SYNCHROS

WORK-PACKAGE 6
Strategic Agenda for Better Coordination of Cohorts Globally

D 6.3

SYNergies for Cohorts in Health: integrating the Role of all Stakeholders

Authors:
Anastassja Sjalm, SPF, Nottwil, Switzerland
(www.paraplegies.ch)
Andreas J. Weser, NTNU, Trondheim, Norway
(www.ntnu.edu/hunt)

Contributors:
Ellen Vorstenbosch, PSSJD, Barcelona, Spain
(www.pssjd.org)

Reviewers:
Members of the SYNCHROS Advisory Board
(www.synchros.eu/our-advisors)

Grant Agreement No. 825884
Start Date: 01/01/2019
Duration: 36 months

DOCUMENT INFORMATION

Authors	**Anastassja Sialm (SPF)** **Andreas Weser (NTNU)**
Contributors	Ellen Vorstenbosch (PSSJD)
Reviewer	Members of SYNCHROS External Advisory Board
Responsible Partner	SPF
Dissemination Level	Public
Nature	Report
Keywords	Strategic agenda; Cohort optimalisation; Strategic Tasks, Resource Based View; Knowledge Based View
Due Date of Deliverable	31/05/2022
Actual Submission Date	31/05/2022
Version	1.0

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

TABLE OF CONTENTS

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

1. INTRODUCTION/ SCOPING OF THE TOPIC AND THE REPORT

Personalised medicine, fighting epidemics, optimising healthcare systems, development of treatments to fight cancer or rare diseases are just few examples of a broad range of topics where answering research questions requires multi-cohort research in order to reach the statistically relevant number of cases in the specific categories of participants in studies or surveys. This is the starting point and motivation for SYNCHROS [1], a European H2020 project with the goal of providing support for the harmonisation of data across cohorts, to identify the practical, methodological, ethical and legal challenges, to compile and suggest solutions and to develop a strategic agenda to tackle these challenges. Institutions and experts representing epidemiological and clinical research, patients, legal/ethical competencies and European infrastructures initiated the project. SYNCHROS demonstrated once more that research based on data collected in different cohorts by different researchers, institutes and under different organisational, legal and ethical governances is still difficult and partly impossible; even for cohorts listed in repositories like the SYNCHROS-repository [2], Maelstom [3] and others. The reasons for this state of affairs were systematically explored and the findings are summarised in SYNCHROS-Strategy-Briefs, [3] [4].
The core issues concern ethical and legal obstacles, standardisation of variables and metadata and the sustainability of infrastructures. Although we see promising developments - for instance the initiative "GAIA-X Domain Health"[1], which aims at a federated, open data infrastructure based on European values – the current state of art (2022) is still characterised by a significant heterogeneity in the governance structures of cohort data stored in silos, i.e. data collections held by one organisation that is not easily accessible by researchers from other organisations, particularly, if their institution is situated in another country. And, even if access is finally granted, these data collections are rarely prepared for multi-study-cross-cohort analyses, i.e., harmonisation of variables, meta data and data remains challenging and partially impossible.

Based on SYNCHROS' findings, expert observations and stakeholder's feedback the focus of this report is on HOW the coordination of cohorts could be further improved in order to make multi-study integrative research across different cohorts easier. In this report we develop a vision, suggest initial strategic tasks and measures, and identify the stakeholders who are expected to be in a position to facilitate the implementation of these activities. The strategic tasks may be tackled independently depending on resources and health-political conditions, but it is important that measures are being initiated and implemented complementarily on the basis of an agreed vision as a roadmap. All suggestions are embedded into a resource- and knowledge focused strategic managerial model, which functions as a sustainable approach for further developing, complementing and amending our initial strategic agenda.

[1] https://www.bmwk.de/Redaktion/DE/Publikationen/Digitale-Welt/211116-pp-health.pdf?__blob=publicationFile&v=6

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

2. THEORETICAL BACKGROUND AND CONCEPT

Within the SYNCHROS-project various waves of stakeholder consultations were carried out and the aforementioned issues related to multi-cohort integrative research were discussed and analysed systematically. Stakeholders of the project were representatives of the following communities:

- Researchers / PIs and research projects that use data from different cohorts,
- Experts in harmonisation methodologies, ethical and legal domains[2], and representatives of infrastructures,
- Funders, e.g., National Ministries of Health, the European Commission or private funders like the Welcome Trust.

Detailed information about the affiliations, participation, recruitment and representativeness of the stakeholders can be found in Annex 6.2. The theoretical background used for both the analysis of stakeholder's feedback as well as the development of a strategic agenda relies on the instruments "Resource Based View" (RBV) [5] and "Knowledge Based View" (KBV) [6]. The rationale of these two approaches is to uncover the resources and capabilities needed to ensure the sustainability of data infrastructures in methodological, ethical and legal terms. For further details on RBV and KBV refer to the glossary at the end of this report, section **¡Error! No se encuentra el origen de la referencia.**.

Based on issues known, in a first step, we start with outlining a hypothetical and ideal organisational model, the resources, the knowledge, the capabilities and the services that would be desirable to get closer to an optimal coordination of cohorts globally.

In a second step, we match this model with the findings from on-going and past projects and initiatives in order to identify existing building blocks to the sustainability of data infrastructure in cohort studies in methodological, ethical and legal terms.

(I) Resource Based View (RBV)

Initially, the resource-based view (RBV) derives from the strategic management field. RBV focuses on the internal resources of an institution in order to identify those capabilities and competencies that are likely to generate competitive advantage. In the context of cohort research in general and SYNCHROS in particular, RBV refers to the identification of resources and capabilities of research institutions that are the most likely to facilitate cohort research processes. RBV provides a theoretical lens that will be applied to the outcomes of stakeholder dialogues

[RVB Core Concepts can be found in Annex I]

(II) Knowledge Based View (KBW)

In strategic management, the Knowledge Based view (KBW) explores how employees are increasingly involved in the formulation and administration of the operational goals of their firms. For cohort research in general and SNYCHROS in particular, the focus is on how research institutions (and their members) integrate knowledge and capabilities in order to conduct cross-cohort research successfully. Capabilities refer to the recurrent patterns in creating, transferring, or otherwise "managing" knowledge. KBV provides a theoretical lens that will be applied to the outcomes of stakeholder's dialogues and to the results of the strategy briefs.

[KBW Core Concepts can be found in Annex II]

[2] Data protection (GDPR), members of ethical committees responsible for large cohort data collections

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

And - in a third step – we derive a strategic agenda, i.e. the set of strategic tasks thought necessary to complement missing "building blocks", structures and functions, embedded in an overall model for a better coordination of cohorts in a real world. In doing so we provide strategic steps and tools for ensuring the sustainability of data infrastructure and interoperability within the methodological, ethical and legal domains in cohort research.

2.1 Applying KBV and RBV lenses to the coordination of cohorts

According to the Resource Based View (RBV) a firm is a bundle of productive heterogenous resources that generates products and leverages knowledge related to these products [5]. Each firm possesses different bundles of resources that allow to achieve a strategic advantage in terms of innovation, production and supply chain in specific market environments. In the context of research institutions involved in the coordination of cohorts, this means that **organisations, stakeholders and the researchers' and health data communities need particular patterns of resources in order to manage cohorts effectively**.

Resources can be classified according to two dimensions namely their degree of operationalization and their scope of application [5] [6]. First, resources are **either tangible or intangible**. Tangible resources have physical attributes and can be both observed and quantified. In cohort resources, examples of such resources may be existing infrastructure, database architecture, funding and datasets. By contrast, intangible resources are not visible and cannot be quantified or operationalized. In cohort research, examples of such resources may be reputation, academic relationships, informal communication channels, trust and research culture. Both types of resources are crucial for coordinating cohorts (e.g. common datasets cannot be curated without trust between data controllers and data providers). The difference is that tangible resources are easier to transmit across institutions while intangible resources are easier to share within specific institutions

Second, resources are either **specialized or versatile**. Specialized resources apply to a narrow range of contexts while versatile resources operate in broad contexts. Specialized resources are thus used to solve local issues (e.g., standards from national datasets) while versatile resources are related to international ones (e.g. measurement standards across EU, trans-European legislation such as the GDPR). In uncertain settings, specialized resources can be more effective than versatile ones because they are more robust [5]. **Resources position barriers** refer to situations where individuals cannot have access to the resource while **resource immobility** means that a specific resource cannot be used, shared and applied effectively.

Thus, research actors and research institutions need specific resources patterns of different types for coordinating cohorts globally. The question is **how resource types should be distributed and which knowledge capabilities can be used for leveraging these resources**. According to KBV, knowledge capabilities refer to competences and operational routines that allow firms to compete in their business environment [6]. In the present context, knowledge capabilities **represent knowledge operating capacities for leveraging the resources needed for coordinating cohorts**. **Integrative** knowledge capabilities are needed when cohort research activities require the coordinated efforts of individual specialists who possess many different types of knowledge (e.g., data controllers and data contributors working in parallel). As such, they include aggregation, transferability and appropriability capabilities (cf. Annex 6.1). **Combinative capabilities** associate existing methods, practices, data and infrastructure in different ways. Development in the coordination of cohorts is then defined by carrying out new combinations of "old" capabilities (e.g. using existing datasets for new research purposes). Finally, **knowledge integration mechanisms** ensure that knowledge on the coordination of cohorts can be shared,

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

applied and implemented across domains and actors (e.g. data sharing between European countries between different institutions with different data infrastructure) . Depending on their types (i.e., rules and directives, routines, sequencing, problem solving and decision-making cf. Annex 6.1) they may require intensive temporal and coordination resources.

In the present report, we will classify domains of actions derived from the Stakeholders' Dialogues according to RBV and KBV concepts outlined above. In practice, this means that we will evaluate the representation of resources types and the number of capabilities needed for such resources.

3. MAPPING OF STAKEHOLDER POSITIONS INTO A STRATEGIC VISION

3.1 Central issues identified by stakeholders - Strategic aims and vision

In the SYNCHROS project, a lot of care and effort was put into a representative selection of stakeholders and the implementation of effective dialogues with them. **¡Error! No se encuentra el origen de la referencia.** and **¡Error! No se encuentra el origen de la referencia.** illustrate this process and the iterative way of covering the stakeholder's feedback in the project results.

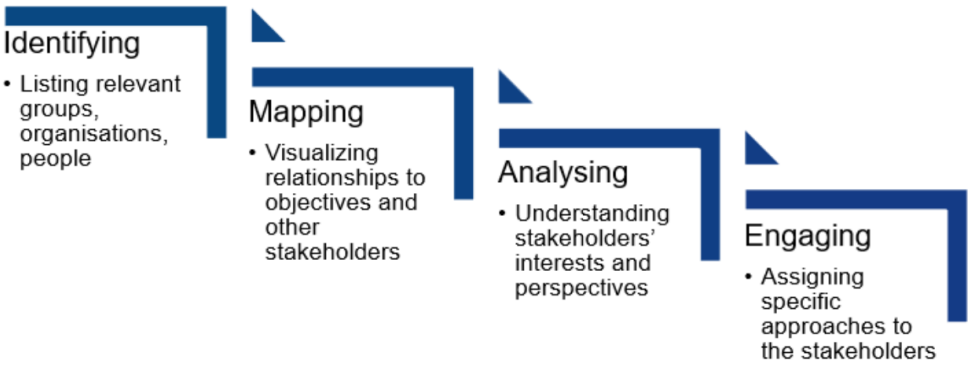

Figure 1 – Four steps of the stakeholder involvement process

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

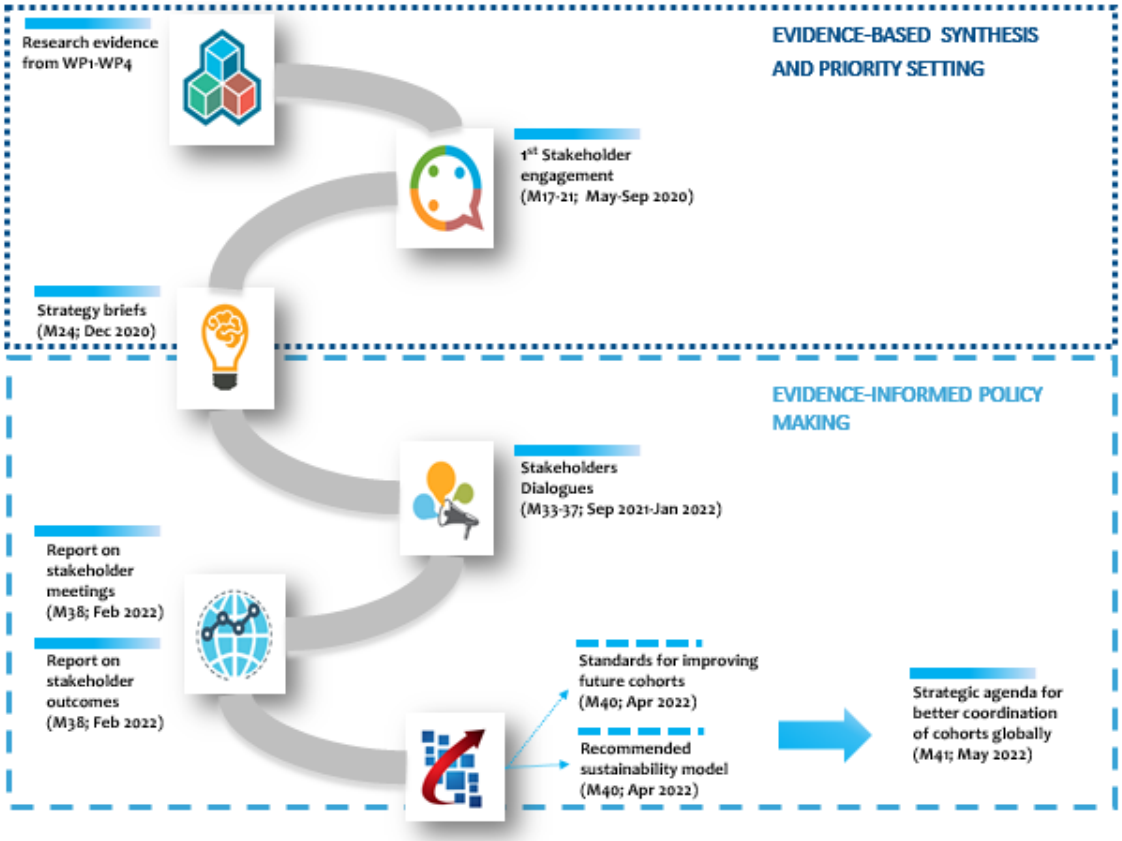

Figure 2 - Embedding of stakeholder's feedback into the project iterative production of project reports and results

The following paragraphs mainly cover and outline issues brought forward by stakeholders. Not surprisingly, the stakeholder consultations disclosed lots of issues and fields for improvements in the coordination of cohorts globally, the details of which are explained in the following sections.

