ASTAHG ALPINE SPACE TRANSNATIONAL GOVERNANCE ON ACTIVE AND HEALTHY AGEING

WP T2

REPORT ON AHA INNOVATION EVALUATION METRICS

D.T2.2.2

Vienna, July 2020

PARTNERS

Responsible partner for work package 2:
European Centre for Social Welfare Policy and Research (Austria)

Authors of this report:
Rahel Kahlert[1], Christian Ernst Heinrich Boehler[1]*, Annemarie Müllauer[2], Elisabeth Kapferer[2], Andreas Koch[2]

1. European Centre for Social Welfare Policy and Research (Austria)
2. Centre for Ethics and Poverty Research at University of Salzburg (Austria)

***** Corresponding author: boehler@euro.centre.org

Acknowledgements:
The authors would like to acknowledge the invaluable input from all ASTAHG project partners, observers and external contributors and they take responsibility for remaining errors in this report.

ASTAHG I Project Partners
FVG | Autonomous Region Friuli Venezia Giulia (Italy, Lead Partner)
AREA | Area Science Park (Italy)
PAT | Autonomous Province of Trento (Italy)
AULSS1 | Local Health Authority n.1 Dolomiti (Italy)
PLUS | Centre for Ethics and Poverty Research at University of Salzburg (Austria)
European Centre | European Centre for Social Welfare Policy and Research (Austria)
PSP PACA | Professional network of home care service providers in Provence-Alpes-Côte-d'Azur (France)
NIJZ | National Institute of Public Health (Slovenia)
GINA | Geneva International Network on Ageing (Switzerland)

Funding
Total eligible costs: 1.078.659 EUR
ERDF grant: 916.860 EUR

Project duration
Start date: 2018-04-17
Closure date: 2021-04-16

More information on the project:
https://www.alpine-space.eu/projects/astahg

Content

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

List of Figures

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

List of Tables

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

LIST OF ACRONYMS

AAI	Active Ageing Index
AAF	Active Ageing Framework
AFC	Age Friendly City Framework
AFCI	Age Friendly City Indicators
AFEE	Age-friendly environments in Europe
AHA	Active and Healthy Ageing
AS	Alpine Space
ASTAHG	Alpine Space Transnational Governance of Active and Healthy Ageing
DAC	Development Assistance Committee
EUSALP	EU Strategy for the Alpine Region
GAWI	Global AgeWatch Index
HTA	Health Technology Assessment
ICT	Information and Communication Technology
MCDA	Multicriteria Decision Analysis
MoU	Memorandum of Understanding
OECD	Organisation for Economic Co-operation and Development
SoSe	Sozialplanung für Senioren (Social planning for older persons)
TGB	Transnational Governance Board
TOC	Theory of Change
UNDP	United Nations Development Programme
UNECE	United Nations Economic Commission for Europe
WHO	World Health Organization
WP	Work package

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

1 INTRODUCTION

Demographic change constitutes a major societal challenge in most industrialised countries that requires combined efforts from different stakeholders, including public authorities, industry, academia and civil society across policy areas to support Active and Healthy Ageing (AHA) (e.g. Rechel et. al., 2013; WHO, 2002; 2013). This challenge is amplified in the Alpine Space (AS) region by its distinctive characteristics, including considerable regional variation both in demographic change and population growth projections, ultimately calling for tailored interventions to foster Active and Healthy Ageing. In addition to that, the AS area is composed of regions that belong to different countries which, thus far, has limited the scope for trans-regional and transnational cooperation to tackle the ageing challenge. Further, AHA policies are often restricted to a few areas of public service provision, such as healthcare and welfare authorities. Potential synergies from cooperation across sectors, for instance, cultural, economic or housing policies, are thus often neglected (WHO, 2012; 2013; 2017; OECD, 2015).

1.1 THE ASTAHG PROJECT AT A GLANCE

The Alpine Space Transnational Governance of Active and Healthy Ageing (ASTAHG) project aims to tackle this challenge by following a *multisectoral, transnational, and multilevel* approach to improve AHA in the AS. It is *multisectoral* as it aims to facilitate innovation across sectors, such as social care, healthcare, long term care, independent living, mobility and transport, as well as culture and tourism; and it follows a *transnational* approach as it brings together stakeholders from different regions of the AS to exchange experiences, ideas and innovations, streamline strategies to address the ageing challenge and to share knowledge and best practices across geographically and/or politically defined contexts. The project's *multilevel* approach aims at

cooperation between stakeholders on local, regional, and national level to identify, implement, evaluate and improve upon successful AHA policies and to harvest potential synergies through efficient cooperation along all stages of the policy cycle.

