D6.9 Mid-term report on data and knowledge management and sustainability

116019 – RESCEU
Respiratory Syncytial virus
Consortium in Europe

WP6 – Project management and outreach to stakeholders

<table>
<thead>
<tr>
<th>Lead contributor</th>
<th>Jeroen Aerssen (17 – Janssen)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><a href="mailto:JAERSSEN@its.jnj.com">JAERSSEN@its.jnj.com</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other contributors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harish Nair (1 – University of Edinburgh)</td>
</tr>
<tr>
<td>Harry Campbell, Rachel Reeve (1 – University of Edinburgh)</td>
</tr>
<tr>
<td>Philippe Beutels (2 – University of Antwerp)</td>
</tr>
<tr>
<td>Louis Bont (3 – University Medical Centre Utrecht)</td>
</tr>
<tr>
<td>Andrew Pollard, Simon Drysdale, (4 – University of Oxford)</td>
</tr>
<tr>
<td>Hannah Robinson, Joseph Mcginley, Gu-Lung Lin (4-UOXF)</td>
</tr>
<tr>
<td>Eva Molero (5 – Synapse)</td>
</tr>
<tr>
<td>Peter Openshaw (6 – Imperial College of London)</td>
</tr>
<tr>
<td>Adam Meijer (9 –NIPHE)</td>
</tr>
</tbody>
</table>

Reproduction of this document or part of this document without RESCEU consortium permission is forbidden. Any use of any part must acknowledge the RESCEU consortium as “RESCEU Real world Outcomes across the AD spectrum for better care: Multi-modal data Access Platform, grant agreement n°116019 (Innovative Medicines Initiative Joint Undertaking)”. This document is shared in the RESCEU Consortium under the conditions described in the RESCEU Consortium Agreement, Clause 9.
<table>
<thead>
<tr>
<th>Description of Action</th>
<th>Version</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoA amendment#1</td>
<td></td>
<td>5/06/2018</td>
</tr>
</tbody>
</table>
# Table of contents

Document History .................................................. 4  
List of abbreviations .............................................. 4  
Definitions .......................................................... 5  
Publishable Summary .............................................. 6  

1. INTRODUCTION .................................................. 7  

2. WORK PACKAGE SPECIFIC UPDATES ...................... 8  

2.1. WP1 – Systematic literature review on RSV and current estimates of burden of disease 8  
  2.1.1. Aggregate data ......................................... 8  
  2.1.2. Inventory ............................................... 8  
  2.1.3. Research explorer ...................................... 8  
  2.1.4. Edinburgh DataShare – description and data security 11  

2.2. WP2 – Consolidation of health care systems data 11  
  2.2.1. Strands of analysis ..................................... 11  
  2.2.2. Datatypes and location ................................. 12  
  2.2.3. Data management procedures ......................... 12  
  2.2.4. Data sharing ............................................ 13  

2.3. WP3 – Retrospective resource use analyses from existing databases/ networks 14  
  2.3.1. Pediatric Intensive Care data in The UK and the Republic of Ireland 14  
  2.3.2. Lille University hospital data in France ........... 14  
  2.3.3. Process and procedure ................................ 15  
  2.3.4. Screenshot .............................................. 16  
  2.3.5. Description of the main variables ..................... 17  
  2.3.6. Preliminary data on the WP4 cohort studies shared by WP4 with UAntwerpen 17  

2.4. WP4/5 - Prospective data collection and presumed risk factors and biomarkers for RSV-related severe disease and related sequelae. 18  
  2.4.1. Cohorts ................................................... 18  
  2.4.2. Data protection and reuse .............................. 19  
  2.4.3. Risk mitigation ......................................... 21  
  2.4.4. Pilot shipments and pilot assays ...................... 22  
  2.4.5. Data Management proposal ............................ 22  

2.5. WP6 - project governance and project afterlife management 23  
  2.5.1. FAIR ...................................................... 23  
  2.5.2. Sustainability ........................................... 23  

3. CONCLUSION AND NEXT STEPS ................................ 23  

ANNEXES .................................................................. 24  

ANNEX I. Directory Data Storage - WP5 24
Document History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>V0.1</td>
<td>15/05/2019</td>
<td>First Draft</td>
</tr>
<tr>
<td>V0.2</td>
<td>5/06/2019</td>
<td>Reviewed by PMO</td>
</tr>
<tr>
<td>V0.3</td>
<td>17/06/2019</td>
<td>External Review</td>
</tr>
<tr>
<td>V2.0</td>
<td>24/06/2019</td>
<td>Steering Committee revision</td>
</tr>
<tr>
<td>Final</td>
<td>16/07/2019</td>
<td>Final edits</td>
</tr>
</tbody>
</table>

