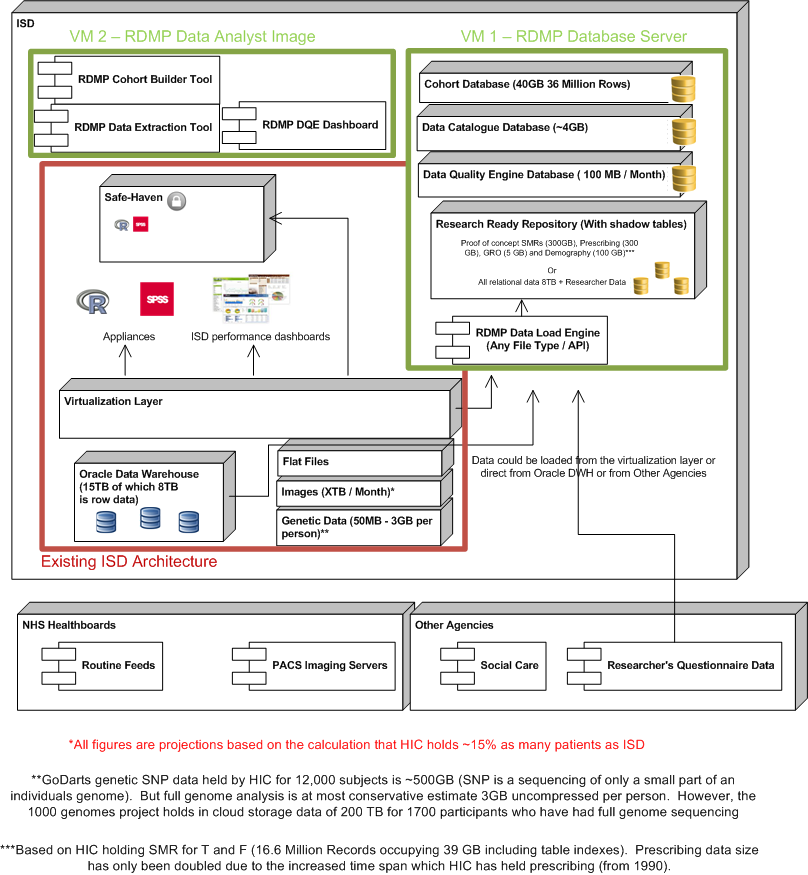
ISD RDMP Feasibility Investigation 2016

# Background

This document was created following a meeting on the 19-20th of January 2016 with HIC in which Carole Morris and Neil Sinclair outlined the ISD existing architecture and proposed virtualization layer to be implemented in 2016. During the second day of this meeting it was decided that it would be worthwhile investigating the potential of the HIC RDMP software to support the research data management lifecycle within ISD as a Research Extraction and Cohort Identification tool.



# Proposed VM Specifications and Roles

## VM 1 – RDMP Database Server

VM 1 is a Microsoft SQL Database Server. In addition to storing the RDMP system databases (Cohort Manager, Extraction Configurations, Logging etc) it stores a research ready subsection of the main ISD Oracle datasets.

## Requirements\*

* Microsoft .Net Framework 4.5
* Microsoft SQL Server 2014 (Standard or Enterprise)
* Microsoft Office (for data loading Excel files / generating reports)
* CPU Cores 8-16
* RAM: 32 – 128 GB
* OS: Windows Server 2012 or later
* Disk: 1TB for data + 300GB for OS / Software

## Advantages

A research ready version of a dataset can be optimised (with indexes) for cohort identification queries and research project aggregate generation.

The Shadow Copy feature of the RDMP software data loads allows for visualisation of changes in the main ISD dataset, reproduction of old extracts and explaining to a researcher why their data has changed since their last data release.

RDMP Software allows for more transparent cohort creation by rapidly linking datasets and showing cohort previews. This reduces errors due to poorly configured filters or miscommunication of researcher requirements.

## VM 2 – RDMP Data Analyst Image

VM 2 is a standard workstation image. Data analysts will use this PC to run the RDMP software tools (Catalogue Manager, Cohort Manager and Data Quality Dashboard) in addition to any tools they regularly use for research data extraction (e.g. Microsoft Office).

## Advantages

The RDMP tools are designed to support the research data management lifecycle and provide more support for cohort identification and data extraction than the current tools used by ISD (SPSS, R, Buisness Objects etc).

A dedicated Data Analyst Image with the RDMP analyst software tools allows for multiple system users at once and horizontal scaling of RDMP users. It also divides the roles of data loading into the research repository (done on VM1) and linkage/extraction (done on VM2).

## Requirements\*

* Microsoft .Net Framework 4.5
* Microsoft SQL Management Studio (for querying server and SMO)
* Microsoft Office
* CPU Cores 4
* RAM: 16 GB
* OS: Windows 7 or later
* Disk: 300GB for OS / Software / Working space

\*Requirements are based on a feasibility study loading the SMR datasets, GRO, Demography and Prescribing records for all datasets nationally. Storage requirements have been calculated based on the datasets held by HIC representing 15% of the nationally held data size.