The overall objective of the project is to improve capacities and coordinating efforts in support of AHA between sectors and different levels, and to respond with tailored initiatives to AS territorial needs. It aspires to enhance governance capacities related to regional AHA policies, foster the transfer of innovation for AHA in the AS, and to develop a social innovation framework for generating and adopting innovative solutions for AHA involving both public and private actors (ASTAHG, 2018). To achieve these objectives, ASTAHG will establish a *Transnational Governance Board* (TGB) for AHA to bring policy makers and other stakeholders in the AS together, to develop a network, and to foster the exchange of successful AHA policies, initiatives and innovations. The TGB is defined as '*an open network and the participation of members is free of charge and voluntarily*' (MoU, 2019). Whilst all ASTAHG partners are founding members of the TGB (Managing Committee), other interested organisations and stakeholders may apply to join at any time. (MoU, 2019). The TGBs main objective is '*to promote an 'age-friendly' Alpine Space Area creating synergies between interested stakeholders and governance levels and helping the Alpine Space local, regional and national authorities and other stakeholders to collaborate in promoting innovative solutions that address the needs of the ageing population'* (MoU, 2019).

To this end, ASTAHG will also develop a portfolio of good practices in AHA governance and establish an AHA innovation observatory which classifies AHA initiatives and solutions with context and efficiency indicators (ASTAHG, 2018). A framework for AHA innovation based on the Quadruple Helix model (Carayannis & Campbell, 2009) will foster collaboration between different actors from local, regional and national

governance, industry, as well as academia and civil society (ASTAHG, 2018). ASTAHG will also align its efforts and results with the EU Strategy for the Alpine Region (EUSALP) so to further enhance the level of transnational governance throughout the AS.

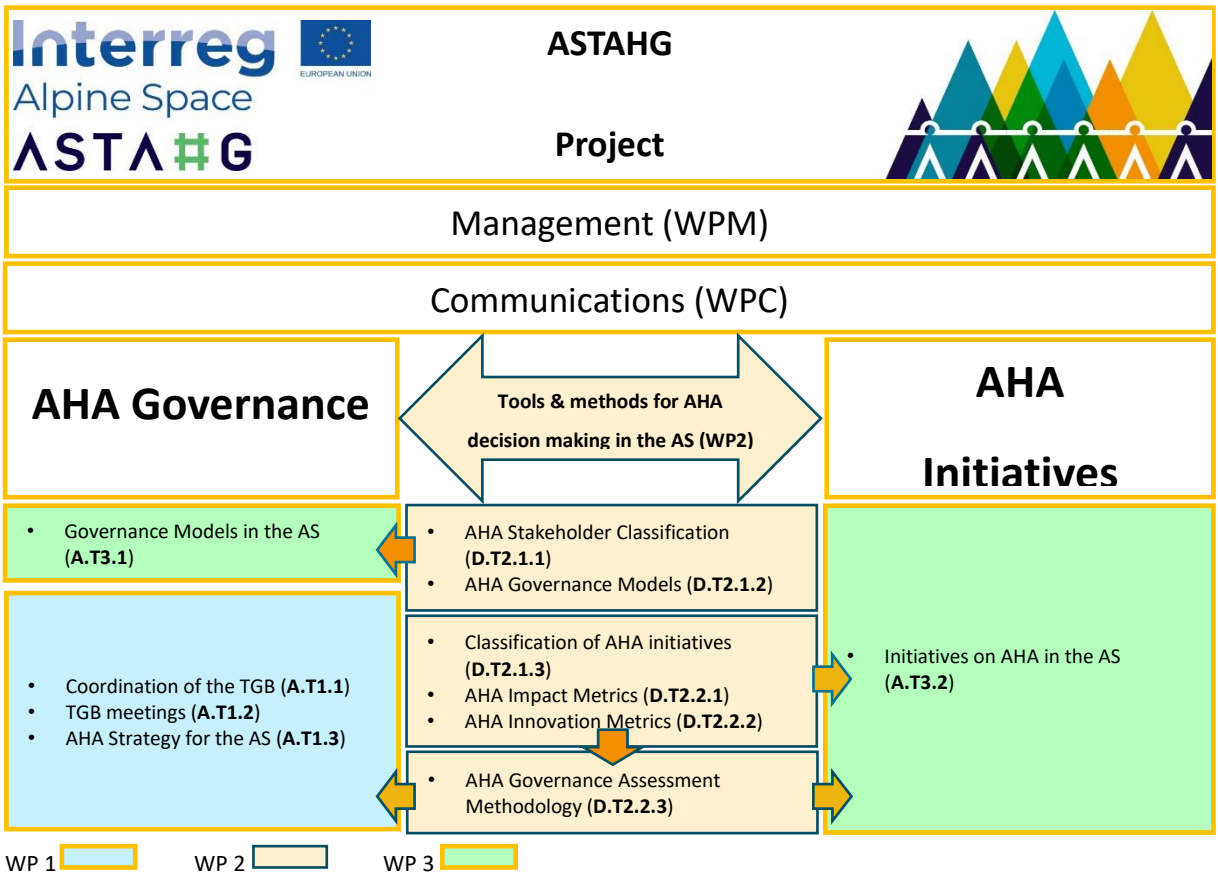
The ASTAHG project has been designed in several Work Packages (WPs), each of which contributes towards the common aim and objectives (Figure 1). Horizontal activities are concentrated in WPM (Management) and WPC (Communication). Whilst WPM is concerned with overall project management and ensures sound and smooth project implementation, internal communication between partners and with the funding organisation, WPC is dedicated to the development and execution of an efficient communication strategy, engagement with Quadruple-Helix actors in the TGB; exchange with other AHA initiatives, in particular EUSALP; dissemination of project outcomes as well as engagement with AHA stakeholders and a wider public audience.