List of abbreviations

AES. Advanced Encryption Standard
CHRUL. Centre Hospitalier Universitaire de Lille (Lille University hospital)
COPD. Chronic obstructive pulmonary disease
DRG. Diagnosis-Related-Groups
ENC. Echelle Nationale des Coûts
GCP. Good Clinical Practice
ICU. Intensive Care Unit
LOS. Length of stay
LRTI. Lower respiratory tract infection
PICANet. Paediatric Intensive Care Audit Network
PICU. Paediatric intensive care unit
PMSI. Programme de Médicalisation des Systèmes d'Information
RCUK. Research Councils United Kingdom
RSV. Respiratory syncytial virus
SD. Standard Deviation
UA. Universiteit Antwerpen (Belgium)
UEDIN. University of Edinburgh (United Kingdom)
UMCU. University Medical Centre Utrecht (Netherlands)
UOXF. The Chancellor, Masters and Scholars of the University of Oxford (United Kingdom)
URI. Upper Respiratory Tract Infections
UTI. Urinary Tract Infection
VAS. Visual Analogue Scale
WHO. World Health Organization
WHO-CHOICE. World Health Organization – Cost-effectiveness and strategic planning
Definitions

Participants of the RESCEU Consortium are referred to herein according to the following codes:

- **UEDIN.** University of Edinburgh (United Kingdom)
- **UA.** Universiteit Antwerpen (Belgium)
- **UMCU.** University Medical Centre Utrecht (Netherlands)
- **UOXF.** The Chancellor, Masters and Scholars of the University of Oxford (United Kingdom)
- **SYNAPSE.** Synapse Research Management Partners S.L. (Spain)
- **Imperial.** Imperial College of Science, Technology and Medicine (United Kingdom)
- **SERGAS.** Servicio Galego de Saúde (Spain)
- **TUCH.** Varsinais-Suomen sairaanhoitopiirin kuntayhtymä (Finland)
- **RIVM.** Rijksinstituut voor Volksgezondheid en Milieu - National Institute for Public Health and the Environment (Netherlands)
- **SSI.** Statens Serum Institut (Denmark)
- **UMCG.** Academisch Ziekenhuis Groningen (Netherlands)
- **PENTA.** Fondazione PENTA for the treatment and care of children with HIV-ONLUS (Italy)
- **AZ.** Astrazeneca AB (Sweden)
- **Pfizer.** Pfizer Limited (United Kingdom)
- **GSK Bio.** GlaxoSmithKline Biologicals S.A. (Belgium)
- **SP.** Sanofi Pasteur (France)
- **JPNV.** Janssen Pharmaceutica, N.V (Belgium)
- **Novavax.** Novavax Inc. (United States of America)

- **Grant Agreement.** The agreement signed between the beneficiaries and the IMI JU for the undertaking of the RESCEU project (116019).
- **Project.** The sum of all activities carried out in the framework of the Grant Agreement.
- **Work plan.** Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried out, as specified in Annex I to the Grant Agreement.
- **Consortium.** The RESCEU Consortium, comprising the above-mentioned legal entities.
- **Consortium Agreement.** Agreement concluded amongst RESCEU participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties’ obligations to the Community and/or to one another arising from the Grant Agreement.
- **DMP.** Data management plan
Publishable Summary

This report entails the updated data and knowledge management plan, including screenshots, identified risks, mitigation steps and implemented procedures for data entry, safety, data protection and folder structures.

This deliverable reported how data has been stored, shared and prepared to be made publicly available. In summary WP1 uses Edinburgh Datashare, WP2 uses strict data sharing procedures set in place by the participating organisations and data sharing is limited to tables and reports, WP3 mainly uses publicly available data on burden of disease and cost. Clinical data based on questionnaires and case-report forms are being collected for the four clinical studies in WP4. During the writing of the deliverable report, around 5,500/10,000 participants have been recruited for the infant cohort study, 1,040 (complete recruitment) for the older adult study, 182/630 for the infant case-control study and 339/500 for the COPD study. Data are collected in Research Online and OpenClinica used to store and share the clinical data from case-control and cohort studies. Data file exports are being set using Surffilesender. WP5 aims to put in place a data storage and management plan to use the data generated from clinical samples. Clinical data are managed centrally while for the sample data information is captured by each individual site. Although not present yet, a commonly accessible folder structure for both clinical data as well as sample information is desired and is currently being investigated. WP6 identified the roadmap for different scenarios on sustainability measurements and data management beyond the project. A short intro on the work in progress is given in the final sections of the report.
1. Introduction

