

Deliverable 6.3: Data and Knowledge Management Plan

**Action Full Title:
REspiratory Syncytial virus Consortium in EUrope**

Action Acronym: RESCEU

- **Grant Agreement no.: 116019-3**
- **IMI2 Call topic: IMI Call 2020-JTI-IMI2-2015-06-02: Establishing impact of RSV infection, resultant disease and public health approach to reducing the consequences**
- **Name of the coordinating person: Harish Nair**
- **e-mail address: harish.nair@ed.ac.uk**

Document version 1
Date: 26.06.2017

Lead contributor	AZ (13)
	DillonL@MedImmune.com
Other contributors	Harish Nair (1 – University of Edinburgh)
	Harry Campbell (1 – University of Edinburgh)
	Philippe Beutels (2 – University of Antwerp)
	Louis Bont (3 – University Medical Centre Utrecht)
	Andrew Pollard (4 – University of Oxford)
	Eva Molero (5 – Synapse)
	Peter Openshaw (6 – Imperial College of London)
	Lieke Sanders (9 - National Institute for Public Health and the Environment, Bilthoven, the Netherlands)
	Judith Hackett (13 – AZ)
	Charles Knirsch (14 – Pfizer)
	Antonio Gonzalez Lopez (15 – GSK)
	Scott Gallichan (16 – Sanofi)
	Jeroen Aerssens (17 – Janssen)
	Brian Rosen (18 – Novavax)

Due date	30/06/2017
Delivery date	26/06/2017
Deliverable type	R
Dissemination level	PU

Description of Work	Version	Date
	V1	26/06/2017

Table of Contents

Table of Contents	3
Document History	4
List of Abbreviations	4
RESCEU Data and Knowledge Management Plan (DKMP)	5
1. Data Summary and FAIR (Findable, Accessible, Interoperable, Re-usable) data standards	5
2. Allocation of resources	8
3. Data security	8
4. Ethical aspects.....	9

Document History

Version	Date	Description
V0.1	16/05/2017	First Draft
V0.2	25/05/2017	Comments from ISAG
V0.3	26/06/2017	Review from SC
V0.4	26/06/2017	Final Version

List of Abbreviations

COPD	Chronic obstructive pulmonary disease
DKMP	Data and Knowledge Management Plan
DoA	Description of Action
EDC	Electronic Data Capture
EU	European Union
FAIR	Findable, Accessible, Interoperable, Re-usable
GCP	Good Clinical Practice
IE	Internet Explorer
IMI	Innovative Medicines Initiative
MA	Medically attended
MS	Microsoft
RESCEU	REspiratory Syncytial virus Consortium in EUrope
RCUK	Research Councils UK
RSV	Respiratory Syncytial Virus
WP	Work package

RESCEU Data and Knowledge Management Plan (DKMP)

The RESCEU Consortium is committed to open data access as described in the Horizon 2020 Guidelines for Research Data Management in order to optimize the use of the RESCEU resource for the broader scientific community. As multiple data types will be produced by the RESCEU Consortium, it is inappropriate for the project to use a one-size-fits-all approach to data management. Instead, the management of data from each Work Package (WP) will be handled by that WP and led by the respective lead institutions for each WP according to the specific objectives and needs of the WP. All WPs will seek to adhere to FAIR (Findable, Accessible, Interoperable, Re-usable) data standards to the extent possible, as described below for each WP. As the clinical samples obtained by clinical studies in WP4 will be used for biomarker analyses in WP5, data management for WP4 and WP5 will be considered together.

1. Data Summary and FAIR (Findable, Accessible, Interoperable, Re-usable) data standards

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

The literature reviews to be conducted as part of WP1 are being done to understand the global epidemiology of RSV and 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 through studies funded by research councils, charitable trusts / foundations, tax payers and industry. Existing data will be re-analysed using common case definition and pre-specified analytic approaches. Data are expected to be at least 5 GB in size. Data would be useful to scientists, industry and policy makers.