WPs 1 to 3 are concerned with project implementation. In this context, WP1 aims to establish and manage the TGB that will be composed of public and private actors, pertaining to different levels (regional/local) and sectors as well as representing AS territorial characteristics (ASTAHG, 2018). The TGB is organised in different thematic groups and meets regularly in order to share experiences, knowledge and expertise and to develop a sustainable AHA strategy for the AS based on intersectoral, transnational and multilevel cooperation. The activities in WP1 range from the coordination of the TGB (A.T1.1) to the organisation of regular TGB meetings (A.T1.2) and to develop an AHA strategy for the AS (A.T1.3).

WP2 develops and provides tools and methods for the project, in particular a classification of AHA stakeholders (D.T2.1.1), a model for AHA governance in the AS (D.T2.1.2), a classification of AHA initiatives (D.T2.1.3), as well as AHA impact

evaluation metrics (D.T2.2.1), AHA innovation evaluation metrics (D.T2.2.2, this report) and an AHA governance assessment methodology (D.T2.2.3). WP3 is concerned with the application and use of tools and methods developed in WP2: data gathering and analysis of AHA governance models (A.T3.1) and the identification and monitoring of innovation in AHA in the AS (A.T3.2).

Figure 1: Components of the ASTAHG project and WP2 in context



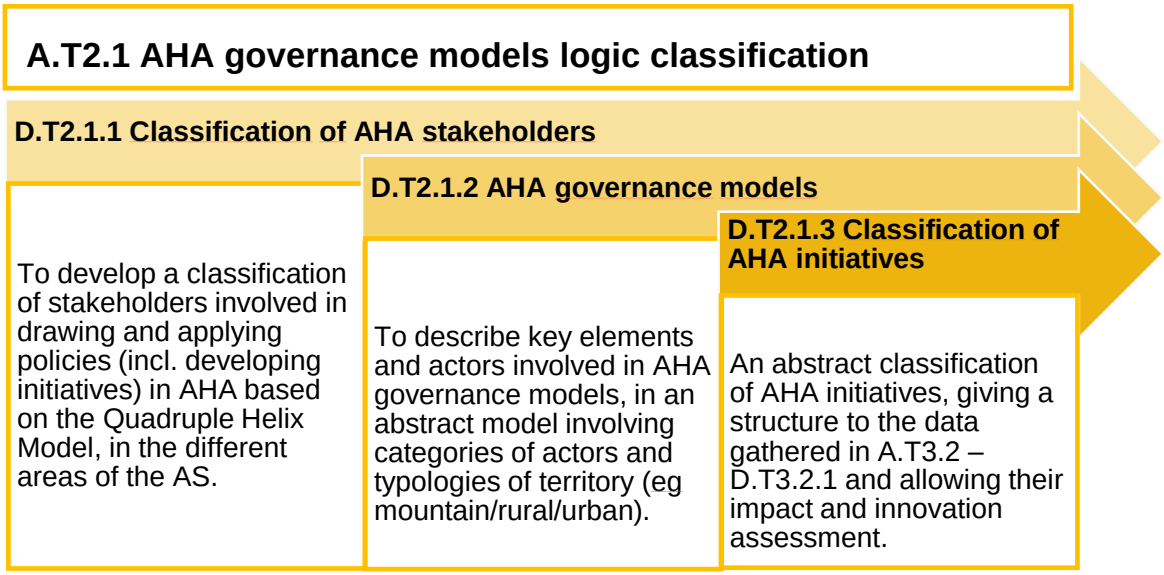
Source: Own drawing based on ASTAHG (2018).

