Reviewer's report

Title: A European inventory of common EHR data elements for clinical trial feasibility

Version: 3  Date: 4 December 2013

Reviewer: Felix Koepcke

Reviewer's report:

The content of the manuscript has improved significantly since the first version. It could still gain from professional proof reading.

- Major Compulsory
  none

- Minor Essential Revisions
  1) hardly used data elements are moved to the wish list. Still some data elements in the data inventory show less than 10% usage (e.g. substance abuse 2%). Should these not therefore be moved to the wish list?
  2) In the background you introduce fully structured, semi-structured and free text data elements, but semi-structured elements never appear again in the manuscript. The distinction should be referenced or explained.
  3) There is also still no definition for availability. (available yes/no) is not sufficient.
  4) I think frequency and usage mean the same thing. One of the terms should be replaced by the other throughout the manuscript. (Personally I would prefer usage.)
  5) I would still like to see which site searched a subsystem and which their EHR. general categories are sufficient, e.g. EHR, intensive care, cardiology, pathology, ... . This information could easily be added as a header line in the heat map.
  6) English language should be improved.

- Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)
  7) by a group of up to 30 people > by a group of 30 people
  8) standard definitions were identified and added > definitions were added
  9) To verify their availability > To measure their availability
  10) Rankings of data elements were created from the results of the data exports. > Data elements were ranked according to usage.
  11) In conclusion: The data elements in the Data Inventory were identified with the knowledge of domain experts from pharmaceutical companies. > This is not a conclusion. It could be better to elaborate on your second sentence: What is
available? What not?

12) Generally, the background could use more references.

13) Table 1 is unnecessary. It is the same as figure 2. If you want to keep it, data item and data group part should be indicated.

**Level of interest:** An article of importance in its field

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I participated in a research project with M Dugas between 2010 and 2012. University hospital Erlangen is one of the data providers in the EHR4CR project, but I am not personally involved.