Horizon 2020

Project title: AuTonomous intraLuminAl Surgery

Clinical data-gathering

Deliverable number: D2.1

Version 1.1

Funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 813782

Project Acronym: ATLAS

Project Full Title: AuTonomous intraLuminAl Surgery

Call: H2020-MSCA-ITN-2018

Grant Number: 813782

Project URL: https://atlas-itn.eu

Deliverable nature:	Report (R)
Dissemination level:	Public (PU)
Contractual Delivery Date:	October 30, 2019
Actual Delivery Date	October 30, 2019
Number of pages:	8
Keywords:	Data gathering
Authors:	Benoit Rosa, UNISTRA
Peer review:	Alicia Casals, UPC Paolo Fiorini, UNIVR Arianna Menciassi, SSSA Florent Nageotte, UNISTRA Johan Philips, KU Leuven Emmanuel Vander Poorten, KU Leuven

Abstract

This deliverable 2.1 is a step-by-step guide for clinical data gathering. It aims to provide a starting point for the ESRs in the ATLAS project who wish to gather clinical data within the scope of their PhD project.

Contents

List of Figures

List of Tables

1 Definitions and scope

1.1 Scope of this document

This document concerns the act of clinical data gathering. Clinical data gathering can be defined as collecting data during or after a medical or surgical procedure performed on a human subject by a qualified physician. One can distinguish three types of clinical data-gathering procedures :

- Data collection which does not involve any modification of the OR or of the clinical workflow (e.g. collecting medical images after the surgery)
- Data collection which does involve modification of the OR or of the workflow, but no pyshical intervention on the patient or on the surgical instruments (e.g. placing RGBD cameras in the OR)
- Data collection which involves modification of the surgical instruments or of anything which is in contact with the patient (e.g. placing sensors on an endoscope)

The third case has implications on the safety of the patient, and would involve certification procedures for asserting that the instruments are sterilizable (and indeed sterile after undergoing the sterilziation process) and safe for the patient. It is therefore out of the scope of this document. The first two cases, in which modifications to the OR or clinical worflow may be done, but without direct impact on the patient's safety, are discussed in the following. Note, however, that some workflow modifications may have indirect impact on the patient (for instance taking a supplementary fluoroscopy image induces an increase in X-ray doses).

This document will detail the process of conducting and documenting a clinical data gathering pipeline. We will detail the general steps required in the preparation and collection of the data gathering. Please note that this document is intended to be a step-by-step guide for helping students setting up data collection experiments. The document has no legal value, especially regarding the ethical approval procedures and regulation bodies to contact, which are country-specific and will not be detailed in this document.

1.2 Types of data collection

There are several types of data collection in clinical trials. The type of data depends on the data collection process and the person collecting the data:

- Data collected by the patient: questionnaires (e.g. patient-reported outcome)
- Data recorded by the investigator (clinical staff or external): electronic patient records, medical images, laboratory data (e.g. analysis of biopsy samples)

Given the technical scope of the ATLAS project, we will focus on data collected by the investigator (clinical staff or engineers) during surgical procedures. Data can be of several types, below are a few common examples:

- Pre-operative medical images (e.g. MRI or CT scan)
- Intra-operative medical images (e.g. intraoperative MRI, Ultrasound, Endoscopic camera recording)
- Other sensory data (e.g. sensory information from a navigation system, kinematic information from a robotic device, RGBD cameras placed in the OR ..)

2 Preparation

Before starting the data collection, a number of steps must be taken in order to ensure that the data are collected properly. The steps are listed below in a succession of questions to answers.

2.1 Which data do we need, and why?

The very first question to ask is why do we need data. What is the hypothesis or the research question being tested? Once the answer to this question is clear, answering the below list of questions is essential in order to determine which data are needed:

- What is the intended use of the dataset? (e.g. surgical instrument recognition in endoscopic images)
- Which type of data are needed? Do I need to record separately the configuration (i.e. settings of the ultrasound machine, for instance) ?
- What is the minimum number of cases or datapoints required?
- Does the dataset require the intervention of multiple surgeons? Should it be recorded at several clinical sites?

2.2 Do we *need* to collect data?

Often, similar datasets may be available. Consulting public data repositories or domain specific repositories with published datasets should be done before considering collecting new data. Examples of public data repositories are Zenodo (https://zenodo.org/) or the Harvard Dataverse (https://dataverse.harvard.edu/). Data may also already have been collected by surgeons during previous procedures (e.g. pre and post-operative MRI scans for a given procedure). Ask your local surgeon (and the consortium) whether such data have been collected during clinical routine.