1.2 CONTRIBUTION OF WORK PACKAGE 2

As depicted in Figure 1 above, the overall aim of WP2 is to provide tools and methods for the ASTAHG project to bridge the gap between AHA governance and AHA initiatives and to enable efficient AHA decision making in the AS. WP2 thereby aims at supporting activities both in the context of implementing a Transnational Governance Board (WP1) as well as activities in WP3, which will gather data and information on AHA initiatives and governance models in the AS. Whilst deliverables D.T2.1.1 (AHA stakeholder classification) and D.T2.1.2 (AHA governance models) play a particular important role in the conceptualisation, design, and composition of the TGB by contributing both theoretical models and structuring the space of relevant stakeholders in accordance with the Quadruple Helix Model (Carayannis & Campbell, 2009), they also provide tools for WP3 to collect context specific data on relevant AHA actors and governance models prevalent in the AS region. Deliverable D.T2.1.3 (classification of AHA initiatives), on the other hand, is more concerned with developing a tool to gather information on policies, initiatives and innovations which aim at improving Active and Healthy Ageing in the AS. This tool will, in turn, provide a framework for WP3 to collect and analyse relevant information from each project region, and help structuring the evidence on cross-sectorial AHA policies, initiatives, and innovations which may have the potential to:

- support AHA of the population in the respective project regions
- improve the sustainability of social, health and care systems, as well as other areas of public service provision, and
- contribute towards the competitiveness of local economies by encouraging innovation for AHA in the AS.

Figure 2: Deliverables in Activity T2.1 - AHA governance logic classification



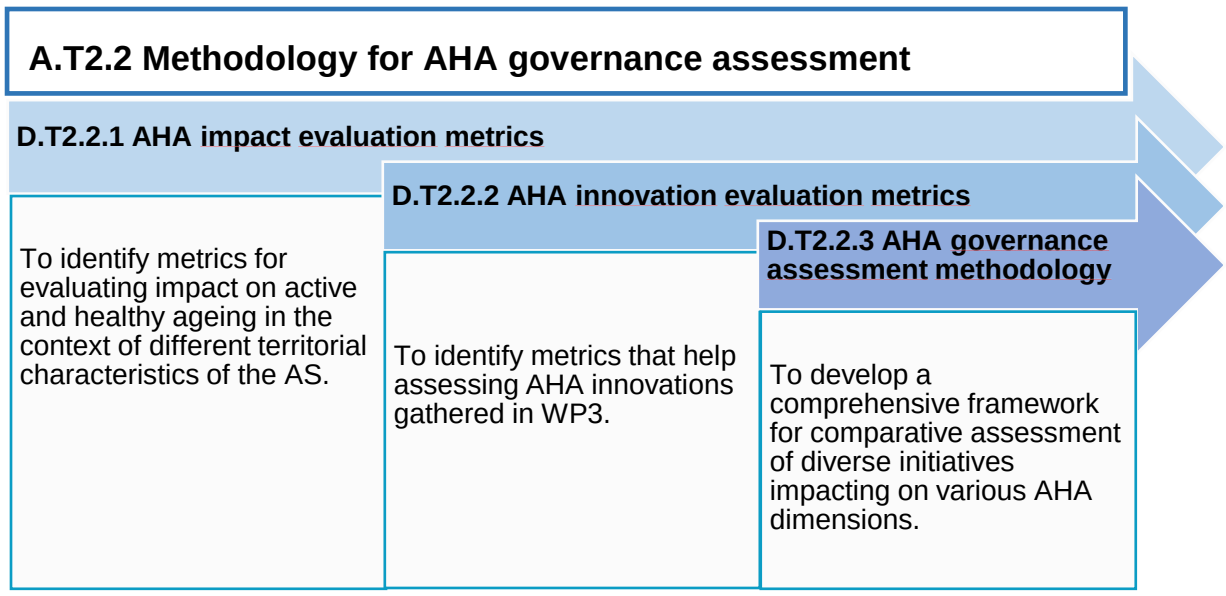
Source: Own drawing based on ASTAHG (2018).