The points made by stakeholders can be clustered into the following three fields, which we will call pillars of action:

I. Standards and de facto standards for variables, metadata and metrics of data collected,
II. Guidelines for harmonising legal, ethical and organisational standards and rules for data access, and
III. An efficient way of providing knowledge and support regarding all issues related to cross-cohort research.

We ordered the specific issues identified by stakeholders along these three pillars, ¡Error! No se encuentra el origen de la referencia..

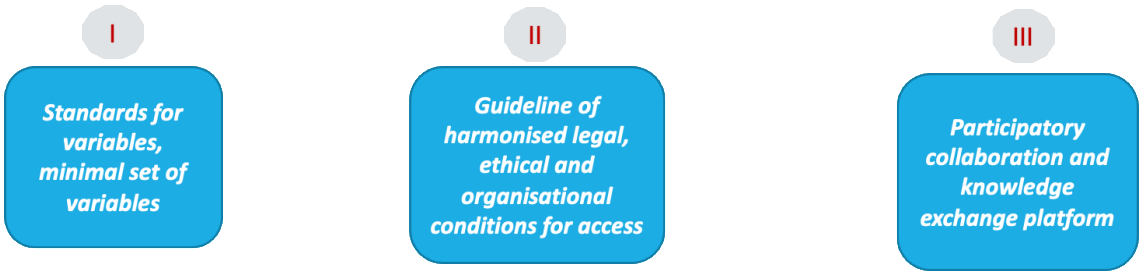

Figure 3 - Stakeholder dialogue inputs ordered by three pillars

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Based on the analysis of the points expressed by stakeholders and the multidisciplinary expertise represented in the SYNCHROS consortium, we created an expandable model of a hypothetical, ideal functional environment, which would match all or at least most of the expectations of stakeholders. This model shall serve as a "template" that allows to reflect the current status on the visionary target status.

In such an "ideal world" where the major issues of integrative multi study research were solved and interoperability of data infrastructures was given, researchers can - with reasonable effort:

- Identify the cohorts that has the data / variables needed
- Find the information on their definitions, metadata, data metrics and stratification, the number of participants and corresponding metadata
- Variables, metadata, metrics are widely standardised or de-facto standardised and/or information on how to convert data into these standards is provided
- Find information on the study designs and methods of data collection
- Get advice and support from the cohort management
- Find information on how to get access to the data, and
- Request and get the access based on comparable and mutually accepted organisational, legal and ethical rules.

In such a world, variables were defined following international standards or broadly accepted practices, and the data can be harmonised for the subsequent statistical analyses using proven, state of the art harmonisation methods - with or without the help of specialists. Furthermore, institutions collecting the data are trusted parties in their societies, citizens participating in studies can rely on an ethically and legally correct use of the data for purposes, they are aware of and agree with.

Naturally, this theory is an illusion; simply, because in many longitudinal studies data are collected over decades, and variables and their definitions change over time, as well as the methods of data collection and the accuracy of data collected. And, even for data collected today some variables may have different meaning and interpretations in different contexts - just think of variables to cover the often-needed socio-economic-status of a participant, usually described by annual income and education. This would be completely different for countries in Africa, parts of Asia or Europe/America. Harmonising data with such heterogeneity requires special approaches, but one does not need to look at such extreme examples to understand the difficulties of multi-cohort studies. Even across EU countries, cohort research still suffers from governance heterogeneity. Particularly for variables that are not measured or measurable with biophysical instruments and methods, we still see too few reliable standards that could ease multi-study integrative research.

Nevertheless, the above vision of an ideal environment for multi-study integrative research allows assessing the status quo, identifying gaps and designing the structure and processes of resources and knowledge necessary to get closer to an "ideal world" in methodological, legal and ethical terms.

In the following structured sections, we describe in more details the positions expressed by stakeholders.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

3.2. Pillar I: Standards for variables, minimal set of variables

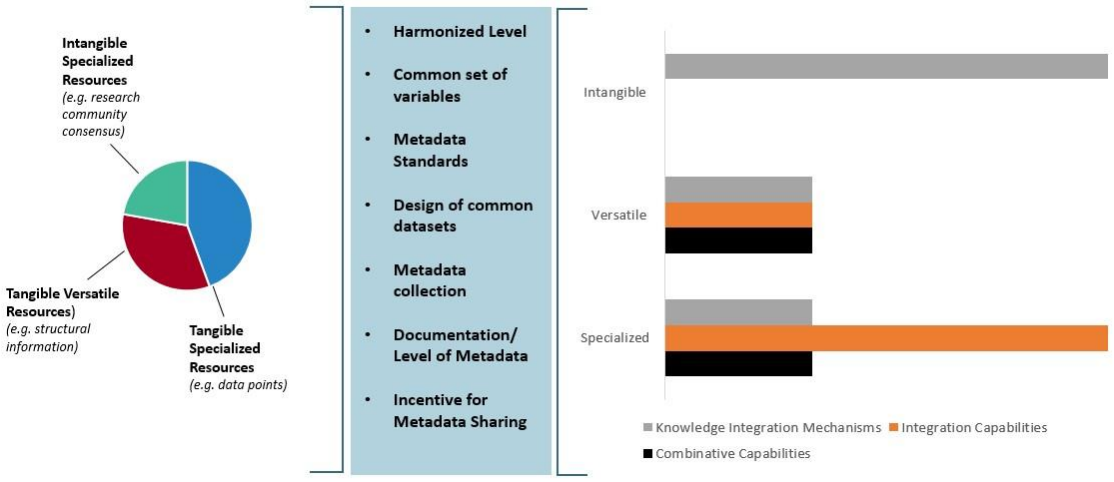

Figure 4 - Pillar I: The components required for Pillar I are outlined in the central column. The representativeness of resources types is shown in the chart on the far-left side (blue for tangible specialized resources, red for versatile tangible resources and green for intangible resources). The amount of capabilities needed for each resource type in Pillar I is represented in the bar chart at the far-right side (grey for knowledge integration mechanisms, orange for integration capabilities and black for combinative capabilities)

1. Harmonisation Levels

In order to determine the level of harmonisation needed, it is necessary to first determine the kinds of resources concerned. On the one hand, the harmonisation level depends on the specific purposes for which the research is being carried out. On the other hand, it is crucial to understand what is harmonised in the first place. That is, harmonising data is far more difficult than harmonising metadata. This is because **data harmonisation does not directly concern data per se; hence, it is impossible to fix the data items themselves**. What is actually harmonised are **data metrics and organisation, where suffixes and prefixes link data points to definitions**. Data points are thus needed to determine the harmonisation level, but data points are highly specialized resources, which limits their application. That is, each study has its own data points with specific data items. This may lead to **resource immobility: the organisation of data and data points cannot be transferred to other cohort research projects**. There are no established standards for data harmonisation levels, especially when innovative questions are concerned.

Thus, the question is, which knowledge capability can best confront such specialized and potentially immobile resource so that the harmonisation level can still be determined. We suggest that integrative capabilities of knowledge aggregation and knowledge transferability are needed. They won't fundamentally change data points, but they will allow to **categorize, structure and define data points without changing them** (because they do not include knowledge transformation steps).

2. Common and minimum set of variables

What are the resources and capabilities needed to determine common and minimum sets of variables? A common set of variables is only possible with the active **engagement and agreement of the researcher community**. This means that the resources involved are both **tangible** (i.e., consensus about a specific and quantifiable number of common variables to be retained) and

intangible (i.e., the active engagement of the research community that is based on values such as the attainment of the public good). Surprisingly, it is the tangible, quantifiable resource that is the most difficult to exploit. That is, there are significant problems in reaching a consensus about a minimum set of variables. The number of domain-specific elements remains low because each database has a very strong focus on those data elements that are not redundant.

What is needed, is thus a set of capabilities that could best leverage specialized types of resources. In the present context, combinative (rather than integrative) capabilities may be suitable because they allow **combining existing resources in different ways**. In practice, this means that the **scope of consensus finding should be limited to few variables**. Medical societies may help to define such as consensus scope, but the task remains difficult because the specificity of research questions prevents overlap between datasets.

3. Metadata Standards

The first step is to identify tangible resources that can foster the establishment of standards for variables and common metadata in particular. **There are already existing, established tangible resources for metadata standards**. Existing and recommended standards and modules such as the DDI [7] and the EU-SILC (for surveys) [8] are available. Catalogues for the collection of recommended standards are already generated at the WHO level (e.g., GATHER guidelines for global health estimates) [9].

The difficulty, however, resides in maximizing and leveraging such existing resources in an appropriate way. That is, **these types of resources have to be versatile** in order to have value. In the present state however, existing standards present significant **barriers to entry** for researchers, which prevents their identification and use. Namely, there is **not enough empirical data about the standards in use** and it is thus difficult to determine what standards researchers are using in practice. Moreover, researchers have difficulties in finding these standards, as their use is both wide and fragmented.

It is thus crucial to identify the kinds of capabilities needed to maximize the use of existing metadata standards and alleviate entry barriers. Combinative capabilities (that use combinations of existing knowledge to obtain results and actions that can be applied across contexts) play an important role. An **effective combinative capability is to generate a catalogue for collecting standards inputs**. This would enhance standardization and harmonization since researchers would no longer have to generate new modules from scratch and would be able to use already existing modules instead.

This combinative capability can generally be associated with integrative capabilities such as the **aggregation of standards** (for better transfer) and knowledge integration mechanisms through **rules and directives**. Namely, generating a catalogue for the collection of standards inputs will necessarily create **procedures and standardized information that will regulate the application** of **knowledge** and collaboration in cohort research practices. **Hence, a standard collecting catalogue should specify the scope of the standardization to be done and the type of research community concerned.**

4. Minimum and common datasets

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Metadata standards may be especially helpful for the design of minimum and common datasets. However, the resources required for common datasets differ from those needed for metadata standards. Namely, at the present stage, common and minimal datasets are not versatile but rather specialized resources, which means that they can be applied only to **a limited range of contexts**. That is, common and minimal datasets are generally very small and are approved by the research institutions conducting the research. Minimal datasets are also **dependent on the particular field of the research.** For instance, epidemiologists are likely to require different variables (and thus different fixed data items) than chemical specialists. Common datasets are possible only for certain kinds of data, for instance adverse events.

It is thus crucial that specialized resources should be coordinated effectively, so that robust and effective common data sets could be generated. For this endeavour, we need both integrative capabilities (where coordinated efforts of individual specialists with different types of knowledge are needed) and knowledge integration mechanisms. A key integrative capability here is the one of knowledge aggregation (i.e., the efficiency of knowledge transfer depends on its aggregation). This kind of capability allows structuring the data despite inter-study differences so that the data will be organised in a similar manner and therefore being more comparable. For each adverse event, it is possible to choose a common data element that indicates the start date, the end date and the on-going status of the event. **This common data element becomes the main code for structuring all adverse event data**. The **integrative capability in this context, is thus to obtain fixed data items** and apply them to structure the data for further comparison.

However, such activities can be done only by the research community and thus require **knowledge integration mechanisms**. This is because a single researcher cannot fully know what a minimal dataset should be. In order to identify which knowledge integration mechanism is needed, it is necessary to identify the emergence context of the resources for common datasets. Minimal datasets do not emerge out of anywhere: it **is only possible to generate them once there is sufficient metadata** about what other researchers are using (in terms of common data elements).

Thus, the knowledge integration mechanism should be relatively **simple** and operate through existing procedures and standardized information (i.e. rules and directives) that will regulate the generation of minimum datasets. The appropriate knowledge integration mechanism is thus to see **how others use metadata and generate common codes for data structure** and make an informed selection of them for a minimal dataset. However, it is also possible on some occasions to start from a common minimum dataset. The JRC (Joint Research Center for rare disease in EU) [10] defined common data elements that are now referenced by all the 24 European reference networks for data comparison.

5. Metadata collection

An important issue in metadata collection is that **academic health data is rarely recorded with detailed structural information** (unless they are transferred to the FDA or EMA). This is a significant problem because the level of structural information is crucial as it allows data to be structured and compared in consistent ways. That is, in order to do a consistent comparison, it is necessary to know what each of the researchers involved means by certain terms (e.g., “dizziness”).

Structural information is related to tangible, versatile types of resources and their effective leverage and use can be achieved through integrative capabilities and integrative knowledge mechanisms. That is, knowledge integration mechanisms are needed when cohort research activities require the coordinated efforts of individual specialists who possess different types of

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union’s Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

knowledge. For metadata collection, a lot depends on the **use preferences from statisticians**. While the IT staff and the data managers are often willing to use the proposed standards for data (recording), **statisticians are more conservative and tend to use standards they are familiar with**.

Hence, we need to apply a **knowledge integration mechanism that will coordinate the collaboration of specialists of different domains, e.g., IT, eHealth, statistics, clinical research**. In general terms, knowledge integration mechanisms should be used to **train and prepare** the staff with different backgrounds so that the acceptance for standards could be increased. In more specific terms, training should include group problem solving and decision making with isolated **knowledge transformation steps** (where standards are increasingly accepted) and personal communication-intensive forms of knowledge integration.

Such a process cannot be implemented in a top-down manner, however. Instead, the acceptance for standards should come from the bottom up or more precisely, **from the research community**. That is, if metadata collection and websites are commonly used, it is because researchers, IT-people and statisticians are **committed to them through practice**.

6. Documentation/Level of Metadata

A central domain for the first pillar (and thus for the sustainability and interoperability of variables standards) concerns the level at which metadata should be documented. The question is not trivial because **the level of documentation depends on the nature and use of harmonised data**. While it is expected that there should be a basic minimum metadata documentation for other cohort studies to follow, the reality is that even such minimum documentation is determined by funders' needs and expectations. The question is **which kind of details the funders should specify when they put out the request for proposals**.

We should first identify the resources needed for determining the level of documentation for metadata. The resources concerned are **descriptive metadata and contextual metadata** (cf. the discussion above about standard metadata content). In both cases, the resources are tangible (i.e., they can be quantified) and versatile (they concern most types of cohort data). **However, descriptive metadata does not present barriers to entry while contextual metadata do.** In the current context of cohort research, contextual metadata include entry barriers and display resource immobility because they are not documented in a consistent way. There are **no real existing schemes for the documentation of contextual metadata** for cohort research and hence, researchers have difficulties in fully leveraging contextual metadata resources. This is a strong contrast to the situation for clinical trial studies where researchers can rely on an established consistent system for documentation (CDISC[3] define-XML or OMOP[4] system for observational data) [11] [12].

Descriptive and contextual metadata are co-dependent. That is, while descriptive metadata can provide a detailed description, it is generally advised to describe how this data has been collected in the first place. The method for data collection includes objective measurements and contextual metadata that require even more detailed descriptions (e.g., the kinds of devices and calibration used, if they were machine-generated etc.). Thus, we **need specific knowledge capabilities to (i) document the data collection process appropriately and (ii) remove barriers to entry and resource immobility arising from contextual metadata**.