Halfway through the project, WP6 is leading the implementation of the strategy for knowledge and data management. Partners have done the exercise on how data will be shared and made publicly available. How these assets will be protected and accessed. Complex dataset and sample governance requirements as well as ethical approvals to access data and facilitate conduct data linkage are necessary and should follow proper regulations and procedures. This deliverable report determines mechanisms on how generated data and samples are handled during and beyond the project.
2. Work Package Specific updates

2.1 WP1 - Systematic literature review on RSV and current estimates of burden of disease

2.1.1 Aggregate data

Literature reviews are being conducted as part of WP1 to understand the global epidemiology of RSV and to assess disease burden in various risk groups. Aggregate data (through literature review or establishing investigator groups) or individual patient data (through investigator groups) are being collected in a pre-designed format to permit assembling these data into a larger pooled database. These data are global in origin, collected from studies funded by research councils, charitable trusts / foundations, tax payers and industry. Existing data are being re-analysed using common case definition and pre-specified analytic approaches.

2.1.2 Inventory

Data extraction from published literature is done by two independent reviewers. For unpublished data provided by the primary study investigator part of the investigator group, the data are entered in a common MS Excel template by a local data assistant and validated by a RESCEU researcher. When there are inconsistencies in data, queries are sent to investigator and data are “locked” only when the queries are resolved satisfactorily. The data are stored on Edinburgh Datastore which is a server accessible only to University staff (using their unique login credentials) and only when on University’s network.

To date, 9 of the 13 deliverables from this WP have been completed and submitted to IMI. Three of the submitted deliverables have been published. The underlying data for these publications would be useful to scientists, industry and policy makers if they wanted to replicate the analysis / update these in future.

2.1.3 Research explorer

Therefore, the final versions of the data and associated metadata and documentation associated with these publications have been deposited in an institutional repository – Edinburgh DataShare (http://datashare.is.ed.ac.uk/) and are housed and available in accordance with the FAIR data standards. The data are uploaded to Edinburgh DataShare at the time of submission of manuscript to a peer reviewed journal but are under embargo until publication of the manuscript (Fig 4). The data are made available publicly when the manuscript is first published online (Fig 1-3). The data are searchable online in Edinburgh Research Explorer under RESCEU project.
The individual datasets can be seen below. Each dataset has an unique identifier (DOI). The data made publicly available are in MS Excel format and freely downloadable.
Figure 3: Dataset for Deliverable D1.2

Figure 4: Dataset for Deliverable D1.3

Figure 5: Data related to deliverable D1.10 (under embargo). The data are not publicly accessible at the moment.
Data bear the name “RESCEU” and search keywords, data dictionaries, case definitions and underlying analytic codes are provided. All data included in the analysis are provided in an aggregate format by narrow age bands (and gender or other groups if applicable).

2.1.4 Edinburgh DataShare – description and data security

The Edinburgh DataShare repository supports open access and data are accessible through any web browser (IE, Chrome, Firefox, Safari etc.). No specific software other than MS Office (Excel) is required. Where data are solely / primarily from literature review, there are no restrictions on use and a data access committee is not needed. However, where substantial amount of unpublished data (from investigator groups) are included, the data requests from third parties will be processed through a data access committee, as some of the investigators may have provided data that are hitherto unpublished (i.e. from ongoing studies). Once the contributing studies have published their data, the full dataset will be made openly accessible. The data are made available through a Creative Commons license – the user will need to acknowledge RESCEU. The data will be licensed using an open data license to the dataset; Creative Commons Attribution 4.0 International license (CC-BY 4.0) is the default license.

Edinburgh DataShare is an online digital repository of multi-disciplinary research datasets produced at the University of Edinburgh, hosted by Data Library in Information Services. It is designed to be scalable and highly resilient. User support is provided by Data Library staff. The Edinburgh DataShare repository contents will be backed up to tape daily and database snapshots will be available on disk for two days. Daily tape backups are retained for 30 days and a monthly archive copy is kept for a year. The Edinburgh Datashare is compliant with the guidelines of Research Councils, UK (RCUK).