Activities in A.T2.2 (Methodology for AHA governance assessment, Figure 3), are concerned with developing tools and methods for efficient cross-sectorial AHA decision making in the AS. In this context, Deliverable D.T2.2.1 (AHA impact evaluation metrics) gathers indicators that may help quantifying the impact of AHA policies, initiatives and innovations on various dimensions of AHA with the aim to support decision makers identifying promising AHA interventions in their respective contexts. To better understand the innovative character of AHA policies, initiatives and innovations, deliverable D.T2.2.2 (this report) further proposes how to identify innovation evaluation metrics from the long-list of indicators gathered in Deliverable D.T2.2.1, whilst both deliverables ultimately feed into the development of an AHA governance assessment methodology (deliverable D.T2.2.3). The latter is based on the concept of multicriteria decision analysis (MCDA) and will help decision makers in prioritising amongst policy alternatives that may all lead to various favourable effects across relevant sectors but generally compete for limited resources. The three deliverables also form the basis for data collection and analysis in WP3, with the

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

ultimate aim to identify and monitor innovation in AHA in the AS through the development of an AHA innovation observatory.

Figure 3: Deliverables in Activity T2.2 - Methodology for AHA governance assessment



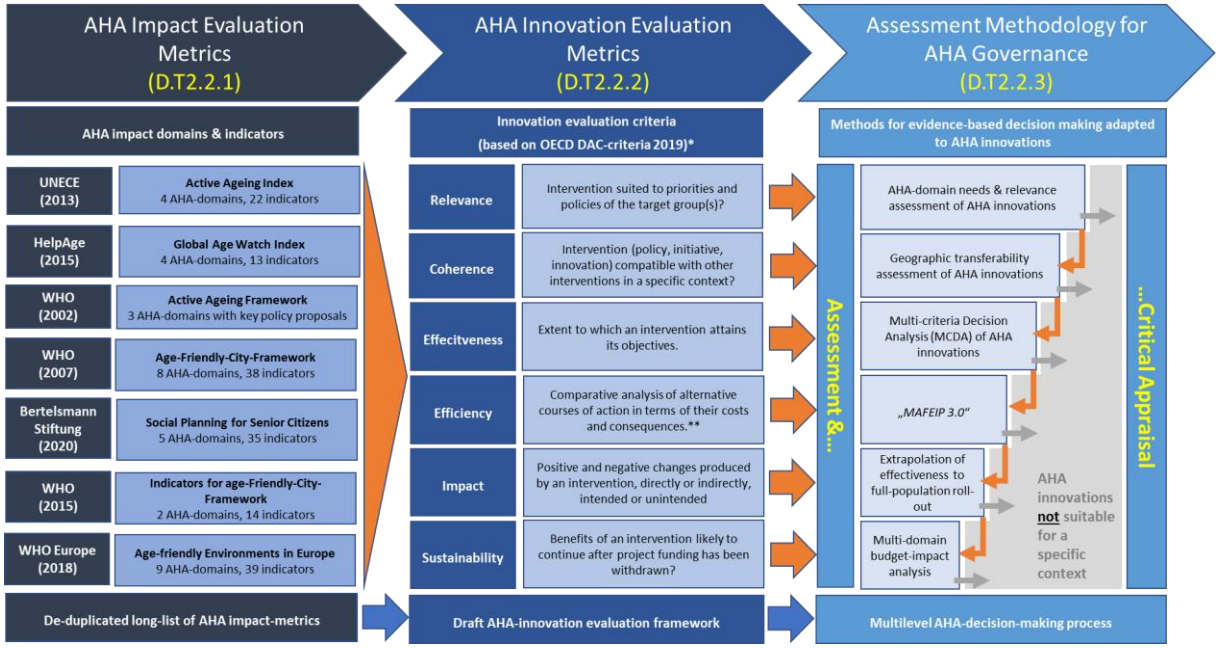
Source: Own drawing based on ASTAHG (2018).