[3] Clinical Data Interchange Standards Consortium, https://www.cdisc.org/standards
[4] OMOP Common Data Model, https://www.ohdsi.org/data-standardization/the-common-data-model/

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

In order to document the data collection process adequately, we need **combinative capabilities that will link the objective measurements with contextual metadata** in a seamless way. For this endeavour, we need to rely on relatively **simple, already existing tangible resources** (too much complexity may result in too much detail). Namely, there are already standard (descriptive) metadata descriptions available such as the DDI (e.g., Data Documentation Initiative) [7]. The combinative capability in this case, is thus to **list all the methods for standard metadata descriptions used currently in different catalogues** (e.g., in the BBMRI, the EUPHA) [13] [14]. These examples would allow researchers to determine an adequate standard of metadata descriptions and can be used by the (EU) commission in their strategy for standards definition.

7. Incentives for metadata sharing

Incentives are needed for sharing and documenting metadata and minimum datasets. The resources needed are both tangible and intangible because they are not fixed and pre-established (unlike resources for metadata standards). Tangible resources concern the costs and funds needed to set a minimum dataset. Currently, **funds** are specialized resources (i.e., they are associated to specific projects) that suffer from **resource immobility** (i.e., it is a resource that is quite inelastic in its application). This is because funders do not allocate resources adequately while researchers tend to neglect metadata documentation altogether.

Thus, using **funding as an incentive for sharing and publishing metadata description may be dangerous** mainly because there is a potential for the emergence of resource position barriers for researchers. That is, funding based on metadata would exclude all research projects that cannot provide a whole set of metadata. The solution in this context is that research **infrastructures should step up their efforts to provide services for researchers** so that they could create consistent metadata. While funding remains an issue, such a measure will ensure that expertise is concentrated in one place and that researchers can rely on a central service for assistance. Funding will thus no longer suffer from resource immobility and can be leveraged adequately.

However, research infrastructures need a range of integrative capabilities in order to provide appropriate services for creating metadata and to **actively support and credit researchers**. First, transferability and aggregation capabilities are needed so that explicit knowledge about metadata can be communicated to researchers (and vice versa). In practical terms, this means that **research infrastructure should give researchers references for the data description they provide**. Second, research infrastructures need an **appropriability capability** that will allow researchers to receive a return equal to the value of the knowledge they bring in (in this context the metadata description). In practical terms, this means that research infrastructures should support researchers in **uploading their data descriptions** in a way that will ensure that they will be consistently credited for their work.

It should be noted however, that, while research infrastructures in social sciences libraries of education research (e.g., ERIC [15]) already use such integrative capabilities, there is still no similar strategy in the health domain/health cohorts.

Incentives for publishing and sharing metadata can also rely on intangibles resources. Unlike tangible resources, **intangible resources** cannot be quantified, at least directly. The intangible resources for metadata publication/documentation incentives concern the **consensus of the researchers' community**. That is, there is the possibility to convene a research community so that

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

a common researcher-based consensus can be reached. However, there should be clarity about the type of standards that researchers want to reach a consensus on.

In order to achieve this, we need knowledge integration mechanisms centred around **rules and directives that will structure standards information and communication systems for metadata publication and implementation**. For instance, this consensus from the researcher community can **be reinforced by transnational and international agencies** (such as the World Health Organisation or the European Commission). Such agencies have the mandate and the power to implement this researcher community consensus as they have clear procedures in place. Moreover, if there are established directives for journals to publish the harmonisation process in the detail, this may incentivize researchers to do the same.

Section Summary:
The components of Pillar I and its associated resources and capabilities are listed in Table 1.

Since Pillar I concerns standards for variables, it requires **mainly tangible resources** that can be observed and measured. This is because standards need to be concretely implemented with clear rules, directives and application domains in order to have any relevance. A standard can be characterised as such only to the extent to which it is implemented and used by researchers and communities of practice. Resources for standards need to be both **specialized** (i.e. so that they could be applied to local and national contexts) and **versatile** (i.e. so that they could be extended to wider international contexts). However, the **relationship between local specialized resources and global versatile resources lacks specification**. For instance, minimal data sets are dependent on specialized resources and domains (e.g., specific research fields differ in terms of the variables they require). As a result, a conversion to a versatile resource such as common data sets remains difficult.

In this context, Pillar I **converts** specialized resources into versatile ones through integrative capabilities (e.g., aggregation). In doing so, it ensures that **both types of resource retain equal weight**. Integrative capabilities are mostly applied to specialized resources because standards from different contexts need to be aggregated together into a coherent whole in order to be **transferred** to international contexts where versatile resources are required. In general terms, **standards are achieved by maintaining a balance between specialized and versatile resources.**

Example of the conversion of specialized tangible resources into versatile specialized resources:

*Data from multiple cohort studies (specialized tangible resources) is subjected to **ex-post retrospective harmonization** so that it can be compared and accessed (i.e. data becomes versatile because it can be used across contexts). Ex post retrospective harmonisation combines data from cohort studies that were not specifically designed to be comparable, but, even though no standard formats or protocols were used, variables can be assessed and formatted through an agreed semantic strategy in order to achieve commonality through data processing procedures.*

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Table 1 – SYNCHROS Strategic agenda – Pillar I

Component	Resource and Resource Type	Capability
Harmonisation Level	Data points: Specialized tangible resource, potential for resource immobility	Integrative capabilities: Aggregation and transferability
Common and minimum set of variables	Consensus about the number of common variables to be retained: Tangible specialized resources: Active research community engagement: Intangible specialized resources	Combinative capabilities: Limiting the consensus scope to few variables
Metadata standards	Existing resources for metadata standards (e.g. DDI and the EU-SILC): Tangible and versatile resources, contain barriers to entry.	Combinative capabilities: -Generating a catalogue for collecting standards inputs -Aggregation capabilities (integration) and Rules & Directives (knowledge integration mechanisms): -A standards collecting catalogue with specifications about the standardization scope and research community type.
Design of common and minimum datasets	Common datasets: Specialized tangible resources -Emergence Context of resource (i.e. common datasets): Occur only once there is sufficient metadata about the common data elements other researchers use.	Aggregation capability (integration): -Generating fixed data items for data structure and comparison Rules and directives (knowledge integration mechanisms): -(i) Evaluating how others researchers use metadata (ii) Generating common codes for data structure

		and (iii) Making an informed selection of these codes for a minimal dataset
Metadata Collection	Structural information : Tangible, versatile ressource	Group problem solving and decision making (knowledge integration mechanisms): - Increasing the acceptance of standards though training and preparation of staff with different backgrounds.
Documentation/Level of Metadata	Descriptive metadata and Contextual metadata: Tangible and versatile resources, contextual metadata presents barriers to entry and includes resource immobility	Combinative capabilities: -Listing all the methods for standard metadata descriptions used currently in different catalogues
Incentives for metadata sharing	1) Costs and funds: Tangible resources, present resource immobility issues. 2) Research community consensus: Intangible resources	1) For Costs and Funds: -Aggregation and transferability capabilities (integration): Research infrastructure giving references for provided data descriptions to researchers -Appropriability capabilities (integration): Research infrastructures support and credit researchers in uploading their data description 2) Research community consensus: -Rules and directives (knowledge integration mechanisms): -Consensus is implemented through standard information and communication systems from the

		researcher community and enforced through by transnational and international agencies.

3.2. Pillar II: Guideline of harmonised legal, ethical and organisational conditions for access

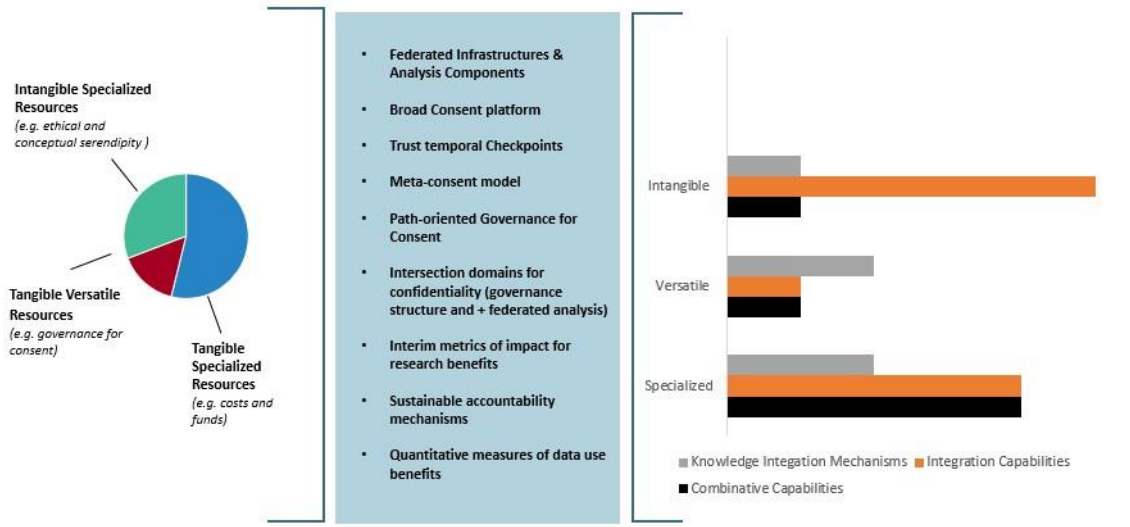

Figure 5 - Pillar II: The components required for Pillar II are outlined in the central column. The representativeness of resources types is shown in the chart on the far-left side (blue for tangible specialized resources, red for versatile tangible resources and green for intangible resources). The amount of capabilities needed for each resource type in Pillar I is represented in the bar chart at the far-right side (grey for knowledge integration mechanisms, orange for integration capabilities and black for combinative capabilities)

1) Federated Infrastructure and Analysis Components

Pillar II should contain the capacity for federated infrastructure and federated analysis, as they are instrumental in ensuring the interoperability of data infrastructures. In order to implement a federated approach strategically however, we need to evaluate the kinds of resources required.

Costs and funds resources for federated structures are unstable because they **present potential resource position barriers**. That is, it is not always possible to determine who will assume the costs, pay for the software or recruit managers for using the software on the national nodes. Even if the cost contributors are identified, one should prevent that this does not block other potential collaborators to participate in the federated structures. In other words, established cost contributors should not have exclusive "first mover" advantage. Unless such issues are solved, it is difficult to develop infrastructures that will assure data quality. We thus need to use **combinative capabilities** in order to divest professional and data quality management to **individual resources at the research institutions level**. Costs and funds resources will thus become more specialized to the individual characteristics of the projects.

However, costs and resource funds depend on **what federated analysis and decentralized structures can or cannot do**. That is, some elements are practically not feasible in federated analysis (there are for instance, interoperability and compatibility issues in relation to data queries). We thus need to **use transferability and aggregation capabilities** to determine the boundaries of usefulness for federated analysis. Only then will we be able to know if the investment required is worth it.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

If we consider **federated analysis as a resource** for Pillar II, then it should be noted that it is relatively easy to handle and thus **does not present resource immobility problems**. Despite its scalability problem, it is a versatile resource that is not particularly costly, unlike centralized solutions such as data lakes. However, federated and centralized analytical arrangements present significant **problems for anonymization conversion**. As a **resource, data** in cohort research has **limits to its versatility:** data cannot be converted in any formats and anonymized indefinitely. For this reason, data in cohort research is **pseudonymized rather than fully anonymized**.

However, any **aggregation and transferability capabilities** needed for the conversion of data into a pseudonymized format should take into account that the legal basis for such conversions is still unclear. Moreover, because the versatility of data is limited, any manipulation of it for **anonymization** purposes may degrade its scientific value. First, full anonymization techniques use algorithms that make data unusable. Second, the value of cohort data is often dependent on the amount of location and temporal information (such as dates). Suppressing such kind of information in a longitudinal context is impossible unless one is willing to sacrifice the scientific value and impact of the cohort study.

Hence, if one intends to pseudonymize data in a federated, centralized and cohort study context, it is necessary to ensure that **the aggregations and transferability capabilities used will not degrade data as a resource**. Hence, we need to bind aggregation and transferability capabilities for pseudonymisation to the longitudinal characteristics of cohort studies (i.e., temporal and location dimensions).

(2) Broad Consent Platform

As a resource, broad consent is a **relatively versatile resource but with intangible aspects**. It is tied to the social value of research (an intangible concept that cannot be quantified but which underpins research practices)) and there are too many uncertainties related to future data re-use (e.g., unexpected discriminatory issues can emerge). Such an uncertainty means that broad consent can present important **resource position barriers** as well as **barriers to entry**. Broad consent presents barriers to entry because it **does not necessarily reflect the preferences of data subjects** and as such, hinders their ability to fully participate in consent arrangements. For instance, study participants may object to specific categories of research for ethical and/or personal reasons.

It is thus crucial to fund those **integrative knowledge mechanisms** that would allow breaking these entry barriers. **Group problem solving and decision-making processes** are particularly relevant in this context. They should be implemented in a **platform** where participants have the possibility to opt-out from studies they consider ethically objectionable. However, participants' control over their data erodes when broad consent (rather than specific consent) is adopted. Thus, a broad consent platform needs to have built-in integrative **appropriability capabilities** so that the control is not diluted but rather, equally distributed to research institutions that participants can trust.

(3) Trust temporal checkpoints

In a participant-researcher relationship, trust is a **highly intangible resource for ensuring consent and data re-use**. This is problematic because intangible resources (such as trust) do not apply to all situations equally. Trust is specific to the initial consent arrangement and to the particular research relationship concerned. While it can evolve over time, trust remains a **rare resource**

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

because of its specialization: trust between the researcher and the participants remains **unique and cannot be replicated** because it is specific to the scientific, legal and social factors of the research relationships concerned. It is thus not clear how such specialized trust can be applied across different contexts of data reuse.

On the surface, trust can be quantified by consent arrangements and confidentiality guarantees. However, the origins of a resource such as trust are marked by **unique historical conditions** (e.g., people are more willing to trust researchers in critical situations such as during a pandemic), **causal ambiguity** (i.e., there is still no clarity regarding the motivations of governments, pharmaceutical firms and the health sector) and **social complexity** (i.e. there is no guarantee that research institutions won't turn rogue later on). Hence**, trust is a ~~rare and~~ not substitutable resource**: once it is lost, there is no more possibility for cohort research.

Such characteristics mean that trust is a **highly specialized** but volatile. A way to mitigate this volatility is to **add quantifiable, measurable dimensions to trust** (i.e. **converting trust into a tangible and quantifiable resource).** This can be done through knowledge integration mechanisms such as **rules and directives**. According to the knowledge-based theory, rules and directives generally regulate the applications of knowledge and the collaboration between those who hold specialized knowledge. In the context of cohort research, this means that we should install **checks and barriers to data misuse**. This can be done through a governance system operating in the long term, with regular control mechanisms over time. Examples of such control mechanisms can be institutionalised arrangements for whistle-blower protection as well as the legal enforcement of individual responsibility and liability for data misuse.

