The Ten Commandments for Translational Research Informatics

Tim Hulsen[1]

Abstract

Translational research applies findings from basic science to enhance human health and well-being. In translational research projects, academia and industry work together to improve healthcare, often through public-private partnerships. This "translation" is often not easy, because it means that the so-called "valley of death" will need to be crossed: many interesting findings from fundamental research do not result in new treatments, diagnostics and prevention. To cross the valley of death, fundamental researchers need to collaborate with clinical researchers and with industry so that promising results can be implemented in a product. The success of translational research projects often does not depend only on the fundamental science and the applied science, but also on the informatics needed to connect everything: the translational research informatics. This informatics, which includes data management, data stewardship and data governance, enables researchers to store and analyze their 'big data' in a meaningful way, and enable application in the clinic [1]. The author has worked on the information technology infrastructure for several translational research projects in oncology for the past nine years, and presents his lessons learned in this poster (and a published paper[2]) in the form of ten commandments. These commandments are not only useful for the data managers, but for all involved in a translational research project. Some of the commandments deal with topics that are currently in the spotlight, such as machine readability, the FAIR Guiding Principles and the GDPR regulations. Others are mentioned less in the literature, but are just as crucial for the success of a translational research project.

1: Create a separate data management work package

Researchers often do not think about data management, except for when it is already too late: the data sit on several computers scattered over different organizations, and nobody knows how to combine them and make sense of them. The solution: create a separate work package or work stream on data management, with its own goal, milestones and deliverables.

2: Reserve time and money for data entry

Instead of leaving data entry to trial nurses, who usually already have a high workload, reserve money to hire people who can do data entry for a certain amount of hours per week. By spending relatively little money on data entry, one can save a lot of time and money by not having to redo analyses because of missing or erroneous data.

3: Define all data fields up front together with the help of data analysis experts

Create a 'codebook' with an extensive list of data fields designed to answer all research questions that we could think of at the start of the project. Involve the data analysts or statisticians, because they have the clearest insights on what is needed here. Connect to standards and ontologies wherever possible.

4: Make clear arrangements about data access

Data access can be a sensitive issue. Therefore, it needs a clear arrangement up front. Of course, data access needs to be arranged in the informed consent as well, as patients are the data owners, and the GDPR (and HIPAA) have strict regulations about the patient's privacy. The Data Processor Agreement (DPA) should be very clearly defined, and enough time should be reserved to arrange it.

5: Agree about de-identification and anonymization

The responsibility for proper de-identification of the data often lies with the organization that collects the data (usually the hospital). They should de-identify (or anonymize where possible) the data and keep a mapping table at the hospital. De-identification and anonymization should also be arranged very clearly in the informed consent and the data processor agreement.

6: Reuse existing software where possible

There is usually no need to develop tools for data capturing, data management, data quality control, etc. from scratch, because there are many open source tools available for this, which can not only be used freely but also developed further. Within the CTMM-TraIT project (now part of Health-RI), a list of suitable tools was created, which can be viewed at https://trait.health-ri.nl/trait-tools/.

7: Make newly created software reusable

There might be cases where study-specific software needs to be created, for example to perform novel analyses. If there are no intellectual property issues, this newly created software (together with metadata) can be submitted to repositories such as GitHub, SourceForge or FigShare. This way, future translational researchers can reuse the software and do not need to reinvent the wheel.

8: Adhere to the FAIR Guiding Principles

The FAIR Guiding Principles should be applied to both data and software created in a translational research project, to achieve transparency and scientific reproducibility. FAIR means Findability, Accessibility, Interoperability, Reusability. FAIR does not only support reuse of data by individuals, but also puts emphasis on enhancing the ability of machines to automatically find and use the data.

9: Make sure that successors are being instructed correctly

Translational research projects usually take 4 to 5 years. Clinicians, researchers and data managers, but also trial nurses, might come and go during this period. These people spent quite some time learning how to fulfill their task; know-how that will be lost if not enough time is spent on correctly instructing the person that will take over the job.

10: Make it sustainable: what happens after the project?

What will happen when the project is finished? For example: who will pay for the continued storage of left-over biomaterials? Who will keep the database running? At the start of the project, the researchers should already make a plan for what happens at the end of the study, to avoid that data and biomaterials are lost for future research. This planning should also include a financial paragraph.

Reference projects: MOVEMBER GAP3 LIMA relimagine

[1]*Philips Research, department of Professional Health Solutions & Services, Eindhoven, the Netherlands*

1. T. Hulsen, S.S. Jamuar, A.R. Moody, J.H. Karnes, O. Varga, S. Hedensted, R. Spreafico, D.A. Hafler and E.F. McKinney, From big data to precision medicine, Front Med (Lausanne) 6 (2019), 34. doi:10.3389/fmed.2019.00034.
2. T. Hulsen, The ten commandments of translational research informatics, Data Science 2 (2019), 1-2, 341-352. doi: 10.3233/DS-190020

Please contact tim.hulsen@philips.com if you have any questions

See for more info:
https://www.philips.com/a-w/research/home.html
https://gap3.movemberprojects.com
https://www.lima-project.eu
https://www.reimagine-pca.org

PHILIPS