2.2 WP2 - Consolidation of health care systems data

2.2.1 Strands of analysis

WP2 seeks to assemble large-scale (e.g. national) healthcare data on RSV healthcare burden with the goal of developing a conceptual framework to assess RSV healthcare impact across multiple groups. Task 1 in WP2 focuses on assessing the healthcare burden of RSV in at least seven EU countries (Denmark, Netherlands, Finland, Scotland, England, Italy, and Norway). This task includes four main strands of analysis:

1. Direct measurement of RSV healthcare burden in young children (<5 years).
2. Estimate RSV disease burden in all age groups and certain subgroups and risk groups (e.g. children and adults with chronic medical conditions) using time-series modelling.
3. Estimate the association between RSV/bronchiolitis episodes in young children and subsequent episodes of wheeze, asthma, lower respiratory tract infection (LRTI) episodes and
pneumococcal disease (invasive pneumococcal disease or pneumococcal pneumonia) in childhood.

4. Estimate the healthcare costs associated with RSV infection in hospitalised and primary care patients in all age groups (including children, the elderly, and other high-risk groups). This work will feed into Work Package 3 (WP3) of the RESCEU project (economic burden of RSV).

Secondary research questions will also be developed and agreed by the contributing partners.

### 2.2.2 Datatypes and location

WP2 will not prospectively collect new data, but project partners will access data routinely collected by national health services, health providers, national registries for births and deaths, and national clinical audit groups in the seven EU countries. These data include individual-level data on hospital admissions, GP consultations, mortality data, laboratory surveillance data, birth records, maternity data, prescribing data, and financial data. The majority of this data is at the national level, except in some cases it is regional only or a nationally representative sample.

Routine health data within each partner country will, subject to approvals, be accessed by RESCEU partners within that country under strict governance and security regulations. Permission to access this type of data is typically granted under license, which can only be used for the research specified in the research plan attached to the application for authorisation. The approval process within each country is different but requires ethical approval and usually some review of data users (such as, for example, training in information governance) before approval is granted. Any significant changes in the research plan or research organisation are reported to the health data controller organisations, who will consider whether the change requires a new license. Linkage of datasets is usually done through a trusted third party. In some countries, approval for the requested RESCEU data was required from multiple organisations or data controllers.

### 2.2.3 Data management procedures

RESCEU WP2 will make use of employer (e.g. university) and/or health data controller organisations’ standard data management procedures when accessing routine health data within each partner country. Each partner country in WP2 will have to adhere to the strict data security standards and regulations set in place by their data controller organisations. In general, this will include:

- Anonymised data records,
- Controlled access to the anonymised research datasets within a safe setting within the data controller organisation,
- Datasets will be stored on secure servers with regular back-up by the data controllers.

Data used by the RESCEU project for health care systems analyses will remain under the control of the data controller organisations and will not be owned by RESCEU WP2 researchers. The data will not be findable or be made openly accessible by default as health data remains confidential and controlled by responsible national bodies or organisations, which balance public privacy and patient confidentiality with the re-use of routinely collected health data for secondary use such as research.
2.2.4 Data sharing

Data exchange between RESCEU partners will be limited to summary reports and tables only; individual-level or potentially identifiable information cannot be shared between countries. There are also strict data protection rules on reporting and publishing of results using this type of data. Reporting and publishing results must follow these rules, as well as general ethical guidelines. In most countries, all summary reports and tables will also be subject to scrutiny by the national data controllers before release of results from the safe settings within the data controller organisation. Many countries place a restriction on the minimum data points per cell in tables extracted from the safe settings: in some countries no cells with $n<5$ are able to be released, with some countries placing stricter restrictions of no cells with $n<10$.

The data will not be available for exchange with any RESCEU external researchers or organisations and will not be interoperable outside of RESCEU project partners. The datasets accessed and used by RESCEU WP2 researchers within each partner country will use standard data variables, definitions, format and coding systems to be compliant with a standardised analysis application. Data dictionaries, case definitions and detailed analysis plans created by RESCEU will be made available online to encourage uptake and use by other European countries (and other international groups) in the future. The results of the WP2 analyses will be made available at www.resc-eu.org.

The datasets may be useful to RESCEU researchers within each partner country for further detailed analysis relating to RSV, in accordance with their granted approvals. Data will be safety stored in certified repositories for long-term preservation and curation. The details of this will depend on each partner country’s governance and legislation for agreed length of time for storage of each research dataset. The processes will be strictly defined by the data governance authorities in each country. Usually, after the expiry of the authorisation the research material enabling the identification of an individual (i.e. all individual-level data) must be destroyed.
2.3 WP3 - Retrospective resource use analyses from existing databases/networks

Most data handled in WP3 until now are publicly available data on burden of disease and costs whereas some of the economic models shared by individual companies are not yet published or publicly available. In 2019, three non-public database extractions are accessed by academic WP3 partners and affiliate partners.