Another solution is to structure the intangibility and volatility of trust through **temporally based routines as well as sequencing processes.** That is, in the control-trust interaction, the issue of control gains more importance over time. For instance, data subjects are likely to trust their institutions (and their ~~alleged~~ respect for human dignity and democratic values) and thus give their consent for the use of their personal data. However, this does not mean that they will trust the same institutions a decade later. Thus, in order to **preserve trust over time, data subjects' control over their data is essential**. The main question, therefore, is how we can ensure that **data subjects have the right tools to fully exert their right to data control**.

This question is not easily solved given the increasing use of emerging digital data collection technologies (i.e. eHealth) in cohort research- Modern digital technologies problematize the issue of control because their structures and modes of operation distribute control away from users and dilute personal agency and awareness This is partly motivated by participants' lack of awareness of their own autonomy: they tend to readily agree to terms and conditions of the platforms they use without understanding the implications for doing so. Autonomy and control is thus not something that can be taken for granted because it can be easily exchanged for perceived benefits (such as the services of an app). In fact, few users have really full control of their privacy rights in digital platform settings. This gradual loss of control is amplified by the lack of integration of eHealth data into existing national and local medical records.

In this context, temporal trust checkpoints can give participants the means to **exert their rights for data control even in studies with digital data collection technologies**. That is, both participants and research institutions should be involved **in sequencing processes that will integrate data control parameters into the research process with minimal communication**. First, data control is negotiated and checked in a time-patterned sequence where the researchers and the participants' output occur independently (because researchers and participants are assigned to different time slots). Once a temporal sequence for data control is established and **repeated, it**

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

can be turned into a routine. Such a routine will ensure that even a small number of researchers' and participants' choices will activate trust-enforcing processes.

(4) Meta-Consent model

Participants' motivations are crucial resources for ensuring enduring consent and control transfer over personal data. However, such motivations can be characterised as **intangible** resources and thus difficult to handle. That is, participants' decision to give their consent is not value-free. The issue is to determine on which values consent is based on. In some instances, some of these motivations may be ethically objectionable (e.g., a participant refuses to grant his/her consent because of the ethnicity of the researcher). **Since such underlying reasons are not identifiable, ethically objectionable biases can shape data use** for some studies at the expense of others.

A way of controlling for such underlying biases is to use a **meta-consent model** where the consent status reflects ethical preferences. In practice, this means that choices for giving or not giving consent to data use is **structured according to a pre-defined set of options**. This requires **combinative capabilities** (e.g., recombination of options) and **aggregation capabilities** (e.g., moral preferences are aggregated and adjusted to consent options).

An example of applying combinative and aggregations capabilities for meta-consent is to include explicit participants' moral preferences in the consent form from the onset. For instance, there may be a consent option for studying the characteristics of vulnerable groups (such as refugees). The participant may then either accept or decline this option according to his/her ethical preferences. Meta-consent should also give participants to specify the type of questions and the frequency at which they wish to be re-contacted

A meta-consent model is also important for integrating **vulnerable populations** in the research process. In this context, the motivation and the consent of vulnerable populations are **highly specialized, fragile and heterogeneous resources characterised by high social complexity**. As such, it does not require combinative capabilities that simply use established resources to protect participants from harm. Namely, vulnerable populations may be more reserved in granting their consent for many reasons, including stigmatization risks. As a result, vulnerable populations are not fully represented in cohort studies. By taking reservations of vulnerable populations into account, health research actually compromises the development of appropriate treatments for them.

What is needed is a meta-consent model with **integrative capabilities and with knowledge integration mechanisms**. First, integrative capabilities of **aggregation and transfer** should be used to identify the purpose of data use (i.e., academic use versus for-profit use). Then, knowledge integration mechanisms (i.e., **group problem solving and decision-making processes**) should give patients the option to set their preferences from the onset. Meta-consent will then allow participants to specify the type of questions and the frequency at which they could be re-contacted. Such solutions imply technical and even ethical challenges (e.g., can the re-contacting be made through an app?; does the participant have access to the app?; can the participant use an app?) but they have the merit to target patients' preferences.

(5) Path oriented governance for consent

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

As a resource for consent, governance is a **tangible but heterogeneous resource**. The reasons for this are that governance is too versatile in terms of its application and too unclear in its execution. In cohort research, governance generally includes access committees and patient representations but in general terms, it is necessary to evaluate the current landscape to determine which effective governance structures are already in use. In such situations**, combinative integration capabilities** are enough: dataset types can be classified in terms of whether they require broad consent and associated governance structures or not. Some data and datasets (such as data with a high social value or aggregated and anonymised data) do not require broad consent and use a different legal base.

However, governance is a complex resource to handle because it originates from **specific historical conditions** namely, the GDPR. It is the GDPR (rather than purely ethical considerations) that shape current governance for consent. As a result, large institutions do not want to take the role of data controllers and put the burden of data controlling and consent back to data providers. Governance in cohort research thus presents resource **positions barriers** (i.e. data providers have to take data controlling action tasks that were previously determined by others) and **barriers to entry** (i.e. large institutions cannot timely implement governance because they have delegated data controlling duties elsewhere).

The solution is thus to find a **path-oriented and purpose-led model of governance**. **Combinative capabilities** would identify tools and pathways for data sharing while **sequencing processes** will define controlled data access arrangements. For instance, when a hospital receives data (and consent) from patients, it has to ensure that this data will be used only for legitimate purposes and by legitimate users. Doctors have to provide controlled access arrangements that monitor access applications by data users. Such an approach relieves infrastructures from administrative burdens and reinforces data-sharing safety. The only purpose of infrastructures in this context is to maintain data safety by minimizing risks (e.g., putting restrictions on data downloads).

(6) Intersection domains between federated analysis and governance structure for confidentiality

In order to enforce **confidentiality** requirements; the consensus is that **a federated analysis and governance structure to coordinate private and public interest are complementary solutions**. The Cancer Registry Advisory Board, for instance, uses two assessment levels. First, there is the purely technical level where data is kept in a confidential but scientifically usable form. At the second level, the main task is to determine which datasets are outliers and whether they should be suppressed from the data query system. Data stewards assess the impact of the exclusion of such datasets on the research value while having full discretion to act according to the privacy and confidentiality interests of the data subject. Governance in this context can be used to weigh and design an environment that both protects the data subject and ensures the scientific value of data.

In such a context, **federated analysis is a tangible, stable and specialized resource**. It is specialized because it applies to situations where data does not have to be moved from the place where it is stored. Such an arrangement ensures confidentiality and **does not require complex governance structures.** Researchers have no contact with the data per se and it is the institutions responsible for data storage that are legally liable for the security of personal data. In this sense, federated analysis has built-in **entry barriers** for researchers (i.e., they have no direct access neither to the input nor to the output data), which reinforces data security and confidentiality.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Thus, federated structures, while efficient in themselves cannot replace governance provisions in preserving confidentiality. This is because federated structures are characterized by **resource heterogeneity**. Federated approaches are not created equal since some systems rely on more resources than others. Under-resourced federated approaches will thus suffer from **resource immobility** and will be hard to leverage.

As resources, federated approaches are characterised by **causal ambiguity, which may hamper data re-use**. This means that the fact that researchers assess the data without ever seeing it, prompts the question about their ability to identify the source of potential biases. Data re-use is not neutral and the wording of the law in the privacy domain can lead to discriminatory limitations on data sharing. A preliminary solution in this context is to use **aggregate capabilities to determine the degree of centralization needed to mitigate resource immobility from federated analysis** (e.g., governance can be centralized in relation to an expert or should it be devolved to the local data host). This is a way to engage with the **anonymization requirements**: while federated governance may enforce anonymization on data by design, some studies (especially in cohort settings) cannot be done with anonymized or de-identified data.

The issue becomes clear if **confidentiality itself is evaluated as a resource**. Confidentiality is **a very specialized resource** because it is always used within the scope of consent. This means that the confidentiality issue is different when output (as opposed to input) data is concerned. That is, some kinds of **output data** can be aggregated in such ways that personal information about the data subject will remain hidden. In such cases, only **aggregation capabilities** are needed. Other types of output data can still reveal personal information about persons, even after aggregation. What we need in this context are **protective capabilities**, that is capabilities that preserve privacy rights.

In general terms, **the intersection between federated analysis and more centralized governance structure is defined by the goals and aims to be achieved**. Goals in this context, are **tangible and highly specialized resources**. The degree of governance complexity is determined by what researchers intend to do with the data and the extent to which they are willing to share it with others.

If a researcher intends to transfer data outside his institutions toother countries (for instance outside the EU), he/she would need a variety of legal tools such as data access, data use and data material agreements (i.e., **extensive integration capabilities and knowledge integration mechanisms**). If by contrast, the data stays inside the researcher's institution and federated analysis is used, then fewer instruments are needed (i.e., **combinative capabilities** are sufficient). In internal research paradigms, the researcher can simply rely on a data access community that ensures that subsequent data uses align with the consent's content. This means that **routines** for data access are already established.

(7) Leveraging the social value of research: sustainable data re-use mechanisms

Identifying what constitutes the social value of research is crucial, since there is no ethical or scientific justification for conducting research for its own sake. It is on the basis of specific civic responsibilities (rather than abstract concerns about justice) that we are able to ensure participants' involvement in research in a way that would be beneficial both for the participants themselves and the public as a whole. In order to identify quantifiable parameters of social value we need to stop considering the **social value of research as an intangible resource. Instead, social value should be specific in scope and linked to tangible benefits**. Social value should thus become

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

a **tangible, specialized and quantifiable resource**. As a result, the capabilities needed for fully leveraging the social value of research should be specialized and quantifiable as well.

*A way to bind the social value of research to specific civic responsibilities and thus to transform it into a tangible resource is to identify **future rather than present research benefits**. Judgements on the social value of research do not only rely on a clear scientific consensus of research value but also an **ethical consensus on the impact of research in the future** (usually implemented through Research Ethics Committees). The tangibility and specification of social value can thus be determined by Research Ethics Committees.*

However, quantifying and identifying the social value of researcher remains difficult in practice. The only currently available method is to ask researchers from selected case studies to outline the perceived impacts and benefits. Such a practice is biased both in terms of scope and in terms of the studies concerned (i.e. it is not clear on which basis such case studies are selected). As long as social value remains an intangible, ambiguous resource, some actors in the research community will be tempted to use conceptual shortcuts. In practice, this means that the social value of research is mainly understood in terms of the **extent to which data can be re-used for both scientific and public benefit**. However, the problem is that the **data community does not have a measure for data use**. Social value is thus tied to a resource (i.e. data use) that presents **significant resource immobility problems (i.e. it is difficult to use and apply)** .In such a context, it becomes difficult to use "data use" as a resource for determining the social value of research. Namely, available data might often not be re-used because there are no clear measurements and understanding of the benefits of doing so.

The solution in thus to engage in **knowledge translation by generating quantitative measures of data use benefits**. In order to achieve this, we need a set of integrative knowledge capabilities. First, **aggregation capabilities** will link **various existing knowledge translation frameworks in a consistent way**. Since these knowledge frameworks are already established and can be characterised as existing resources, **combinative capabilities** are also needed so that new combination of existing knowledge frameworks can be carried out. Second, **metrics to identify whether there is a return of investment from data uses can be created through appropriability and transferability capabilities**. In practice, appropriability and transferability capabilities will include **big data analytics and machine learning methods** that would help us to identify pathways of how research outcomes are eventually translated into practice. Hence, to determine the social value or benefits of research.

(8) Distributing research value: sustainable sharing of research burdens and research benefits

Another way to leverage research value is to interpret it in terms of research burdens and research benefits. Benefits are unavoidably connected to burdens because burdens are mostly carried by people who do not benefit from research. It can be argued that these burdens can turn into benefits later on but the issue remains that the distribution of burdens and benefits is not the same across groups. Burdens and benefits are thus meant to be **highly specialized and heterogeneous resources** if they are to be distributed correctly. That is, it is crucial to first **identify those groups that are the most likely to carry the burdens associated with research results**. Burdens and benefits can thus become **tangible resource because they will be adequately targeted to specific groups**.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

A central consideration in identifying groups that are the most likely to carry the burden of research is to improve their inclusivity in the research process. For instance, many treatments cannot be directly tested in cohorts related to disability for both ethical and legal reasons. As a result, individuals with disabilities are de facto excluded from appropriate treatment because the outcomes of research (e.g. medicines, drugs etc.) do not reflect their needs. In such contexts, vulnerable population carry the burden of research (they are the ones being targeted by the research aims) but they are excluded from research benefits.

Another important issue concerns **how the social value of research could be distributed**. Since distribution of social value is a matter of justice, we need to adopt broad measures on the international level that will ensure that **research does not always benefit the same groups**. We have to define specific regulations that would prevent international organizations to fund projects on the basis of their potential to bring financial profit (e.g. there is generally less funding for malaria research). This will require a strict **application of combinative capabilities** (i.e. recombining existing funds so that they could be evenly distributed) as well as **knowledge integration mechanisms (i.e. rules and directives that will regulate the application of knowledge about research benefits)**.

In general terms, we need a **multi-level sequencing knowledge integration mechanism**. At the patients' level, social value concerns ethics reviews, the study design and participants' benefits. At the policy level, social values are implemented through funding programs within the Chatham House framework [16] (e.g. Horizon Europe funding and outcomes is determined by the potential for societal impact). At the impact level, researchers generate KPIs (i.e., Key Performance Indicators) for impact monitoring during **group solving and decision-making processes**.

Such knowledge integration mechanisms should be integrated **with integrative and combinative capabilities** so that specific metrics of research impact could be identified. Namely, we should measure research impact by looking at the **networks' trajectories and the diffusion of information**. In order to achieve this, it is necessary to adopt a long-term view and to develop **unbiased interim metrics of impact**. This allows to evaluate specific researcher networks in order to see if they could have a potential innovation value in the future.

Much like confidentiality and consent, the overarching aim is to turn social values of research from intangible resource **into a tangible one, with its own quantifiable metric**. **Integrative knowledge capabilities** are needed to determine how the social value of research can be measured, addressed and implemented.

(9) Determining the social value of research: Sustainable Accountability mechanisms

Once the leverage and the distribution of the social value of research is determined, it is necessary to generate clear attributions of accountability. Namely, who is responsible for determining what research value is and who should enforce it? As it stands, **researchers face considerable resource position barriers and resource immobility problems when they try to determine research value**. First, research impact and social value are difficult to demonstrate when it concerns longitudinal studies. Second, an important way to demonstrate impact is to increase the re-use of data during data sharing activities. However, current funding does not favour data re-use and tends to attribute resources to primary data instead.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

In general terms, researchers' identification of research value can only be unstable and provisory. This is because researchers operate in **a context of ethical and conceptual serendipity.** That is, machine learning and AI-based analytics tend to generate serendipity both in terms of unintended findings and in terms of the impact on patients' lives. Deep learning tools guarantee that researchers will be confronted with results that they have not predicted and for which they have no theory. In determining research value, researchers thus increasingly rely on **resources with unstable emergence contexts**. They are causally ambiguous (i.e. findings cannot be fully predicted) and socially complex (i.e. the research impact on participants' lives cannot be fully foreseen).