1.3 AIM AND STRUCTURE OF THIS REPORT

Whilst evidence based decision making is a well-established process in some areas of public policy making (such as healthcare), there is still ambiguity as to how to prioritise innovations which are competing for limited public resources across traditional silos of governance, especially when innovations are likely to yield diverse (and sometimes perhaps even conflicting) outcomes. Nevertheless, it is important for AHA-stakeholders to engage in a transparent process so to identify innovations that provide not just good value for money but are also tailored to the needs and preferences of

the population in their respective target settings. Some attempts have been made in order to support decision-makers in allocating scarce resources towards cost-effective AHA innovations, most notably with the development of the Monitoring and Assessment Framework for the European Innovation Partnership on Active and Healthy Ageing (www.MAFEIP.eu), developed by the European Commissions' Joint Research Centre, Institute for Prospective Technological Studies (Boehler et al., 2015, Boehler & Abadie, 2015, Boehler & Abadie 2016). However, as of today, there is ambiguity as to how to identify the most beneficial technologies from a diverse set of AHA innovations, and a lack of understanding on how to embed evaluation and critical appraisal methods within a structured and transparent process of multi-stakeholder AHA governance.

Figure 4: Relationship between deliverables D.T2.2.1, D.T2.2.2 and D.T2.2.3



*Sources: Own drawing based on OECD 2002 & OECD 2019. **Drummond et al., 2005.*

This report (D.T2.2.2) summarizes the work carried out to gather and classify AHA innovation evaluation metrics from the long list of indicators identified in Deliverable

D.T2.2.1 for assessing how various cross-sectorial AHA policies, initiatives and innovations may be suited:

- to address the needs of their respective target populations (**relevance**);
- to fit into the target context, with existing policies, initiatives and / or innovations (**coherence**);
- to achieve their objectives, and to which extend they do so (**effectiveness**),
- to provide good value for money, i.e. how they compare to existing interventions in terms of their cost and consequences (**efficiency**),
- to assess the greater impact of cross-sectorial AHA innovations (**impact**), and
- to evaluate cross-sectorial financial impact (**sustainability**)

Both deliverables (D.T2.2.1 and D.T2.2.2) feed into the development of an AHA governance assessment methodology (D.T2.2.3), which, based on the principles of MCDA, aims to guide AHA decision makers in prioritising activities, which transcend traditional silos of public policy making. The relationship between deliverables D.T2.2.1 to D.T2.2.3 is depicted in Figure 4. Hence, this report builds a bridge between Deliverable D.T2.2.1 and Deliverable D.T2.2.3, by linking the proposed indicator long list (D.T2.2.1) with frameworks, theories of change and quality criteria to choose appropriate indicators, resulting in a comprehensive approach to AHA decision making, as laid out in the AHA governance assessment methodology (D.T2.2.3).

Chapter 2 introduces OECD DAC evaluation criteria (OECD 1992; OECD, 2002; OECD 2019) and thereby provides a conceptual framework for AHA innovation assessment, which is at the core of Deliverable D.T2.2.3. Based on evaluation criteria such as relevance, coherence, effectiveness, efficiency, impact and sustainability, we construct a governance assessment methodology for AHA innovation assessment that can be further adapted to local settings.

Chapter 3 introduces theory-of-change modelling (TOC) of an innovative intervention as context for indicator development and selection. It aims to provide local AHA stakeholders with a theoretical framework for choosing, from the long-list of indicators presented in Deliverable D.T2.2.1, those that are particularly relevant for assessment purposes in their respective contexts.

Chapter 4 proposes quality criteria for indicators so to prioritise metrics for AHA innovation evaluation. It provides further guidance on indicator selection by explaining desirable indicator properties for innovation evaluation, such as validity, reliability, timeliness, sensitivity to change etc. As part of the AHA impact evaluation metrics, we place special attention on quality criteria for developing AHA indicators and / or generic indicators that may be useful in the context of cross-sectorial decision-making.