2.3.1 Pediatric Intensive Care data in The UK and the Republic of Ireland

This study by affiliate partner University of Cambridge (Prof Caroline Trotter) is described as part of D3.2.

PICANet (Pediatric Intensive Care Audit Network) is an international audit that collects information on all children admitted to paediatric intensive care in the UK and the Republic of Ireland (http://www.picanet.org.uk/). PICANet has permission to collect patient identifiable data under section 251 of the NHS Act 2006 (originally enacted under Section 60 of the Health and Social Care Act 2001). Demographic details from the Republic of Ireland are anonymised in accordance with the requirements of the Irish Data Protection Officer. Researchers wishing to access this data must first apply to the PICANet Clinical Advisory Group, and then to the Health Quality Improvement Partnership (HQIP) using standard forms. Data received for analysis will be anonymised.

The economic cost of PICU bed days will be estimated from standard public sources; including the Unit Costs of Health and Social Care (Curtis & Burns 2016) and NHS reference costs (Department of Health 2016).

A proposal to PICANet for this study was submitted, and access has been granted. The data requested do not include personal identifiers, they are classed as “limited access de-identified data” and must be hosted on a secure server. To this end the affiliate partner at UCAM is using the University of Cambridge’s Clinical School Secure Data Hosting Service. This is an ISO:27001 certified Safe Haven for members of the University to store sensitive data. Given the data security measures in place, it is impossible to take screen shots of the data to share with RESCEU.

The affiliate partner at UCAM will be able to access the data - on the secure server within the Safe Haven only - starting 1st May 2019 and ending 31st December 2019.

2.3.2 Lille University hospital data in France

This study is led by affiliate partner, University of Lille (Prof Benoit Dervaux). The study is described as part of D3.2.

As any French hospital since 1991, the CHRUL has produced information on its activity based on the French Diagnosis-Related-Groups (DRG) system (Programme de Médicalisation des Systèmes d'Information, PMSI). For each patient, all diagnoses and main procedures are encoded throughout the hospital stay. The French DRG classification has been continuously adapted through time. A big change occurred in 2009 with a split of DRGs to take into account the severity of inpatient cases.
More important for the research project, the CHRUL has also contributed since 2007 to the French hospital reference costs database (Echelle Nationale des Coûts, ENC). This database is completed by a small sample of French hospitals (n=100) which have analytical accounting. These hospitals are able to estimate the cost of each stay. Concerning the pediatric unit, 240 beds are installed, and 28000 emergency visits are realized per year. The study population is defined as hospitalized children with ARI consulting the CHRUL pediatric emergency unit or directly admitted to the pediatric intensive care unit. Until 2015, all cases were sampled. From 2016, local guidelines have changed. Virological diagnosis is no longer systematic.

Given above information, the study period is set to 2010-2015. To fit with research objectives, following data need to be collected: biological and medical information (symptoms at admission, diagnoses, virological diagnosis, course of disease, respiratory sequelae...), resources used (total length of stay (LOS), LOS in intensive care units, medical acts and drugs, overall hospitalization costs...). These data will be extracted from: 1) the patients’ medical files, 2) PMSI database, 3) ENC database. Regarding information from medical files, a data collection form has been elaborated. Databases’ matching is possible using patients’ identification number.

The study is monocentric with no transfer of individual data outside the CHRUL, and the final database is encrypted and made anonymous. The protocol of the study has been approved by the local Data Protection Officer (Reference: DEC16-274).

### 2.3.3 Process and procedure

The data controller was Prof. François Dubos who is a pediatrician, infectious disease specialist and the chief of pediatric emergency department of Lille University Hospital.

The data controller was in charge of:

- The collection of clinical data from medical records in a dedicated and secured CRF
- The linkage of medical and cost databases. The primary key was the internal hospitalization identifier
- The anonymization process: all directly or indirectly identifying information was definitely deleted from the dataset

In compliance with French regulation, the database consisted of reused data already collected as part of medical care and was limited to the data absolutely necessary for the completion of the study objectives.

Databases were encrypted with AES-256 keys. They were locally stored in separate folders. Access to the local drive was restricted to authorized individuals and required secured and nominative identification. The local drive was kept in a locked room with key.

The data processor was in charge of data management and data processing. The data processor only had access to the anonymized database. The anonymized database was encrypted with an AES-256 key and stored on a local drive, different from the one used by the data controller. Its access required secured and nominative identification as well.
These various steps are summarized below:

After data management, RESCEU-Lille_Database was made of 1592 individuals (lines) and 61 variables. The size of the table in .csv format was 506 ko.