This means that research institutions should have a stable, tangible resource base for evaluating research value. This can be achieved through knowledge integration mechanisms most notably, through **rules and directives**. Hence, from the onset, research institutions need to engage in an ethical analysis on the possible impact of research and its potential to bring benefits in the short and long term. This ethical analysis should be intergenerational in scope, as it needs to identify possible benefits for future generations. The strategy should be anticipative: we need to have an ethical impact assessment of the end goal before any funding of the research (**hence, transferability capabilities are needed**). Research institutions should balance the risks of unintended findings through a framework that would the risks of research serendipity. In general terms, **a governance framework for unintended findings is needed**.

Section Summary:

The components of Pillar II and its associated resources and capabilities are listed in Table 2.

Specialized tangible resources dominate Pillar II because guidance in methodological, legal and ethical terms requires explicitly quantifiable measures and fields of application. Pillar II is aimed to provide guidance to researchers and institutions with a range of different backgrounds, affiliations and interests and as such, needs to be as explicit in its advice as possible. The dynamic between specialized and versatile resources differs in significant ways from Pillar I. In Pillar I (i.e., Standards) versatile resources are nearly as important as specialized ones. Versatile resources and specialized resources can be converted into each other because they have the same weight throughout the standardization process. In Pillar II however**, versatile resources decrease in importance**.

At a first glance, this relative lack of importance of versatile resources may seem counterintuitive. Pillar II targets a large audience with different research aims and as such, seem to require versatile resources that can be applied to a wide range of contexts. A central characteristic of Pillar II however, is the **increasing importance of intangible resources**. Intangible resources are often (but not exclusively) associated to ethical parameters such as trust. Such intangible resources are often highly specialized because they are **personal and relational.** Participants' feelings of trust are unlikely to be the same across situations and depend on existing research arrangements. Pillar II aims to provide guidance to conditions for access and as such requires **acceptance from its purported audience**. **Intangible specialized resources thus provide a specific and relational anchor into the conditions for access** for participants and researchers alike.

Specialised tangible resources require a lot of integration and combinative capabilities while intangible resources are generally leveraged through integrative capabilities. This may be explained by the interactive relationship between these two resource types. Namely, many components of Pillar II result **intangible resources** made tangible (e.g., "Interim metrics of impact for research benefits"). This is because intangible resources tend to rely on tacit knowledge and

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

as such, cannot be easily shared across context. Thus, in order to use intangible resources effectively in the coordination of cohorts, it is necessary to partly quantify and operationalize them. The important number of capabilities needed for tangible specialized resources can be thus explained by the need to give a more quantifiable and observable aspect to intangible resources. **Intangible resources are thus converted into specialized tangible ones through a combination of integrative and combinative capabilities**.

An example of the conversion of intangible resources into tangible specialized resources:

Participants' motivations for consent (i.e. an intangible resource) are explicitly outlined in consent arrangements (i.e. a tangible specialized resource). This is done by specifying questions types, inserting predefined options and agreeing with the participants on the frequency they wish to be contacted.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Table 2 – SYNCHROS Strategic agenda – Pillar II

Component	Resource and Resource Type	Capability
Federated Infrastructure and Analysis Components	1) Costs and funds: Tangible unstable and specialized resources with resource position barriers: 2) Federated Analysis: Tangible resource with limited versality	1.) For Costs and Funds resources: -Combinative Capabilities: Divesting professional and data quality management to individual resources at the research institutions level. - Transferability and Aggregation capabilities: Determining the boundaries of usefulness for federated analysis. 2) For Federated Analysis: -Aggregation and Transferability capabilities: Conversion of data into a pseudonymized format while preventing data degradation.
Broad Consent Platform	Broad Consent: A relatively versatile resource but with intangible aspects. Includes resource position barriers and barriers to entry.	Group problem solving and decision making (integrative knowledge mechanisms): - Setting a platform with in-built opt-out and opt-in options according to ethical preferences Appropriability mechanisms (integration): - Preserving and distributing participants' control over their data to in relation to trusted research institutions.
Trust temporal checkpoints	Trust: A highly intangible, specialized rare and not substitutable resource.	Strategy 1 - Rules and directives (integrative knowledge mechanisms):

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Page 29 of 58

	Context of emergence: Unique historical conditions, causal ambiguity and social complexity.	-Governance system for checks and barriers to data misuse (includes long term control mechanisms) Strategy 2 – Routines and Sequencing processes (integrative knowledge mechanisms): - Setting sustainable routines for participants' data control: data control is negotiated and checked in a time-patterned sequence.
Meta-consent model	1) Participants' motivations consent: Intangible, highly specialized resources 2) Vulnerable populations motivations for consent: A highly specialized, fragile and heterogeneous resources Context of emergence: high social complexity	1) For participants' motivations -Combinative capabilities and aggregation capabilities (integration): Structuring consent according to a pre-defined set of options. 2) For vulnerable populations' motivations: -Aggregation capabilities (integration): Identifying the purpose of data use -Group problem solving and Decision-making processes (integrative knowledge mechanisms: Setting preferences from the onset: Specifying questions types and re-contacting frequency with the participants.
Path-oriented governance for consent	Governance for consent : Tangible and heterogeneous resource, includes barriers to entry and position barriers	Combinative Capabilities: -Identifying tools and pathways for data sharing Sequencing processes (integrative knowledge mechanisms):

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Page 30 of 58

	Context of emergence: From specific historical conditiond (i.e. the GDPR)	- Defining controlled data access arrangements
Intersection domains between federated analysis and governance structure for confidentiality	1) Federated Analysis and Approaches: Tangible, stable and specialized resource. Include entry barriers, resource immobility issues and resource heterogeneity. Context of emergence: Causal ambiguity 2) Confidentiality: A tangible?, highly specialized resource 3) Goals and Aims for governance: Tangible and highly specialized resources.	1) For Federated Analysis: -Aggregation capabilities (integration): Determining the degree of centralization and the degree of anonymization requirements. 2) For Confidentiality: -Aggregation and protective capabilities (integration): Aggregating output in such ways that personal information about the data subject will remain hidden. 3) For Goals and Aims: - Integration capabilities and knowledge integration mechanisms (rules and directives): Application of legal tools for transferring data to other countries -Combinative capabilities and routines for data access: -Using federated analysis for transferring and storing data within the research institution.
Equitable sharing of research burdens and research benefits (Unbiased interim metrics of impact)	Burdens and benefits: Highly specialized, tangible and heterogeneous resources. Potentially include resource barriers (i.e. research may be benefit some groups rather than others).	Combinative capabilities and Integrative knowledge capabilities: - Recombining existing funds so that they could be evenly distributed across groups

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Page 31 of 58

		-Identifying networks' trajectories and diffusion of information pathways Rules & Directives, Sequencing processes and Problem solving and Decision-making (knowledge integration mechanisms): - Implementing social values at the policy level through funding programs within the Chatham House framework -Researchers generating KPIs for impact monitoring
Sustainable Accountability mechanisms	Context of research/Ethical and conceptual serendipity: Highly versatile, intangible resources. Context of emergence: Causally ambiguous and unstable	Rules and Directives (knowledge integration mechanisms) and Transferability capabilities (integration): -Balancing the risks of unintended findings through a framework that would monitor good serendipity and bad serendipity (i.e. creating governance framework for unintended findings).
Sustainable data re-use mechanisms (Quantitative measures of data use benefits)	Social value: Tangible, specialized and quantifiable resource, includes resource immobility problems.	Aggregation capabilities (integration): - Linking various existing knowledge translation frameworks in a consistent way. Combinative capabilities: - Implementing new combination of existing knowledge frameworks Approprability and Transferability capabilities (integration): -Creating metrics for identifying investment return from data uses efforts.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Page 32 of 58

3.3. Pillar III: Participatory collaboration and knowledge exchange platform

The proposed participatory collaboration and knowledge exchange platform shall be the democratic forum for stakeholders for discussion, to express sound opinions, positions and to find corresponding information. As such, it is a pivotal element to foster trust in the research institutions and research as such. Below we list elements of such a platform that have been suggested by stakeholders to be covered and curated.

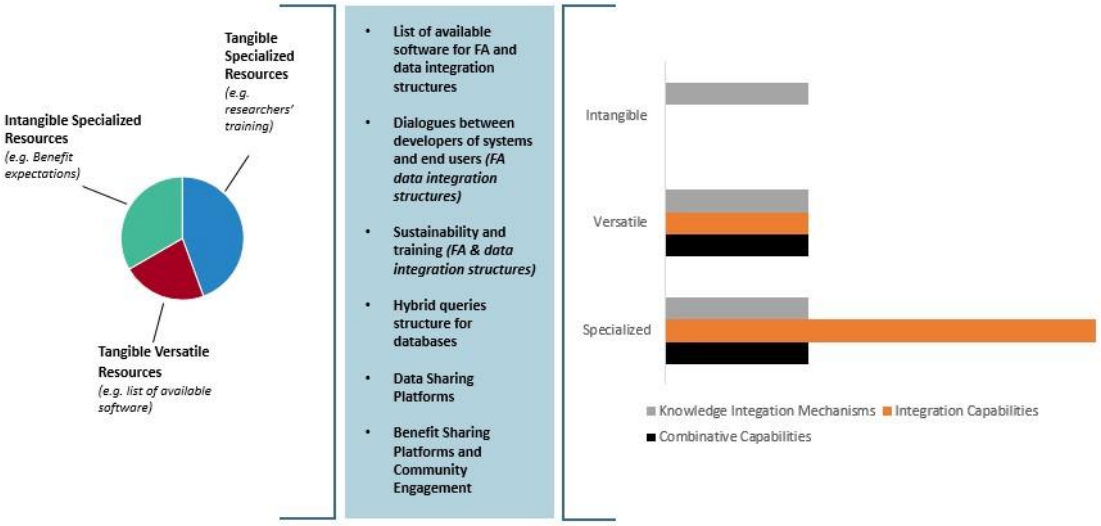

Figure 6 – Pillar III: The components required for Pillar III are outlined in the central column. The representativeness of resources types is shown in the chart on the far-left side (blue for tangible specialized resources, red for versatile tangible resources and green for intangible resources). The amount of capabilities needed for each resource type in Pillar I is represented in the bar chart at the far-right side (grey for knowledge integration mechanisms, orange for integration capabilities and black for combinative capabilities)

1. List of available software for federated analysis and data integration structures (Resource Creation)

It is crucial to have a clear picture of the pros and cons of the available software so that a federated system could be safe by design. That is, it is necessary to have a **list of the software available including their assumptions, properties, analytical possibilitie**s (e.g., Bayesian) and the requirements for central hubs and individual nodes. However, we have to first examine what such a list of software means in terms of resources. A list of available software is an existing, tangible and versatile resource. **Such a resource can be applied to a variety of contexts** (hence its relevance for the design of federated infrastructure and platforms) but is still largely **immobile and present resource position barriers**. That is, the software in use is heterogeneous and not well documented. As a result, it is impossible to know if it is scalable enough for a large number of cohorts and nodes.

What are the required capabilities for designing such a software list so that resource immobility problems could be avoided? The priority should be given to those integrative capabilities that are

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

able to **link cohort data with federated infrastructures**. More specifically, aggregation capabilities would allow reaching the level of standardization required for relating federated systems to cohort data effectively.

Such aggregation, standardization-based capabilities would allow us to create an inventory with an overview of all the software available. An **inventory of federated software** will be a highly versatile resource that would indicate what kind of requirements software have in relation to federated systems and the kind of analysis this software is able to perform. As a resource, an inventory of existing software will rely on **combinative capabilities that will link it to the landscape of use for specific software by different projects**. That is, we should focus on big cohorts (as opposed to small) ones, when identifying the software used and evaluate which standardization strategy is used in each case.

2. Federated Analysis and Data Integration Structures: Dialogues between developers of systems and end users

Federated analysis is a tangible but heterogeneous resource: it requires from the researchers and the research institutions a bundle of technological, financial and organizational resources. **Not all research institutions are equal in terms of the resources they possess for federated analysis**. Some institutions may rely on extensive metadata and established analysis tools, while others have datasets that are not easily susceptible to harmonization. Because of such differences, federated analysis becomes a specialized rather than a versatile resource as it includes barriers to entry (e.g., there is a disadvantage in performing federated analysis when technological resources are missing).

Therefore, **in order to avoid specialization and resource barriers**, any significant collaboration within the research community on federated analysis requires strategic alignment: there should be an agreement about what a practical improvement from federated analysis means in concrete terms. The **practical improvement** from federated analysis can be defined through a dialogue between the **federated system developers and the end-users** (i.e., the researchers). Such a dialogue requires integrative capabilities with coordinated efforts from both developers and researchers, especially in terms of transferability (i.e., the assumptions of federated software that can be directly translated in user-friendly methodological solutions).

3. Federated Analysis and Data Integration Structures: Sustainability and training

As we have shown, effectively using federated analysis as a resource requires creating an inventory of existing software and a dialogue between researchers, statisticians and software developers. However, it is also crucial to understand the extent to which data infrastructure and federated analysis are **sustainable for collaboration and data sharing knowledge creation**. The third pillar should have a mechanism to check **the impact of the federated analysis and structure in use once a research project has been completed**. Only then would we be able to ensure that the scientific strategies used during a project will have a positive impact for further research. In other words, rather than sharing data for its own sake, the focus should be on sustainable knowledge creation through data sharing and access.

Researchers' training guarantees a long-lasting positive impact from research strategies (such as federated analysis) for future projects. **Training for researchers depends on the software they already use** since different tools require different training. It is thus a resource issue because

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

training depends on the **materials and tools already available**. That, the resources for researchers training are (at least initially) **specialized**: they depend on the particular context of practices in which the researcher operates. It is thus crucial to find ways to avoid resource immobility (e.g., the training is ineffective because it is too tightly related to the software used) and barriers to entry (e.g., researchers with less resources and less advanced analytical tools receive a less sophisticated training).

One way to extend and sustain training resources is to relate combinative capabilities with integrative ones. **Combinative capabilities can be used in a short-term perspective**: they can be used to construct **an inventory** of what tools are available and what training needs should be met. **In the long term however, integrative capabilities** are needed. First, transferability capabilities should identify how methodologies developed with federated infrastructures in mind can be extended to other infrastructure types. Second, aggregation capabilities can be used to link these methodologies with ethical and legal frameworks (such as confidentiality and privacy). Finally, appropriability capabilities should use the results of aggregation and transferability processes in order to credit knowledge owners and facilitate knowledge exchange. This can take the form of workable data lakes for a secure hub of private data.