If a similar dataset is available, conditions should be checked to ensure that it can be used. Public datasets are typically free to use for academic purposes, but consent for secondary use should be checked. The license should also be checked to ensure exactly what can be done with it.

For unpublished datasets or data previously recorded by clinicians, the protocol (see subsection below) and ethical approval of the data collection should be obtained. Some cases may require declaring the data collection as a retrospective clinical trial (especially if the intended use was not foreseen in the original ethical approval).

2.3 What is the required administrative framework?

Administrative requirements for clinical data gathering vary from country to country. The general rule is that any collection or use of data concerning patients need to be authorized by an ethics committee (additionally, import authorization might be required, especially if data was collected in non-EU countries such as Switzerland, for instance). A protocol should be written. It is a document that describes the whole data collection procedure. A typical protocol document will include general information about the study (title, investigators, background, objectives and methods, general timeline), about the patients (selection of subjects if relevant, ethical aspects and safety), about the data recorded (data types, data handling, recordkeeping, data processing and anonymization).

The committee will check that the study is indeed necessary, that the patient is aware of the data collection, has given his/her informed consent, that his/her safety is not compromised, and that the patient personal information and data are protected. Local clinical staff are the best persons to help setting up such protocols. Open sharing of datasets is encouraged when possible, but should comply to strict anonymization and privacy rules (see the ATLAS Data Management Plan for more details).

2.4 How will the data be recorded?

This section relates to the technical preparation of the data collection. Some data can be recorded directly on the hospital's data management system. Medical imaging data are typically recorded on the PACS (Picture Archiving and Communication System) of the hospital. They can later be extracted (often with built-in anonymization features) in a standard format such as DICOM. Endoscopic camera images are not always recorded on the PACS. Often, a USB stick can be plugged at the back of the endoscopy tower to extract the videos.

Other data types may require setting up a computer in the Operating Room for data recording purposes. One special case is when several data types are simultaneously recorded and need to be synchronized. This usually requires setting up a specific equipment and recording procedure. These cases must be anticipated and mentioned in the ethical approval.

On top of the actual data used for research, metadata of the *experiment* itself should be collected. Settings of the machine, date and hour, surgeon performing the procedure, type of procedure, etc ... These data are typically sensitive, as some of them may allow deanonymizing the patient (for instance date and surgeon name for uncommon procedures). One should anticipate things, think about all possible cases of misuse of the data, and how to mitigate those. Again, this should be specified clearly in the protocol and ethical approval documents.

2.5 Who will perform the recording, and when?

This seems like an obvious question, but it has implications on the data collection process. If the data are to be recorded by the surgeon directly or the clinical staff, one needs to make sure that they are aware of the study design and of *what* exactly needs to be recorded. Clinical staff have many things to handle during a surgery and may pay attention to your study only marginally during a surgical procedure. IF relying on clinical staff, special care is likely required to ensure the recording is non-obstrusive, simple and error-prone. This may coll for developing a brochure with instructions and/or a graphical user interface that provides a walkthrough through the recording steps. In the latter case, extensive testing of the GUI is recommended.

Being aware of the *when* (as much as possible) can also help organizing things. Some surgeries are planned in advance while others are typical emergency procedures.

2.6 Where will the data be recorded?

This question may also seem obvious, as the answer is "in the OR". The question should however be raised for several reasons. First, surgeons may be performing surgeries in different ORs within a single hospital, depending on the day or of the OR availability. If special equipment needs to be setup, then only operations in that OR can be recorded. Second, if possible one should visit the OR at least once before starting a data collection campaign, even if it does not involve any specific equipment. Often, surgeries are long, and if one wants to be in the room and observe, it is a good thing to know the environment in order not to pose any problem to the clinical staff who is working around the patient.

3 During data collection

3.1 Presence in the OR

In most cases, the investigator will want to be present in the OR when collecting/recording clinical data, at least for the first few surgeries in order to check that the recording procedure is performed correctly. In some cases it is however required, for instance to operate the data recording software or equipment.

Note that it may not always be beneficial to be present in the OR, as the presence of a person may interfere with the clinical workflow. In any case, the presence of engineers in the OR to observe or operate recording equipments should be mentioned clearly in the ethical approval request.

3.2 Checklists

Checklists can be a valuable tool if the data collection pipeline involves several materials or operations. A good practice is to make a step-by-step checklist of the required actions. If possible, one person should be devoted to fill the checklist as the surgery goes and note timings and events, while another person does the actual recording of data.