Final versions of the data and associated metadata and documentation produced by WP1 will be deposited in an institutional repository – Edinburgh DataShare (<http://datashare.is.ed.ac.uk/>), and will be housed and available in accordance with the FAIR data standards. The data will be made available along with the publication of the manuscript. Manuscripts will be published within the timescales detailed in RESCEU DoA. Data will bear the name “RESCEU” and search keywords, data dictionaries and case definitions will be provided. All data included in the analysis will be provided in an aggregate format by narrow age bands (and gender or other groups if applicable).

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 would be 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 will be made available through a creative common licence – 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.

WP2: Consolidation of health care systems data

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. WP2 will not prospectively collect new data but project partners will access data routinely collected by national health services, health providers, national registers for births and deaths, and national clinical audit groups in several EU countries. Routine health data within each partner country will, subject to approvals, be accessed by

RESCEU partners under strict governance and security regulations. The datasets may be useful to RESCEU researchers within each partner country for further detailed analysis relating to RSV.

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. The data collected for WP2 Task 1 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. The health data controllers in most countries will be other national organisations who will create anonymised linked datasets available to the RESCEU researchers, usually within safe haven settings and under strict governance rules. RESCEU researchers in most partner countries will have to submit detailed applications for approval by the respective data controller organisations to gain access to the health data. 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 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 exchange between RESCEU partners will be limited to summary reports and tables only; and even these will generally be subject to scrutiny by the national data controllers before release. However the data will not be available for exchange with any external researchers or organisations outwith RESCEU and will not be interoperable outside of RESCEU project partners.

The results of the WP2 analyses will be made available at www.resc-eu.org. Data dictionaries, case definitions and detailed analysis plans created by RESCEU will also be made available online to encourage uptake and use by other European countries (and other international groups) in the future.

WP3: Retrospective resource use analyses from existing databases/networks

RESCEU WP3 researchers will identify, collect and compile existing data using statistical, mathematical and economic methods. New data will not be collected by WP3. Instead, WP3 researchers will combine multiple data sources (e.g. disease and cost burden estimates, cost-effectiveness estimates) generated in other WPs of RESCEU, as well as freely available existing data (such as EUROSTAT data) and existing data obtained from government, private and administrative databases. The analytical results of WP3 will be useful to the scientific and medical community, pharmaceutical industry and policy makers.

RESCEU-generated data used by WP3 partners will be stored and managed as described by the source WPs. The data WP3 uses from existing data sources will be stored by the researchers who will develop models to compile the data according to the activities described for WP3. Standard data software (R, SAS, Stata, Excel, Access) will be used to store and manage data that can be stored in accordance with the original data sources. For some data sources, data will be consulted on-site, and no data from such sources, except summary statistics, will be stored by WP3 collaborators. For any off-site manipulation of data, standard data formats will be used with a granularity that is bound by the legal contract of data usage from the original source. After compilation, modelled data (i.e. the results of WP3 analyses) will be freely available, and made part of publications, such that they are de facto part of public repositories (i.e. academic publications and their supplementary online material). These publications and their accompanying data repositories will be fully searchable according to guidelines provided in the accompanying publications (including specific terminology where appropriate). Published articles will contain RESCEU (consortium) in their Title/abstract. RESCEU will be noted as a keyword. The results from WP3 will be fully interoperable and re-usable after the project ends. The data inputs will be re-usable in the summary formats in which results are reported.

The results of the WP3 analyses will be made available at www.resc-eu.org.

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

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 cohorts:

- 10,000 children (1000 questionnaires and biological samples, 9000 only questionnaires)
- 1,000 older adults (biological samples and questionnaires)
- 630 case-control children (biological samples and questionnaires)
- 500 adults with COPD (biological samples and questionnaires)

For WP5, the processing and analysis of several thousand biological samples (viral and microbiome nasopharyngeal swabs, stool, urine and blood) will generate very large amounts (several terabytes) of metadata (e.g. transcriptomic, metabolomic, proteomic, genetic, microbiome, viral sequencing, etc.).