4. Hybrid Queries structure for databases

Some EU initiatives build a virtual platform that contains a federated data discovery ecosystem. Such an **ecosystem** is a highly versatile resource because users can get access to a variety of linked disease databases, in this case for rare diseases, at different levels of specificity. The third pillar (participatory knowledge exchange platform) should thus adopt the same approach to databases and database queries. In particular, it should focus on sequencing knowledge mechanisms because they minimize unnecessary communication while integrating knowledge through time-patterned steps. In practice, this sequence of knowledge integration for queries is adapted to **the degree of specificity**. For the lowest specificity level, automatic queries arrangements quickly give a response without asking for detailed permission. At the higher specificity levels, users have to get in contact with the resource and ask for permission from data access committees.

In more specific terms, a participatory knowledge exchange platform (Pillar III) should help researchers making appropriate queries in the databases. In terms of integrative knowledge mechanisms, data holders and data providers can already rely on **established routines**. They have standard analytical tools, good practices derived from federated analysis and Bayesian analysis when they have to engage in data assembly themselves. In other words, they can rely on stable, tangible and versatile resources. Routines related to such resources require less effort because they can support complex actions even if fixed rules are not in place.

However, such established routines and resources have less to say about what the users actually need. As we have seen, researchers have heterogeneous resources, and some are more advantaged than others (e.g., they may have more funds or more advanced technical tools). If we fail to take the **individual position of researchers** into account, then the stable resources and routines for data queries run the risk of presenting significant resource barriers for under-resourced researchers.

In order to avoid this, we need to adopt a **hybrid approach to a query strategy that will use both federated and centralized arrangements** in accordance with the local context. In order to implement such a hybrid approach, we need first to evaluate the resources at our disposal through combinative and aggregation capabilities. Namely, while some data centres are well

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

equipped for complex queries, others are hampered by discoverability problems. Combinative capabilities can be used to **identify which approach database setters have taken in certain local contexts**. Such approaches will then be **converted** through knowledge integration mechanisms into directives for queries organisation. That is, the approaches used in certain contexts will be used as **benchmarking** when input fields in large-scale cohorts are concerned.

It should be noted that hybrid approaches to queries should rely on **national infrastructure resources** because they are both versatile and specialized. In France, Germany, and Wales, researchers are creating national hubs of health data (mixed with social science data) that require various safe haven data schemes. **Safe heaven data schemes** allow users to access the database under certain conditions and preserve the potential use of data resources of national hubs. This relies sequencing a knowledge mechanism process that allows researchers' inputs to occur independently in a timely and coordinated manner.

5. Data Sharing Platforms

As knowledge integration mechanisms, sequencing and automatic queries responses (cf. 3.4.4) are useful as long as the data is not shared. **Sharing data** however, warrants **compliance to the specific requirements of the cohorts** and thus requires the creation of versatile tangible resources. More precisely, a **state of the art of national initiatives** helps to determine what kind of data sharing strategies for queries are needed.

Combinative capabilities can be used to create a state of the art of national initiatives and project collaborative research initiatives. This will allow identifying **bottlenecks and practices of the current initiatives**. Alternatively, combinative capabilities and knowledge aggregation capabilities can be used to **integrate existing cohorts** (with owners' agreement). This will allow identifying the real challenges and facilitators in practice and uncovering unexpected problems.

We then need knowledge integration mechanisms focused on group solving and decision-making. In practice, this means that a **forum and a pilot proposal** (on the basis of specific small user cases) should be created. This would allow identifying possible valuable approaches, the limits of these approaches and the architecture that would best suit users' needs. **Integrative knowledge mechanisms (i.e. groups solving and decision planning)** should be again applied later on so that researchers from different domains could collaborate on knowledge transformation steps. In practice, this means that researchers would **test the concept field**, develop practical solutions and test their feasibility (by for instance, integrating results back into a consortium, in this context, aggregation capabilities are needed).

6. Benefit Sharing Platforms and Community Engagement

Data sharing has to bring benefits not only for the scientific community and the general public, but also for the participants of cohort research. As a resource, a **research benefit is problematic because it contains both tangible** (e.g., scientific quantifiable benefits) **and intangible components** (e.g. the ethical purpose and social value for data and results sharing). Even the tangible aspects of research benefits are ambiguous. Namely, while it is assumed that disinvestment, data sharing and retrospective harmonization bring benefits for research, the nature of these benefits remains unclear. **Researchers do not have metrics for data sharing activities** and therefore, are not able to fully assess the data quality that they need.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

This lack of clarity can be explained by the emergence context of research benefits. Namely, research benefits emerge as resources in **contexts marked by causal ambiguity** (i.e. the benefits of cohort research are not easily determined in advance) and **social complexity** (i.e. ensuring that benefits are evenly distributed, especially across vulnerable populations remains difficult). Research benefits are thus **difficult to share** because they display resource immobility issues, resource position barriers (i.e. participants are disadvantaged because research institutions have access to research benefits before they do) and barriers to entry for participants.

An illustrative example in this regard is WHO's data sharing of Zika virus studies. While its data sharing was valuable for Zika virus research in general terms, it failed to deliver any benefit for the people affected by the Zika virus themselves. Data sharing in itself is not enough to enforce an equal distribution of research benefits.

As we have noted, research benefits have also intangible aspects that are not quantifiable. That is, **participants have expectations about tangible benefits from the future re-uses of their data**. When data subjects provide broad consent for future use, participants have a tendency to think that they will be informed about any incidental findings. If we consider participants' expectations as a resource for the creation of a participative collaboration platform (Pillar III), then we have to consider its source.

First, participants' expectations about potential benefits originate in significant causal ambiguity and social complexity: expectations of benefits from incidental findings are **not necessarily aligned with privacy expectations**. That is, participants are not necessarily prepared to give up their privacy preference, rights and needs in favour of benefits that may simply not be there. As long as this causal ambiguity and social complexity of participants' expectations is not solved, we won't be able to accurately evaluate the trade-offs between privacy, personal benefits and potential harm.

The solution here is to consider **privacy as a tangible and specialized resource** that allows integrating participants' expectation about potential benefits. In practice, we have to focus **on privacy harms rather than privacy per se**: privacy as a legal and ethical resource is supposed to protect participants from harm. As a result, participants' expectations about potential benefits become specialized and tangible as well because they are tied to their study relationship with the researcher.

It is this crucial to make these patients' expectations about privacy and data uses both **explicit and negotiable**. **Tacit knowledge should be thus, translated into explicit knowledge through group problem solving and decision-making** for both participants and researchers. Through communication intensive forms of knowledge integration, there should be room for discussion and consultations about what do participants' expectations mean and how they can and cannot be implemented in their particular context. The participants should be able to negotiate their scope of expectations for privacy with the researcher(s) and base their further agreement on data re-use for it.

In order to effectively use and distribute research benefits as a resource, we also need to develop integrative and combinative capabilities in order to systematically **measure and assess the nature, distribution and potential benefits of data use**. Another way to ensure that research benefits are used and distributed effectively is to reinforce **community engagement**. Community engagement facilitates consent as it clarifies to participants what they are consenting to in the long term. Community engagement is a causally ambiguous resource because of the nature of cohort research. Namely, the nature of cohort research is longitudinal and hence, the full impact of research results may be known only in the distant future.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Community engagement can be reinforced though integrative knowledge mechanisms (such a group problem solving and decision making) that will increase data subjects' literacy about the benefits and nature of research on the long term. Group solving and decision-making processes should involve **joint knowledge transformation steps where goals and future data would be clarified**.

Section Summary:

The components of Pillar III and its associated resources and capabilities are listed in Table 3

Specialized tangible resources continue to play a prevalent role in Pillar III (i.e. Inclusion) and **intangible resources** are well represented. In this sense, the interaction between tangible and intangible resources are similar to what was observed in Pillar 2 (Guideline). The difference is that **there is no direct conversion of intangible resources into tangible ones**. Both resource types operate in parallel: specialized tangible resources ensure that the content of participatory collaboration platform are both measurable and explicit while intangible resources reinforce participation through reinforcing relational aspects (mostly through integrative capabilities).

An example of the intangible resources and specialized tangible resources working in parallel:

Researchers' training (a specialized tangible resource) for federated analysis and for implementing reliable and interoperable data infrastructure ensures that the access, curation and sharing of cohort data is scientifically valid and reliable. In so doing, training increases the social value of research (an intangible resource) because data can be shared in a secure and consistent manner (i.e. research cannot have a social value if it is scientifically not valid).

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Component	Resource and Resource Type	Capability
List of available software for federated analysis and data integration structures	List of available software: a tangible and versatile resource. Displays resource immobility and resource position barriers.	Aggregation capabilities (integration): -Linking cohort data with federated infrastructures Combinative Capabilities: -Linking inventory of software to the landscape of use for specific software by different projects.
Federated Analysis and Data Integration Structures: Dialogues between developers of systems and end users	Federated analysis: a tangible but heterogeneous resource: a specialized resource Includes barriers to entry	Transferability capability (integration): -Implementing a dialogue between software developers and end users for knowledge translation (i.e. the assumptions of federated software that can be directly translated in user-friendly methodological solutions).
Federated Analysis and Data Integration Structures: Sustainability and training	Researchers' training: a specialized tangible resource.	Combinative capabilities: -Constructing an inventory of what tools are available and what training needs should be met. Transferability capabilities (integration): -Linking methodologies for federated infrastructures to other infrastructure types. Aggregation capability (integration): -Linking methodologies for federated Appropriability capabilities (integration): -Crediting knowledge owners and facilitate knowledge exchange (e.g. creating data lakes for a secure hub of private data)

Hybrid Queries structure for databases	Federated data discovery ecosystem: a versatile tangible resource	Sequencing (knowledge integration mechanisms): Adapting queries to specificity levels Routines (knowledge integration mechanisms): -- Relying on existing standard analytical tools, and good for data assembly. **Combinative capabilities and Aggregation capabilities (integration):** -Converting local approaches to datasets database setters have into wider directives for queries arrangements.
Data Sharing platforms	State of the art of national initiatives: versatile tangible resource	**Combinative capabilities:** -Identifying bottlenecks and practices of the current initiatives. **Aggregation capabilities (integration):** -Integrating existing cohorts with owners' agreement **Group solving and decision-making (knowledge integration mechanism)** -Creating a pilot proposal on the basis of specific cases and testing the concept field and feasibility in collaboration with other researchers.
Benefit Sharing Platforms and Community Engagement	1) Research benefit: both tangible and intangible specialized resource. Emergence context of resources: causal ambiguity and social complexity	**Group problem solving and decision-making (knowledge integration mechanism):** -Translating tacit knowledge about research benefits into explicit knowledge through negotiations

	High potential for resource immobility 2) Participants' expectations of research benefit: intangible specialized resource 3) Privacy: tangible and specialized resource 4) Community Engagement: intangible resource	between researchers and participants about the implications of data reuse. - Creating joint knowledge transformation steps between researchers and participants where goals and future data uses should be would be clarified

3.5 General summary: Resources and Capabilities for Pillars I, II and III

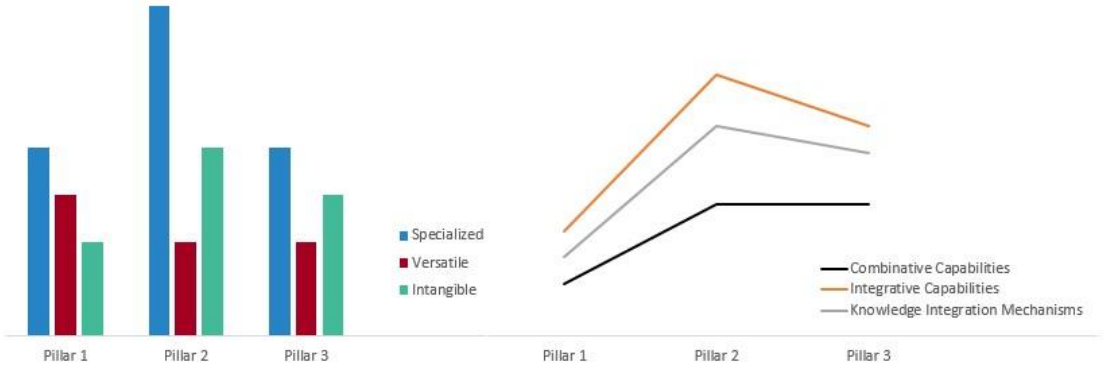

Figure 7 - Evolution of Resources and Capabilities across Pillars I, II and III: The chart on the far-left side represents the amount of resources types needed for each Pillar (tangible specialized resources are in blue, versatile tangible resources are in red and intangible resources green are in green). The chart on the far-right sides represents the amount of the capabilities needed for each of the three Pillars (knowledge integration mechanism are in grey, integration capabilities are in orange and combinative capabilities are in black)

The main outcome of our analysis is that the **role of versatile tangible resources progressively decreases in importance** across pillars. Versatile resources are important in Pillar I (Standards) because they represent the expected target for local, specialized resources. That is, the scope and reach of standards very much depend on the extent to which they can be applied across various contexts. Hence, one of the main aims of Pillar I is to ensure that **local, specialized standards and resources are converted into versatile common ones** so that they could be shared and transferred.

The picture changes as soon as we approach the fields of actions for Pillar II (Guidance) and Pillar III (Inclusion). Here, the **interactions between specialized tangible and intangible resources** take centre stage. In Pillar II, the amount of knowledge needed is more intense (cf. Figure 7) because **converting intangible resources into specialized tangible ones** requires a lot of effort. This dynamic stabilizes in Pillar III because inclusion platforms do not necessarily need quantifiable intangible resources in order to ensure collaboration and participation. This is because acceptance has already been secured in Pillar II through the "tangibilization" of intangible resources. As a result, the participation platform in Pillar III is the product of **complementary outputs** of specialized tangible resources and intangible ones. Specialized resources ensure that the content of the platform is as explicit as possible while intangible resources support sustainable participation.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

4. STRATEGIC AGENDA

4.1 Overview

We consider the three pillars outlined in chapter **¡Error! No se encuentra el origen de la referencia.** as the major strategic domains along which specific strategic tasks can be ordered and formulated. Initiatives and projects can be planned and conducted in a structured, complementary and synergetic manner.
Nevertheless, analysing the stakeholder's positions and looking at the scattered landscape of patient- and population cohorts, the different types of stakeholders, projects, organisations and initiatives focusing on cohorts, synergies and federated data platforms, the pure compilation of information beneath the three pillars will be insufficient until there is a democratic, widely accepted and sustainable organisational structure that coordinates, maintains and curates the information and activities under the three pillars.

Therefore, and first, we suggest as a priority of a strategic agenda a central institutional instrument, meaning the establishment of what we call an Institute for Optimising Multistudy Integrative Cohort Research in Health (working title) governed by an international board of curators. It´s mission should be coordinating and curating the information and activities under the three pillars including the definition, initiation and supervision of further strategic activities. Of course, the role of this institution can be mandated to an existing, well-established institution, if appropriate. We see one of the European Commission's Directorates, e.g., Research and Innovation, as the right place to initiate and steer the process.