2.3.4 Screenshot
2.3.5 Description of the main variables

<table>
<thead>
<tr>
<th>variable</th>
<th>description</th>
<th>type</th>
<th>details</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEP2</td>
<td>sequential number for hospitalization</td>
<td>numerical</td>
<td></td>
</tr>
<tr>
<td>IPP2</td>
<td>sequential number for patient</td>
<td>numerical</td>
<td></td>
</tr>
<tr>
<td>YEAR_OUT</td>
<td>year of discharge</td>
<td>numerical</td>
<td></td>
</tr>
<tr>
<td>logLOS</td>
<td>log-transform of length of stay</td>
<td>continuous</td>
<td></td>
</tr>
<tr>
<td>DRG</td>
<td>diagnosis related group</td>
<td>string of 6 characters</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>patient's age at the entry date</td>
<td>continuous</td>
<td>Days/365.25</td>
</tr>
<tr>
<td>group</td>
<td>virology result</td>
<td>categorical (4 groups)</td>
<td>RSV- other-/RSV+ other-/RSV- other+/RSV+ other+</td>
</tr>
<tr>
<td>preterm</td>
<td>prematurity</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>resp_COMORB</td>
<td>respiratory comorbidity</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>nonresp_COMORB</td>
<td>non respiratory comorbidity</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>ER</td>
<td>admission via the emergency department</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>ICU</td>
<td>admission via the ICU</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>fever</td>
<td>fever</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>sibilants.wheezing</td>
<td>wheezing</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>O2</td>
<td>oxygen therapy</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>resp_support</td>
<td>respiratory support</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>XRay</td>
<td>XRay</td>
<td>categorical (3 groups)</td>
<td>no xray/normal/abnormal</td>
</tr>
<tr>
<td>superinf_pulm</td>
<td>pulmonary superinfection</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>superinf_NONpulm</td>
<td>non pulmonary superinfection</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>complic_pulm</td>
<td>pulmonary complication</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>complic_nonpulm</td>
<td>non pulmonary complication</td>
<td>binary</td>
<td></td>
</tr>
<tr>
<td>COUT.hors_structure</td>
<td>cost per admission, without real estate</td>
<td>numerical</td>
<td>2015 os</td>
</tr>
</tbody>
</table>

2.3.6 Preliminary data on the WP4 cohort studies shared by WP4 with UAntwerpen

RESCEU partners involved in WP4 (and WP5) on prospective data collection have developed protocols for four studies: an active and a passive birth/infant cohort study, an infant case-control study and older adult cohort study, and a study in COPD patients. For the extended protocols of these prospective studies, including ethical issues, we refer to WP4 deliverable D4.1.

UAntwerpen received in April 2019 preliminary anonymized data files from WP4 from two studies (for details on variables contained in these databases we refer to WP4 reporting).

a) The “older adults” database consists of data from season one (2017-2018) and season two (2018-2019). This database includes:

- baseline CRF + questionnaire (both seasons)
- ARTI CRF’s (both seasons)
- diary data (both seasons)
- End-of-season CRF + questionnaire (Season 1 only).

b) The “infant cohort” database consists of data up to the current season and includes data from season one (2017-2018) and season two (2018-2019). The database includes:
The databases are cleaned but not yet checked for errors by WP4, and incomplete: some forms (season 2) have not been entered (yet) by WP4 and are therefore not included in the preliminary database received thus far by UAntwerpen.

These preliminary databases will be used for data checks and preliminary analyses by UAntwerpen. They are stored on the internal FileShare system of the University of Antwerp (see screenshot). This system is protected by the University’s central Firewall and is additionally protected by registered user-only access, that is granted by PI Philippe Beutels. This system is physically kept in a locked room at UAntwerpen. Only registered users (currently only P.I. Philippe Beutels and postdoctoral researcher Lander Willem) can access this folder, and their access times are logged by the system. These data will be analysed on-site in the secure software environment - behind the secure firewall, in a registered user-only environment - to fit statistical models and create summary statistics. None of these anonymous data, except summary model specifications and statistics, will be stored by the registered users outside of the Fileshare system.

