Deliverable D3.1 Target list of initial data contributors – publishable summary

This deliverable includes a spread sheet listing the potential data contributors, contacts and a preliminary assessment of the data contents of the databases. To date PIONEER has identified approximately 70 data bases. This number is expected to grow as we continue to identify new sources and connect with data holders. The spread sheet listing also includes information on the key contact persons for each dataset and their respective contact person within PIONEER, as well as information on contributors data sharing policy's and disease specific information.

Deliverable D3.2 Data needs assessment tool – publishable summary

This is an excel spreadsheet used to identify key data requirements for each research question. The intent of this sheet is to create a "screening" device that will allow us to identify which of the 70+ databases in the PIONEER data catalogue have all (or nearly all) of the elements needed to conduct a specific research project.

A key challenge for PIONEER is to understand what data is in which database. This is important to know, specifically because it will take several years to get all of the various databases into a standard data format and available for conducting research. So, if in the short run, we want to be able to conduct specific research projects, we need to know what databases have the requisite information, and prioritise the mapping of those databases first. This tool is one half of the equation, it lets us know what is needed. The other half of the equation is what is available in each database.

Deliverable D3.3 Data gap assessment tool– publishable summary

The data gap assessment tool is a tool that allows us to determine which datasets have most of the requisite data for a particular research question. There are two components to this tool. The first is the Data Needs Assessment form, deliverable 3.2; and the second is the Data Source Fingerprint or Fact Sheet. At this point, we have completed the Needs Assessment for each the first two research questions, and are in the process of reaching out to each of the potential data sources and gathering information to complete their Fact Sheets. Once the Fact Sheets are completed, we can do a comparison of the data needs to what is available and create a score for each data source for each research question; then prioritise the on-boarding of those data sources which provide the most information. At the same time, we can identify if there are more universal gaps in the data we have, and hopefully identify new data sources or approaches which can fill those gaps and allow the research to progress.

Deliverable D3.4 Data Access Agreement Templates – publishable summary

In order to offer more flexibility to the data providers, the PIONEER consortium has chosen to utilize a mixed of both the Federated (remote data) and Central (importable data) Database Models and will ask the data provider to choose which model they prefer to use or are required to use by local law or other restrictions/regulations.

WP3 together with WP8 have established two standardised Data Access Agreement (DAA) template (Central and Federated) to be used with data contributors. Once data contributors have agreed to share and provide access to their data, we will negotiate a DAA with each participant. Both the Central and Federated DAA outline the policies and procedures under which their data may be accessed and analysed:

a) Who may access their data;

b) What kinds of questions may be addressed;

c) What role the site would play in the analysis;

d) What role they would have in future publications based on their data;

e) What legal issues are to be taken into consideration in each country; and

f) What (if any) remuneration they would receive in return for access to their data.

The DAA templates are a starting point for individual negotiations with potential data contributors for access to their data.

Deliverable D3.5 Value propositions for data contributors – publishable summary

In order to encourage participation in the PIONEER collaboration, we have developed value propositions tailored to each type of data contributors that also satisfies GDPR regulations and allows meaningful research collaboration. These value proposition will form the basis for the Data Access Agreements. Potential data contributors include: academics focused on prostate cancer, large clinical practices and medical centres, life sciences companies, data aggregators and payers/governments. In exchange for signing these agreements, data contributors will be given certain rights and privileges, along with accepting certain obligations. All data contributor will be acknowledged on the PIONEER website as a data contributor, and will be entitled to use the "PIONEER Participant" logo in their materials. PIONEER will also prepare a joint press release announcing their participation in the project.

Deliverable D3.6 Meta data model – publishable summary

The PIONEER meta data model is designed to develop a standard set of on-line meta information about the data sets which can be used to help prioritize which data contributors to focus on and to help identify which data sets can be used to answer specific Research Questions. Templets with required information are used to create a community in the EMIF metadata catalogue and a 'fingerprint' is based on the data contributor Fact Sheet. Ultimately, the PIONEER metadata model will form the basis for the data catalogue.