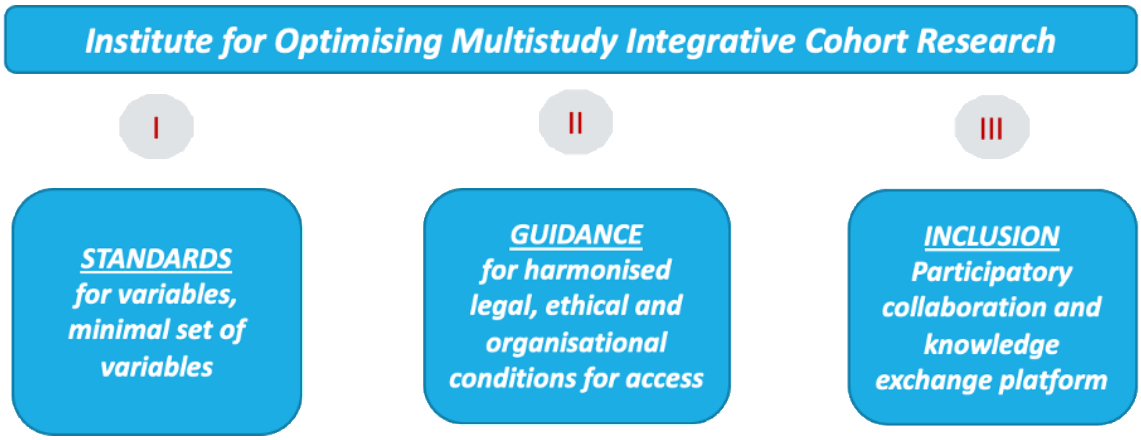

Figure 8 - Three pillars of action coordinated by a sustainable organisational unit.

Mapped into a management- and organisational structure, the strategic agenda for better coordination of cohorts globally consists of a visionary goal, strategic domains of action (i.e., the pillars), strategic tasks derived from the gap-analysis of the status quo and corresponding projects that tackle the specific issues, **¡Error! No se encuentra el origen de la referencia.**.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

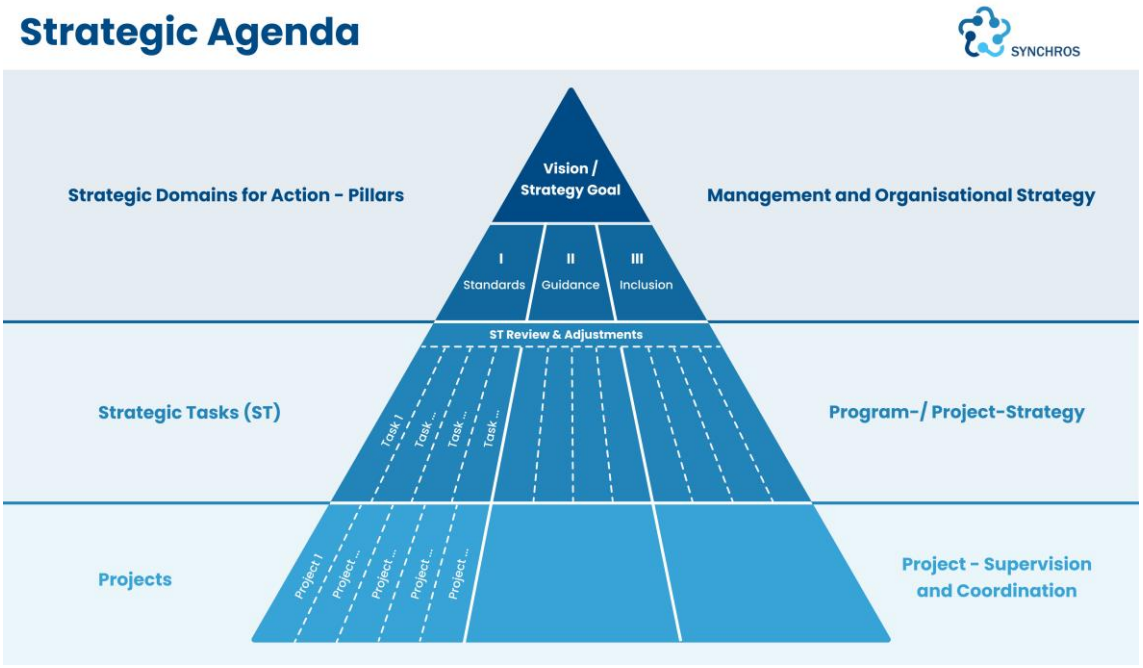

Figure 9 - SYNCHROS Strategic Agenda - from vision to actions

4.2 Strategic Tasks

In chapter 4.1, as a priority, we suggested the establishment of an Institute for Optimising Multistudy Integrative Research in Health governed by an international board of curators, either as a new entity or embedded into an established organisation. If the addressees of this deliverable agree with the basic statements of this report, the following, central and primary strategic task is a logical consequence that should be tackled first:

Strategic Task 0: Elaboration of the details regarding organization, management, financing and implementation of the strategic agenda, particularly of the Institute for Optimising Multistudy Integrative Research in Health. This step includes the consultation and alignment with institutional stakeholders such as the EC-DIRECTORATE-GENERAL Research and Innovation, WHO-Europe, BBMRI, IHCC, Maelstrom etc.

Furthermore, the multitude of points expressed by stakeholders can be condensed into the following list of initial and - obviously - most relevant strategic tasks, ordered by pillars. This list can't be neither complete nor static. Rather, it is an initial approach covering suggested major points of action as seen by stakeholders in the period 2019-2022. It also practically illustrates a possible organizational and managerial strategy suggested to be implemented in order to tackle the tasks. The list of strategic tasks needs to be subject to periodic review, adjustments and updates. Therefore, we consider the activities of review, adjustments and update as a generic and necessary pivotal strategic task in each pillar. We complement the list by examples and use cases, which - as we believe - illustrate the direction of how the issues can be tackled step by step (text boxes).

Pillar I: Standards – Initial Strategic Tasks:

According to our analysis (cf. Section 3), these strategic tasks aim to convert specialized tangible resources into specialized versatile ones.

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

1) **Elaborate, organize and implement the personnel, organizational and technical structures** and means for fostering the process and structures for the standardization and de-facto-standardization, for the sustainable periodic review, adjustments and update of strategic tasks including the details of a sound, sustainable budgetary related to **Pillar I**.

> *Standards are not new; they have proven to be an indispensable prerequisite for further developments in many areas of science and technology. Without standards, our cooperative, collaborative and global world does not function. Compared to this, standardisation in cohort-data collections is still at the beginning. But we can learn from it: For instance, ISO – the International Organisation for Standardisation [17] - has implemented proven processes from proposing standards, elaborating details, evaluation, enquiries for stakeholders, approval and publication.*

2) **Generate a state-of-the-art picture of queries practices and queries organization (de-facto standards). Strong encouragement for the use of existing metadata standards and de facto standard "sets of variables" in all new cohort studies prospectively** and creation of an inventory of existing de facto standard "sets of variables" for specific health-related questions including its definitions, stratification and metadata descriptions. The inventory should be targeted to the needs of the research community. References to findings of the following strategic tasks need to be integrated.

3) **Harmonization of metadata standards for metadata cataloguing.** A number of catalogues and initiatives exist. Examples are Maelstrom Research [18], CLOSER [19], HL7 [20] or projects like EU4Digital [5]. Hereby, Maelstrom - for instance - focuses directly on the issues of cohort-data while HL7 refers to electronic communication processes in healthcare and the application layer (#7) of the OSI-model[6], and EU4Digital provides a guideline for Harmonisation and Interoperability in eHealth in general. The examples illustrate that similar challenges exist in different technical and scientific domains, and initiatives have been established fairly independently. It may be unclear whether metadata is directly exchangeable across existing metadata archives. From that we can derive the strategic task to analyse and evaluate the standardization efforts and outcomes in related scientific and technical domains in order to identify synergies and assets usable for meta data standardization in the context cross cohort research. The action point would be to evaluate differences in standards and to create catalogue harmonization to create a single metadata taxonomy that would permit catalogue integration.

[5] https///eufordigital.eu/wp-content/uploads/2021/03/Common-Guidelines-for-eHealth-Harmonisation-and-Interoperability.pdf

[6] OSI (open systems interconnection), e.g. for clinical, radiological and administrative information systems

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

The process of effective implementing structures of a true standardisation will foreseeably be a long-lasting challenge. Research cannot wait until this is completed. Practically, there is a broad range of areas where ***existing metadata standards and sets of variables*** *are being used in research and clinical practice alike. I.e., the broader adaptation of internationally frequently used methods and sets of variables for data collections, its measures and stratifications / classifications are a pragmatic step towards broader compatibility of cohort data. Examples are:*

- ***Respiratory questions*** *- ISAAC (International study of Asthma and Allergy in Children) [21]*
- ***Eating habits*** *- EAT (Eating Attitudes Test, P.E. Garfinkel, and A. Newman 2001) [22]*
- ***Self-esteem questions*** *- Rosenberg Self Esteem Scale (RSES) [23]*
- ***Emotional symptoms*** *measured using the SCL-5 (Hopkins Checklist) [24]*
- ***Personality Check*** *adapted from EPQ (Eysenck Personality Question) [25]*

New studies, and studies included in new consortia, should be strongly encouraged (through funding, requirements) to make use of the metadata standards in developing their codebook and data files that would immediately enable interoperability with current metadata catalogues and individual studies for collaborative, cooperative, comparable, and replicable research.

4) **Transfer of metadata taxonomy (variable labels) to individual studies to create common data-level variable naming conventions and improve conditions for immediate data use in consortia and preparation of data archival/sharing**.
Metadata cataloguing of cohort studies has progressed rapidly over the last 5+ years and many EU and international studies have completed metadata cataloguing of their codebooks, with additional studies being added currently. An important next step is to move the metadata standards back to the original study—providing the map for idiosyncratic naming conventions to the common variable labelling (metadata standard). This seems like a tractable next step towards FAIR data principles [26] closing the gap between metadata catalogues and mappable access to individual datasets. It would increase opportunities and efficiency for multistudy data sharing and analysis at the study-level and will permit consortia to immediately move to analysis rather than spend years of effort in evaluating harmonizability, etc. The realization of metadata standards at the cohort data level is the necessary step to archive data that will permit subsequent data use and advance towards the ideal SYNCHROS goals.

5) **A comprehensive overview of metadata/codebooks will help to guide selection of tests/assessments in future research.** The evaluation of harmonization potential and harmonization procedures/results, as well as identification of available variables in particular studies, are essential for guiding the design and measurement protocol for new cohort studies. The use of "best" measures and/or those used in studies, which can then be compared in future work are further outcomes of accessible data of this type.

6) **Fostering the use of existing respectively emerging European federated data infrastructures**[7] for storing and managing cohort data. To develop standard measurements for data sharing activities and the impact of such activities. In order to achieve this aim, we need to

[7] https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12663-Digital-health-data-and-services-the-European-health-data-space_en

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

consistently determine the content of data sharing activities, their domain of application and implementation as well as the actors involved.

> *An example of emerging federated data infrastructures is the EU Health Data Space. It provides a consistent set-up for the use of health data for research, innovation, policy-making and regulatory activities. The activities of the Health Data provide the context, parameters and impact for data reuse.*

Pillar II: Guidance:

According to our analysis (cf. Section 3), these strategic tasks aim to convert intangible resources into specialized tangible ones.

1) **Elaborating, organising and implementing the personnel, organizational and technical structures** and means for the sustainable periodic review, adjustments and update of the strategic tasks related **to pillar II**, Guidance. Task II-1 needs to include the details of a sound, sustainable budgetary basis.

> *Examples: WHO (World Health Organization level) supports GATHER guidelines for global health estimates. GATHER guidelines are now de facto requirements for publication in high impact journals. [9]*
>
> *The EU Health Data Space (Institute for Optimising Multistudy Integrative Research) should participate in the process and adopt appropriate infrastructures for cohort. [27]*

2) **To develop a catalogue of infrastructure that would allow institutions at the EU level to steer and structure a catalogue for recommended standards.** This means establishing and maintaining a meta registry of larger (previous, ongoing and new) clinical and population cohort data collection initiatives/projects and maintain / further develop repositories of cohort harmonization initiatives such as Maelstrom, and the SYNCHROS Repository. The task includes a definition of the inclusion and exclusion criteria for data collections being registered. The aim is thus to identify the boundary conditions for a constant registration of data in cohort registries namely (i) what kind of project should be registered in such registries and (ii) what kind of strategies should be used for registering initiatives. This means that we also need to determine the threshold of the amount of population data in these registries.

> *For Clinical trials we already have this kind of registry – the EU- and U.S-. Clinical Trials Registers (https://www.clinicaltrialsregister.eu; https://clinicaltrials.gov). If these or a similar registry would cover not only clinical data collections but also population data collections it was a source for researchers to check whether data collection related to a specific research question was carried out elsewhere*

3) To encourage PIs of relevant data collections to **publish cohort profiles** and – in case of changes or new data collections **- cohort profile updates**. These cohort profiles must cover at least who is in the cohort (number of participants, ages, gender), what has been measured (variables and definitions, metadata) and study designs. An example of a cohort profile update is [30]. Furthermore, all information about the cohort and the access conditions should be available online (*https://hunt-db.medisin.ntnu.no/hunt-db/#/studypart/1*).

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

4) To develop **an widely usable and adjustable curriculum for students who are interested or foreseeably engaged in medical research, epidemiology and statistical analyses of cohort data. This curriculum should cover among others** standards and de-facto standards for variables and metadata, metadata harmonisation, available software use and resources for support in these matters (e.g. a future "**Institute for Optimising Multistudy Integrative Cohort Research").** Such a curriculum should be applicable to the actors involved in cohort research (e.g. students of medicine, health policy practitioners).

Pillar III: Inclusion:

According to our analysis (cf. Section 3), these strategic tasks aim to combine complementary outputs from intangible resources and specialized tangible resources.

1) **Elaborate, organise and implement the personnel, organizational and technical structures** and means for the sustainable periodic review, adjustments and update of the strategic tasks related to **Pillar III**, Inclusion. Task III-1 needs to include the details of a sound, sustainable budgetary basis.

2) Promotion of **trust** in research institutions and research in general:
   - **Counterbalancing and identifying power structures** in commercial and non-commercial research institutions in order **to (i) ensure transparency** and **(ii) distribute belief** and **trust back to research institutions** (addressee: general public).
   - Fostering researchers' **engagement with the individuals behind the research data** in order to ensure sustainable trust and social value of research (addressee: study participants).