Studies will be fully compliant with Good Clinical Practice (GCP). Data collected in the cohort studies will be stored in Research Online and patient data collected in the infant case-control study will be stored in OpenClinica, online secured, GCP acceptable databases. Data will be available for researchers within the consortium during the course of the study. A data access committee will be in place for third party requests. Users of the databases will get role-based access to the system after they have logged-in using their own username and password. The systems will log all data entry steps with timestamps, update reasons and user information. The role-based access to the systems will avoid unauthorised data access and prevent users performing unauthorised actions. De-identified data obtained from the analysis of biological samples (e.g. transcriptomic data) will be stored on University of Oxford servers, which are double backed-up and undergo strict quality control, as well as servers at Imperial College London, UMC Utrecht and industry partners. Articles will be published as open access articles where possible.

Data will be interoperable between researchers of the RESCEU consortium. Research Online and OpenClinica do not require additional software. Data from these databases can be obtained in many formats (e.g. SPSS, SAS, Excel). Data will be uniformly collected from the different sites to make sure data matches when combined between sites and WPs (i.e. WP4 and WP5). Data obtained from the analysis of biological samples will come in a wide variety of formats depending on the sample being tested and the test being undertaken. In some cases analyses will be undertaken using standard techniques while others may require novel techniques. Vocabulary used in the data will be presented as clearly as possible for a general researcher.

Published articles will contain RESCEU (consortium) in their Title/abstract and will elaborate on the study design of the RESCEU consortium cohort studies. RESCEU will be noted as a keyword. Data will be identifiable and locatable by means of a standard identification mechanism. Each subject will get a unique identification code which cannot be directly traced back to the subject. New versions of protocols will be provided with new version numbers. Biological samples will be labelled with a study specific code as well as a unique identification code to ensure accurate tracking of samples between sites and into Biobanks.

All collected data will be useful to anyone with an interest in the incidence, disease severity and basic science of RSV infections in children and elderly adults. This includes, but is not limited to clinical doctors, epidemiologists, vaccine manufacturers, basic scientists, policy makers, etc.

2. Allocation of resources

During the course of the RESCEU project, data produced by each WP will be managed by the respective WP. The costs for data management, including for making data FAIR, will be paid for by the budget of the respective WP. An “RSV Observatory” is envisioned for the long-term preservation of data for the project. This knowledge portal will be designed to allow various stakeholders, including policy makers, clinicians, researchers, and the public, to interact with the RESCEU data and to potentially include additional relevant datasets that may be of use to the larger RSV research community. Additional funding will be sought to design and implement the Observatory.

3. Data security

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

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 will be retained for 30 days and a monthly archive copy will be kept for a year. The Edinburgh DataShare is compliant with the guidelines of Research Councils, UK (RCUK).

WP2: Consolidation of health care systems data

Each partner country in WP2 will have to adhere to the 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 will be safely 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.

WP3: Retrospective resource use analyses from existing databases/networks

WP3 researchers will adhere to the guidance of original data owners. At a minimum data will be used in a protected environment only (behind a secure firewall), and regular back-ups will be performed to enable data recovery. Databases and web servers will be hosted in data centers that meet the highest possible security requirements.

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

Data will be stored in OpenClinica and Research Online, GCP compliant electronic data capture (EDC) systems, to guarantee correct, complete and consistent data collection. Data from the EDC systems will be transferred over the internet using secured data communication protocols. Databases and web servers will be hosted in data centres that meet the highest possible security requirements. De-identified data obtained from the analysis of biological samples (e.g. transcriptomic data collected by WP5) will be stored on University of Oxford servers, which are double backed-up and undergo strict quality control, as well as servers at Imperial College London, UMC Utrecht and industry partners.

4. Ethical aspects

Ethical issues regarding data sharing will be addressed in the RESCEU Ethics policy handbook (D6.1 in the DoA). All data produced by RESCEU will be shared in accordance with relevant international and national regulations to protect patient privacy and confidentiality. Subjects involved in the clinical studies in WP4 will be consented for data sharing and preservation of samples and data. Informed consent forms (D4.1) will be submitted to IMI.