![Screenshot of the protected folder on the University of Antwerp FileShare system containing data from WP4. Only P.I Prof. dr. Philippe Beutels and dr. Lander Willem have access.](image)

**2.4 WP4/5 - Prospective data collection and presumed risk factors and biomarkers for RSV-related severe disease and related sequelae.**

**2.4.1 Cohorts**

To gain an insight into the burden of RSV disease and biomarkers of disease severity and sequelae, data will be collected from prospective cohort studies (infant, older adults, and COPD populations) and an infant case-control study. Prospective data on the incidence of RSV, medically attended (MA)-RSV illnesses and RSV hospitalizations will be collected, as will information on risk factors for RSV disease, health care usage and quality of life. Data will be collected from biological samples, parental questionnaires and clinical data. Medical data will be obtained from hospitals and general practitioner offices and using information directly obtained from the patient. Data will be collected from the
following studies:

- Infant cohort study: 10,000 children (1000 questionnaires and biological samples, 9000 only questionnaires)
- In the active cohort (questionnaires and biological samples) 638/1000 patients are recruited (June 15th, 2019). For the passive cohort (only questionnaires) 4960/9000 patients are recruited. Estimated full recruitment for the active cohort will be reached at the end of 2019 while recruitment for the passive cohort is estimated to be finished by June 2020.
- 630 case-control children (biological samples and questionnaires) Currently, over two seasons 182/630 patients have been recruited. Since this is not sufficient, the study will be extended for another RSV season (2019-2020).
- Older adult cohort study: 1,000 older adults (biological samples and questionnaires) Over the past two seasons 1040 participants have been recruited and the study is due to be finished in the coming three to four months. Study end-visits of the cohort 2018-2019 will be performed in May-June.
- 500 adults with COPD (biological samples and questionnaires) Currently, 339/500 COPD patients have been recruited in the study. Recruitment will continue until October 2019.

Best practices for processing and analysis of several thousand biological samples have been described in D6.3.

Studies will be fully compliant with Good Clinical Practice (GCP). Data collection and security described this D6.3 defined the role-based access to the systems to avoid unauthorized data access and prevent users performing unauthorised actions. With respect to FAIR data and long-term sustainability guidelines are set in place to align and fuel Data interoperability between researchers of the RESCEU consortium. Research Online and OpenClinica databases allow to extract many formats to facility uniform collections on standardized mechanisms

### 2.4.2 Data protection and reuse

The clinical data generated by the infant case-control study will be stored as an excel file (csv format) on a secure network drive within the university servers and will be made available on the RESCEU SharePoint for all collaborators.

For the clinical studies data management to date has proceeded according to plan. Data from the clinical cohort studies are being collected and stored in Research Online.

**Tool used for data collection**

Data from the infant case-control study are being collected and stored in OpenClinica, a commercial open source clinical trial software for electronic data capture (EDC) in clinical data management. For Research Online data are entered by the researchers of each site and, depending on the preference of the local participating site, online questionnaires can also be filled in by the participants themselves. Each participant receives a unique link which they can use to complete their questionnaires. Local investigators have role-based accounts for the online database and can see and enter data for their own local site.
Tool used for data management

The 2 sites of the COPD cohort study are using different databases: Imperial is using their existing COPD-database while UMCG is using Research Online. A common minimal dataset has been established by both sites and is being asked in a similar format, in order to be able to merge this information afterwards.

Tool used for data dissemination with RESCEU partners

For the infant cohort study and the older adult’s cohort study, exports (of raw data) are being made every couple of months with two data exports being performed so far. The most recent export of the infant cohort study and the older adult’s study has been circulated in April 2019 to partners that indicated interest in receiving this data. Data exports are being sent using Surffilesender (a protected online tool by which data can be transferred safely). All exports are password protected and the password is sent through another channel (text message) to those who receive the databases.

Data are exported in Excel compatible (.xlsx) and SPSS compatible (.SAV) files:

Data are consequently coded as zero for “No” and one for “Yes”. Data are sent around with the labels indicating the meaning of multilevel variables. All variables include a variable description in the SPSS files and labels are included in the Excel files distributed. However, we are aiming to standardize coding, making it easier to merge files and easily interpret the data. Currently, there is no commonly accessible folder structure in place to store this generated data exports of clinical data.

Tool for sample shipment

For the clinical samples that are collected in the clinical studies, each site is keeping track of their own collected samples. An overview of the collected samples for the infant cohort and older adults cohort study is accessible through Research Online but for precise amounts of aliquots and number of
processed samples this is insufficient (it merely logs whether the samples are collected at all, e.g. serum “Yes”/”No”). Shipments are being performed currently for multiple types of samples in the four clinical studies. All shipments are coordinated and recorded by WP4/WP5. Standardized Excel sheets are used to log the shipments that take place in which information about the sample (e.g. subject/informed consent/barcodes/date of collection/sample type/visit type etc.) is being logged as well as logistical data (originating site details, receiving site details, type of analysis to be performed etc.). See screenshot below:

As discussed in a recent work package 5 meeting, a folder structure and overall database of the collected and processed samples with their current location is desirable but is not yet in place.

