projects are brought before the AVERT committee for

discussion and for approval before the project is undertaken and the associated project takes advantage of the AVERT infrastructure and database. We would expect the lead investigator for each of the subprojects to liaise with the authorship committee over publication of the paper or papers and for them to take the lead in deciding on eligibility for authorship and order of authors. The main investigators would not expect to be either 1st or senior author on any of these papers, unless they are primarily responsible for that sub-project. All of the associated papers would be expected to include investigators and /or core team members who have made substantial contributions to that project as authors for that particular paper. The final decision regarding authorship on all associated papers will rest with the authorship committee.

All abstracts arising from associated projects should be submitted to the AVERT steering committee as part of the regular communication process, but at least ten working days before any proposed abstract deadlines, in order to ensure there is sufficient time for review and revisions, if needed. Each associated project shall also have an allocated point of contact on the core team, who will be responsible for liaising regularly with the associated study members and should therefore be aware of planned submissions as they emerge. The allocated point of contact with the AVERT core team will be responsible for ensuring the quality of associated studies and will provide feedback on the progress of these studies to the wider AVERT steering committee. It is the expectation that authorship on abstracts will follow the same rules as for the full papers; exceptions to full lists of authors on abstracts may be made in cases where author names and affiliations substantially count toward word/character limits for the abstract text. The final decision regarding authorship on all abstracts will rest with the authorship committee.

We encourage the use of contributorship with an appendix to acknowledge those involved less directly in the study, but would not expect the full contributorship list which we would apply to primary papers to necessarily apply to each associated paper. However, we do require that all those who have been significantly involved in each AVERT paper should be acknowledged. This could be in the form of an appendix to each paper. The AVERT committee will take responsibility for providing a full list of personnel who could be named as authors, contributors or appear in the appendix. This list will be available on the AVERT website. It will be the responsibility of the senior author and the allocated member of the AVERT core team for that particular paper, to edit the list for use in the associated paper or papers as required e.g. if particular names or groups of names should be authors, contributors or simply appear in the appendix. The senior author will be responsible for ensuring the accuracy of these lists and also for making sure that the manuscript complies with the requirements of any individual journal.