Checklists can also be extremely valuable if the data collection involves a specific setup and/or equipment which should be repeated for each procedure. A checklist will help following all the steps and minimizing sources of errors. An example of checklist for a cardiac valve flow image (Doppler) recording is shown on Table 3.1 (this table is an incomplete example, many other details could/should be incorporated in the checklist). In the table, the main actions carried out by the different staff categories (here the clinician and the engineer in the OR) are listed. The time should be noted when every action is performed to help with subsequent analysis/sorting of the data (especially when recording different data types on unsychronized machines). In this example, information about the recording equipment is also noted, as the ultrasound machine and the probe used (in this case) may change from procedure to procedure, and this information may not always appear in the DICOM metadata. Finally, a specific space for notes is left both on each row and at the end of the document.

Table 3.1: Example of checklist for a cardiac valve flow image recording

Date: Procedure type: Protocol number:

Clinical Staff	Engineers	Additional note	Time
Step 0 : initial preparation			
	Prepare USB Stick	Ideally the day before	
Prepare patient			
Start Ultrasound machine	US Machine model: ..... US Probe model: ..... US Machine settings:		
Step 1 : recording images			
Find the valve on echo			
Record a video of the valve	Views taken: ..... Video recorded? Y/N		
Record a Doppler flow video	Views taken: ..... Video recorded? Y/N		
	Verify recording		
Finish operation			
Step 2 : saving images			
	Save data on USB stick		
General notes and observations			

4 After the data collection

4.1 Checking the data

Obviously, checking the data is the first operation which should be performed. One should not wait the end of the data collection campaign, but rather check collected data after each collection day/event. One should check that:

- The data are complete and not corrupted
- The patient approval form has been filled correctly
- The data are fully anonymized

Note about data anonymization: special attention should be paid to data types which may carry information breaking anonymity. Common examples are non-standard metadata in the DICOM files, the name of the patient appearing inside the image (as pixel values) in ultrasound images (see figure 4.1), or camera images showing the face of the surgeon and/or the patient briefly when the endoscope is retracted from the patient's body. As much as possible, data should be anonymized during acquisition. If that is not possible, postprocessing and data checking are two essential steps.

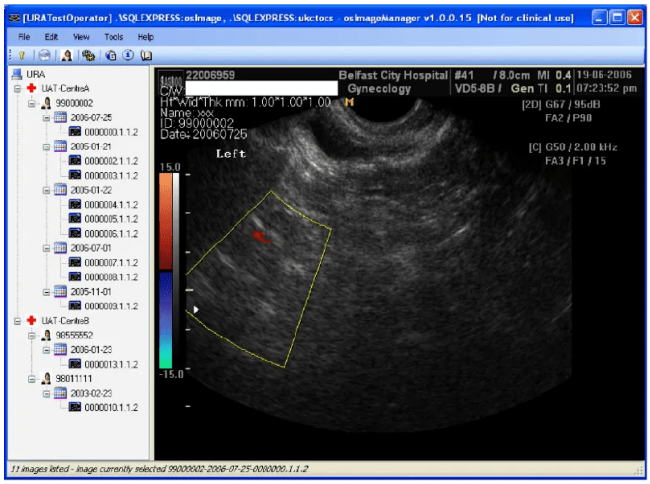


Figure 4.1: Example of ultrasound image with patient identification information contained in the image. In this case, the information is masked automatically with a blank line. Source: OSPACS: Ultrasound image management system, Stott et al, Source Code for Biology and Medicine, 2008. Licence CC-BY

4.2 Storage of the data

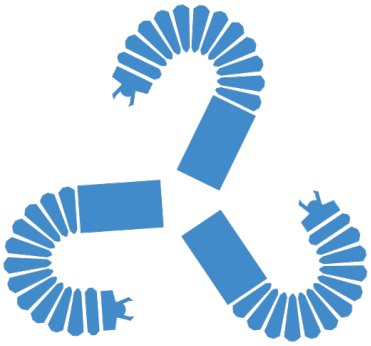
Data should be stored in a secure place. Provisions for data storage are listed in the ATLAS Data Management Plan.

4.3 Postprocessing

Postprocessing of the data may be required for data analysis or formatting. For instance, one may ask one or several expert surgeons to label the data for later use. Another example is to put the data in an open format, if the original data format is proprietary. Any postprocessing should be duly documented, and the original data as well as the postprocessed data should be stored. Ideally, the postprocessing pipeline should be automatic and standardized so that all data points are treated the same way. Completeness of the data should be checked again after any postprocessing step. Postprocessing the data may sometimes break synchronization (especially when converting data formats).

4.4 Publications and public dataset sharing

Before publishing a dataset or a scientific paper using the data, one should check that the protocol and patient consent forms are in order. The ethical approval number will be required for sharing the data publicly. It is also a good practice to communicate this number in the related publications.



The ATLAS project

lastly modified: November 4, 2019

ATLAS-D2.1-1.1

Horizon 2020