### 2.4.3 Risk mitigation

WP5 has identified the risk of sample loss due to data corruption or damage and mitigated this risk by backing up to multiple drives and implementing the encryption of folders protected with passwords. Data was acquired after some preliminary analysis. Multiple compressed files were acquired containing the data. A mechanism has been used for this particular aspect specifically but is not a consortium wide agreed approach yet: Data was stored in a home directory "Pilot_Study_results" which contains multiple subdirectories with varying permutations of samples used, screenshots in Annex 1. Folder containing all results from transcriptomics study thus far, including some outputs of analysis has a total size of 387Mb. During the 3rd RESCEU General Assembly (19-21 June, Utrecht, The Netherlands) a specific session on data management will allow to find a common approach and consortium wide accepted way forward for data storage and hosting.
2.4.4 Pilot shipments and pilot assays

Pilot shipments and pilot assays were performed for transcriptomics (Janssen), pre-F post-F ELISA (Janssen), neutralizing antibodies (Sanofi), and RSVA/B RT-PCR testing (GSK). Shipments are coordinated by WP4 and WP5. Risk of data loss due to data corruption or damage and security have been mitigated with backup of multiple drives and implementation of folder encryptions and security procedures. Sequencing data were provided by the Wellcome Centre for Human Genetics. Data were compressed as bam or gz files accordingly. Primary data are stored in the desktop computer at the Oxford Vaccine Group with multiple backup. Data was acquired after some preliminary analysis by Janssen. Multiple compressed files were acquired from Janssen containing the data. Data was stored in a home directory "Pilot_Study_results" which contains multiple subdirectories with varying permutations of samples used. The size of the sequencing data is 18.4 GB. The outcomes of pilot assays were reported in internal WP meetings and currently stored in WP5 temporary folder (BOX).

2.4.5 Data Management proposal

Janssen prepared a draft proposal on how to store the WP5 data. The storage platform (named BOX) is provided by Janssen and open to be shared with the external parties. This is the same platform where the current pilot data is stored. The platform is intuitive and user-friendly. The platform is suitable for type I data (no patient identifying data) and suitable for larger data files such as the ones generated from omics type of experiments. The proposal of data management and storage requires each partner to provide a minimal set of “meta-data” fields. The meta-data fields (such as study, assay type, platform type, sample type, sample acquisition method, sample collection day) are communicated for each partner to be able to meet the requirement of each data type generated in different labs. This information will be stored in ‘yaml’ files.

The screenshot below shows the draft folder structure and how to name each data.
2.5 WP6 - project governance and project afterlife management

2.5.1 FAIR

As indicated in this deliverable report individual work packages have applied storage techniques that are compliant with good practice in line with ethical and regulatory procedures. The consortium is doing the exercise to find common ground in these mechanisms that would allow to have one single guidance document on storage of the data. WP4/5 data generators have agreed to follow this guidance document enabling to store data decentralized but in the common format that would enable interoperability later on the roadmap. A virtual biobank solution is now topic of discussion to answer the request of live tracking of the decentralized samples.

2.5.2 Sustainability

RESCEU is pulling together a few scenarios to develop a sustainable data platform that should enable to quantify healthcare and economic impact of RSV (all ages and key risk groups) including sequelae at regional and national levels after the funding period has ended. The selected scenario needs to meet the criteria of providing high-quality, sustainable, robust data collection systems that link closely with public health/regulatory bodies/health care providers for informing policy and regulatory processes. The scenarios’ will be shared with full consortium during the General Assembly meeting in Utrecht. Shortly thereafter the consortium will set actions and roadmaps to fuel this initiative and revert on this development during the next DMP and sustainability report.

3. Conclusion and next steps

RESCEU is well in the picture for IMI sustainable approaches and is reaching out to manifest the FAIR implementation of best practices to datasets and samples generated at the beneficiary’s site. At the moment, the partners are coming up with solutions to maintain the acquired knowledge, samples and digital assets beyond the life span of the project. IMI project output sustainability and impact longevity are a key focus point. Thinking of project afterlife management will need emphasis to sustainable financial business model planning in line with ethical, legal and IT buy -in on the consortium level. The intent is to make sure that on fair and reasonable terms the consortium outputs need to be findable, accessible and interoperable to allow data and sample reuse and drive interactions for future partnerships in the field of RSV.
ANNEXES

ANNEX I. Directory Data Storage - WP5