Example: One of the worldwide largest longitudinal population cohorts is the HUNT-Study in Norway [28]. It is a unique database of questionnaire data, clinical measurements and a biobank of samples from the county's inhabitants from 1984 onwards. The HUNT-institute, as part of Norway's largest university, NTNU, was deliberately located about 70 km north of Trondheim in a smaller town, in the middle of an urban, lively structure. Many employees are recruited from the city and the surrounding communities. They are involved in the scientific work and the institute conducts active public relations work about its own activities. The legal and ethical regulations for the collection of data and further handling of it are known and communicated transparently. The integration of the research infrastructure and the close relation between employees and the population in the province of Trøndelag is seen as the major reason for its broad acceptance and a high participation rate in studies of 89% to 54%.

3) To identify the extent to which **the expectations and beliefs about personal rights from the general public** and/or participants can be justified and **balanced** in relation to scientific **research progress** (i.e., accumulation and creation of knowledge).

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Example: Integrating societal preoccupation with personal rights or/and participants' expectations of personal research benefits in the research process will generate mutual trust but it should not be done at the expense of scientific progress.

For instance, the value of retaining and respecting privacy is unclear. Namely, privacy is expected to protect persons from harm. However, what really constitutes harm in relation to information disclosure is debatable. Information sharing is not necessarily harmful in itself especially if the person is not aware that an information breach happened in the first place. This suggests that a governance structure that prevents participants from knowing how their data are used is the way forward. [29]

4) **Developing strategies for the integration and interoperability of data obtained through eHealth and emerging digital communication technologies (EDCTs)** with (i) international ICT platforms (ii) other platform types (EHRs) and (iv) other data types (e.g., clinical) (e.g. the European Medical Information frameworks can be interoperated with a focus on dementia with data from wearables).

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

5. REFERENCES

[1] SYNCHROS.eu https://SYNCHROS.eu (last accessed September 15, 2021).

[2] SYNCHROS Cohort Repository https://repository.SYNCHROS.eu/ (accessed September 14, 2021).

[3] Bickenbach J, Sjalm A. Strategy brief on ethical and legal issues in the optimisation of cohort data in Europe. SYNCHROS Project, 2020.

[4] Bickenbach J, Sjalm A. Strategy brief on harmonization and integration methods. 1–34.

[5] Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of management, 17(1), 99-120

[6] Andreu, R., Baiget, J., & Canals, A. (2008). Firm-specific knowledge and competitive advantage: evidence and KM practices. Knowledge and Process Management, 15(2), 97-106.

[7] https://ddialliance.org/ (last accessed May 20, 2022)

[8] https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions (last accessed May 22 2022)

[9] http://gather-statement.org/ (last accessed May 21, 2022)

[10] https://joint-research-centre.ec.europa.eu/jrc-news/new-user-friendly-tool-european-platform-rare-disease-registration-2022-02-28_en (last accessed May 21, 2022)

[11] https://www.cdisc.org/standards/data-exchange/define-xml (last accessed May 20, 2022)

[12] https://www.ohdsi.org/data-standardization/the-common-data-model/ (last accessed May 20, 2022)

[13] https://www.bbmri-eric.eu/ (last accessed May 22, 2022)

[14] https://eupha.org/index.php (last accessed May 20, 2022)

[15] https://eric.ed.gov/ (last accessed May 21, 2022)

[16] Røttingen, J., Ottersen, T., Ablo, A., Arhin-Tenkorang, D., Benn, C., Elovainio, R., ... & Yang, B. (2014). Shared responsibilities for health: a coherent global framework for health financing. *Final Report of the Centre on Global Health Security Working Group on Health Financing*. (last accessed May 2o, 2022)

[17] https://www.iso.org/home.html (last accessed May 2022

[18] https://www.maelstrom-research.org/ (last accessed May 22, 2022)

[19] https://www.closer.ac.uk/news-opinion/blog/building-closer-metadata-search-platform/

[20] https://en.wikipedia.org/wiki/Health_Level_Seven_International

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

[21] Asher, M. I., Keil, U., Anderson, H. R., Beasley, R., Crane, J., Martinez, F., ... & Stewart, A. W. (1995). International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. *European respiratory journal*, *8*(3), 483-491.

[22] Garfinkel, P. E., & Newman, A. (2001). The eating attitudes test: twenty-five years later. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *6*(1), 1-21.

[23] https://www.sralab.org/rehabilitation-measures/rosenberg-self-esteem-scale (last accessed May 20, 2022)

[24] Strand, B. H., Dalgard, O. S., Tambs, K., & Rognerud, M. (2003). Measuring the mental health status of the Norwegian population: a comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nordic journal of psychiatry*, *57*(2), 113-118.

[25] Alexopoulos, D. S., & Kalaitzidis, I. (2004). Psychometric properties of Eysenck personality questionnaire-revised (EPQ-R) short scale in Greece. *Personality and individual Differences*, *37*(6), 1205-1220.

[26] https://www.go-fair.org/fair-principles/ (last accessed May 20, 2022)

[27] https://ec.europa.eu/health/ehealth-digital-health-and-care/european-health-data-space_fr

[28] HUNT4 - The Nord-Trøndelag Health Study - NTNU. ntnu.eduhttps://www.ntnu.edu/hunt4 (accessed May 25, 2016).

[29] Sialm, A. and Bickenbach, J. Report on Stakeholders' Outcomes. SYNCHROS Project, 2022.

[30] Åsvold, Bjørn et al., Cohort Profile Update: The HUNT Study, Norway; International Journal of Epidemiology, 2022, 1–12 https://doi.org/10.1093/ije/dyac095

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

6. ANNEXES

Annex I: Core concepts of RBV

(1) Tangible vs. Intangible resources

Tangible resources have physical attributes and can be both observed and quantified. In cohort resources, examples of such resources may be existing infrastructure, database architecture and funding.

Intangible resources are not visible and cannot be quantified. In cohort research, examples of such resources may be reputation, academic relationships, informal communication channels and research culture.

(2) Resource Heterogeneity/ Resource Portfolios

A research institution is a bundle of productive resources (both tangible and intangible) and different research institutions possess different bundles of resources.

(3) Resource Immobility

Some resources are either very costly to copy or/and inelastic in demand. In cohort research, resources immobility may for instance, refer to the governance practices that may be hard to implement for other institutions, either because they are too costly or because the manpower is insufficient.

(4) Links between resources and capabilities in the implementation of cohort research

In RBV, both resources and capabilities generate the core competencies of a research institutions. In cohort research, core competencies may for instance, be identified to established "good practices". The core competencies of a research institution determine this institution activity and thus its ability to conduct cohort research.

(5) Resource Position Barriers

If the resource is held by a firm/research institution, then the cost and/or revenues of firms/research institutions acquiring this resource at a later point in time will be affected. In the context of cohort research, this means for instance, that research institution A acquiring a resource (e.g. an automated database architecture) at a later point that research institution B will be negatively affected in the storage and sharing of data, especially in relation to research institution B.

(6) Versatile vs. Specialized Resources

Resources are either versatile, applicable to a broad range of cohort research projects or specialized, applicable to a narrow range of cohort research projects. In an uncertain setting, versatile resources may not be more attractive than specialized ones.

(7) Barriers to entry

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

In strategic management terms, barriers to entry refer to factors that prevent newcomers to enter a certain market. In cohort research, barriers to entry refer to factors that prevent cohort research to unfold smoothly (e.g. too much taken by ethics committees to grant authorization).

(8) Resources Characteristics

There are four types of resources characteristics in a research institution. A resource can be determined in terms of its (1) value (2) rarity (3) imitability and (4) potential for substitution.

Value: The extent to which a resource allows a research institution to neutralize difficulties in research projects and exploit opportunities for cohort research.

Rarity: The extent to which a resource is unique to the research institution conducting the research.

Imitability: The extent to which the resource can be imitated by other research institutions or/and stakeholders.

Potential for Substitution: The extent to which the resource can be substituted by other, not rare and/or imitable, resources.

(9) Context of resource emergence/Origin of resources

In cohort research, resources do not come in a vacuum and are likely to be bound and determined by their contexts. Some dimensions are:

- Unique historical conditions (e.g. the GDPR and "old" databases)
- Causal ambiguity (e.g. harmonisation techniques have no clear reference because of a lack of documentation)
- Social Complexity (e.g. relationships with participants from underprivileged communities)
- Intellectual Property

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Annex II: Core concepts of KBV

(1) Combinative Capabilities

Refer to produce the same methods, practices and data and infrastructure in different ways. That is, materials, outputs and processes produced during a cohort research project are combined differently. Development is then defined by the carrying out of new combinations.

(2a) Integrative Capabilities

Integrative capabilities are needed when cohort research activities requires the coordinated efforts of individual specialists who possess many different types of knowledge.

(i) Transferability:
Explicit knowledge is revealed by its communication, tacit knowledge is revealed through its application. Transfer of tacit knowledge is slow, costly, and uncertain.

(ii) Aggregation
The efficiency with which knowledge can be transferred, depends on its aggregation

(iii) Appropriability
The ability of the owner of knowledge to receive a return equal to the value created by the knowledge

(2b) Integrative Capabilities: Knowledge Integration Mechanisms

(a) Rules and Directives:
Refer to plans, schedules, forecasts, rules, policies, procedures, standardized information and communication systems in cohort research. They regulate the application of knowledge and the collaboration between those who hold specialized knowledge.

(b) Sequencing
A simple process which allows to integrate knowledge while minimizing communication. It includes transformation steps ordered in a time-patterned sequence such that each specialist's input occur independently through being assigned a separate time slot

(c) Routines
Refer to complex patterns of behaviour and practices triggered by relatively small number of initiating signals or choices. When there is no rules, directives and communication in place, they can still support complex patterns of interactions among research team members. Routines allow various interactions to occur at the same time.

(d) Group Problem Solving and decision making:
Refer to isolated transformation steps when cohort research activities may require more personal and communication-intensive forms of integration. The need for group problem solving and decision-making increases with task complexity. Occurs for complex, unusual, and highly important transformation steps

(3) Protective Capabilities

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

In the context of cohort research, it refers to protective organizational arrangements on the one hand, and the protection of participants on the other (e.g. confidentiality and privacy measures). For research institutions, the issue is how much their generated knowledge and data can be protected both in terms of intellectual rights (and data use) and in terms of participants' safety (e.g. GDPR, anonymisation techniques).

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Annex III: Participants in the Stakeholder Consultation Process

Evidence-based synthesis and priority setting

- 93 stakeholders representing 76 projects or initiatives were contacted on several occasions for participation in the stakeholder consultation; candidates who sent their apologies were invited to suggest alternative candidates.
- Overall, 25 agreed to participate; 9 Principal Investigators of harmonization initiatives, 8 methodological experts and 8 ethical/legal experts. All sessions suffered from the sudden Covid-19 pandemic (May 2020) and many stakeholders apologized because they were busy with Covid-19 related research proposals or studies.
- Of the participating stakeholders, 40% was female and 60% male. Males were significantly overrepresented in the session on methodological challenges, whereas the females were considerably overrepresented in the session on ethical and legal challenges.
- Most participants were from the Netherlands (6) and Spain (6), followed by participants from the UK (3), Finland (2), Italy (2), Hungary (1), France (1), Germany (1), Luxembourg (1), Norway (1) and Sweden (1).
- Different health contexts were represented such as Parkinson's disease, oncology, infectious diseases, and cardiovascular diseases. Moreover, participants were involved in biobanks, infrastructures, European Open Science Cloud, and the International Science Council.
- The stakeholder consultations were complemented with interviews with key informants. Experts in the methodological matters (such as members of the CINECA consortium) or legal, ethical and practical matters (e.g. Prof. Madeleine Murtagh and members of the STANDS4PM consortium) have been interviewed to obtain more in-depth information on very specific topics.

Evidence-informed policy-making phase

- 36 participants were contacted for the session on the methodological issues and 33 for the session on ethical issues. Each non-replying candidate was contacted on three occasions. Candidates that sent their apologies were invited to suggest other potential candidates (i.e. snowballing approach).
- Overall, 27 stakeholders agreed to participate (16 and 11 for the methodological and ethical issues, respectively). Both sessions suffered from last minute drop-outs because participants and/or their family members resulted infected with the Omicron COVID-19 variant (5 and 3 cancelations for methods and ethics session, respectively).
- Male participants were overrepresented (63% v s. 37%). In the stakeholder dialogue session on methodological issues 73% was male and 27% female; nonetheless, in the stakeholder dialogue on ethical issues both genders were equally represented (50% and 50%, respectively).
- Most participants were from Belgium (21%) and Germany (21%), followed by Spain (11%), Switzerland (11%), Austria (11%), Czech Republic (5%), France (5%), Finland (5%), Greece (5%) and Italy (5%).
- The following organisations were represented during the stakeholder dialogues:
    - World Health Organisation
    - Research Infrastructure EurOPDX
    - Swiss Personalized Health Network
    - Center of Artificial Intelligence and Medicine - University of Bern
    - Research infrastructure on Population Health Information (PHIRI)

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

- Finnish Institute for Health and Welfare (THL)
- European Joint Program on Rare Diseases (EJP RD)
- European Reference Networks Support Infrastructure (ERICA)
- BBMRI-ERIC
- European Joint Program on Rare Diseases
- Institute of Computer Science of the Czech Academy of Sciences
- European Clinical Research Infrastructure Network (ECRIN)
- Ethics and Patient oriented Care in Oncology (NCT-EPOC)
- European Network of Research Ethics Committees
- Institute of Global Health of the University of Heidelberg
- Zika Virus Individual Participant Data Consortium
- Trilateral Research
- H2020 Project SHERPA
- Health and Digital Executive Agency (HaDEA)
- Bioinformatics Core Group of the University of Luxembourg

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884

Annex IV: SYNCHROS External Advisory Board

External Ethics Advisory Board

Ingrid Klingmann	Chairman European Forum for Good Clinical Practice (EFGCP)	Germany
Evert-Ben van Veen	Partner & Senior Consultant MedLawconsult	The Netherland
Ludwine Casteleyn	Department of Human Genetics – KU Leuven	Belgium

External Scientific Advisory Board

Isabel Fortier (Chair)	Director Maelstrom Research and DataSHaPER program – Research Institute of McGill University Health Centre (RI-MUHC)	Canada
Yaakov Stern	Chief Cognitive Neuroscience Division, Department of Neurology – Columbia University Medical Center	USA
Erik Steinfelder	Director of Biobanking and BioMolecular resources Research Infrastructure (BBMRI-ERIC)	Austria
Vincent Jaddoe	Director Generation R Study – Erasmus MC, University Medical Center Rotterdam	The Netherlands
Martine Vrijheid	Coordinator HELIX (Human Early Life Exposome) project – Instituto de Salud Global, Barcelona	Spain
Scott M. Hofer	Director Integrative Analysis of Longitudinal Studies on Aging (IALSA) – Department of Psychology, University of Victoria	Canada

The SYNCHROS project (www.synchros.eu) has received funding from the *European Union's Horizon 2020 Research and Innovation Programme* under Grant Agreement No. 825884