The final chapter provides a conclusion and next steps for using AHA innovation evaluation metrics and its further development in the context of the ASTAHG project and the future work of the TGB.

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

2 OECD DAC EVALUATION CRITERIA

Report D.T2.2.1 screened available tools and methods for assessing the status quo and progress of Active and Health Ageing. Based on a pragmatic review of the literature, the following tools and frameworks were identified:

- Active Ageing Framework (AAF) by WHO (2002);
- Active Ageing Index (AAI) by UNECE (2013; Zaidi et al, 2013);
- Global AgeWatch Index (GAWI) by HelpAge International (Zaidi, 2013; Mihnovits & Zaidi, 2015);
- Age Friendly City Framework (AFC) by WHO (2007);
- Indicators for Age-friendly City Index (AFCI) by WHO (2015);
- Age-friendly Environments in Europe (AFEE) by WHO Europe (2018) and
- the German tool Social Planning for Senior Citizens (SoSe) by Bertelsmann Stiftung (2020)[1].

Based on this review, the report suggested:

- **AHA domains** (demographic and social structural data; civic engagement and social participation; mobility and transport; communication, information and ICT; housing, outdoor spaces and enabling environment; health and care; security and safety),
- Broadly aligned with **ASTAHG-sectors** (social care, health care, long term care, independent living, wellbeing, culture & tourism, mobility & transport); and

[1] Sozialplanung für Senioren. Planungshilfe für ein seniorenpolitisches Konzept. https://www.sozialplanung-senioren.de/die-themenfelder/ii-sose-planungshilfe-fuer-ein-seniorenpolitisches-konzept/index.html.

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

- a set of measurable **core indicators** (57 indicators and 122 variables) and **supplementary indicators** for qualitative data collection and further development.

Core and supplementary indicators reported in D.T2.2.1 provide a long list of potential indicators for the ASTAHG governance assessment methodology (Deliverable D.T2.2.3).

This chapter draws on this long list of indicators and links them to the evaluation criteria laid out in the OECD Development Assistance Committee framework (DAC): relevance, coherence, efficiency and effectiveness, impact and sustainability. The OECD DAC evaluation criteria constitute a widely adopted reference framework for evaluating projects, programmes, and public policies including innovation technologies around the world since they were developed in 1991. Hence, vast experience accumulated about the implications and limitations of applying the OECD DAC criteria in evaluation in practice. Accordingly, several academic articles suggested revisions, based on the identified gaps, assessment of compliance and room for improvement (Chianca, 2008; Forss & Bandstein, 2008). Based on this literature, in 2019, the OECD revised the definitions and principles of the existing criteria (OECD, 2019) in order to make definitions clearer and more aligned with policy priorities. Furthermore, they specified the context and intended purpose of the criteria, and also further explained the dimensions of each the criteria and how they apply to different evaluations. Considering that the OECD DAC criteria present a very important and relevant evaluation framework, these criteria form the basis for this report, and also provide a conceptual framework for the ASTAHG Governance Assessment Methodology reported in Deliverable D.T2.2.3.

The general purpose of the OECD DAC evaluation criteria is to "*enable the determination of the merit, worth or significance of an intervention*" (OECD, 2019). An "intervention", in this context, is a term broadly used by the OECD referring to the subject of the evaluation, which encompasses all kinds of projects, programmes, policies, strategies, instruments, funding mechanisms etc. The OECD considers that criteria play a normative role as they describe the desired attributes of interventions and support accountability and monitoring results (OECD, 2019). Policy interventions can be regulatory (e.g. self-regulation, quasi-regulation, government legislation etc.), or non-regulatory (e.g. information disclosure).

Two main principles guide the use of the criteria (OECD, 2019):

1. The criteria need to be **guided by the evaluation questions and be contextualised**—understood in "*the context of each individual evaluation, the intervention being evaluated, and the stakeholders involved*". As such they are intended to support high-quality, useful evaluations.
2. The criteria need to consider the **purpose of the evaluation and be applied according to the context** of the evaluation **including stakeholder needs**. Issues such as data availability, resource constraints, timing and methodological considerations may also influence how a particular criterion is covered

For the purposes of the ASTAHG governance assessment methodology (D.T2.2.3), we consider the OECD DAC evaluation criteria as a stepwise process through which the space of potential AHA innovations funnels through (Figure 5). This follows the thought that:

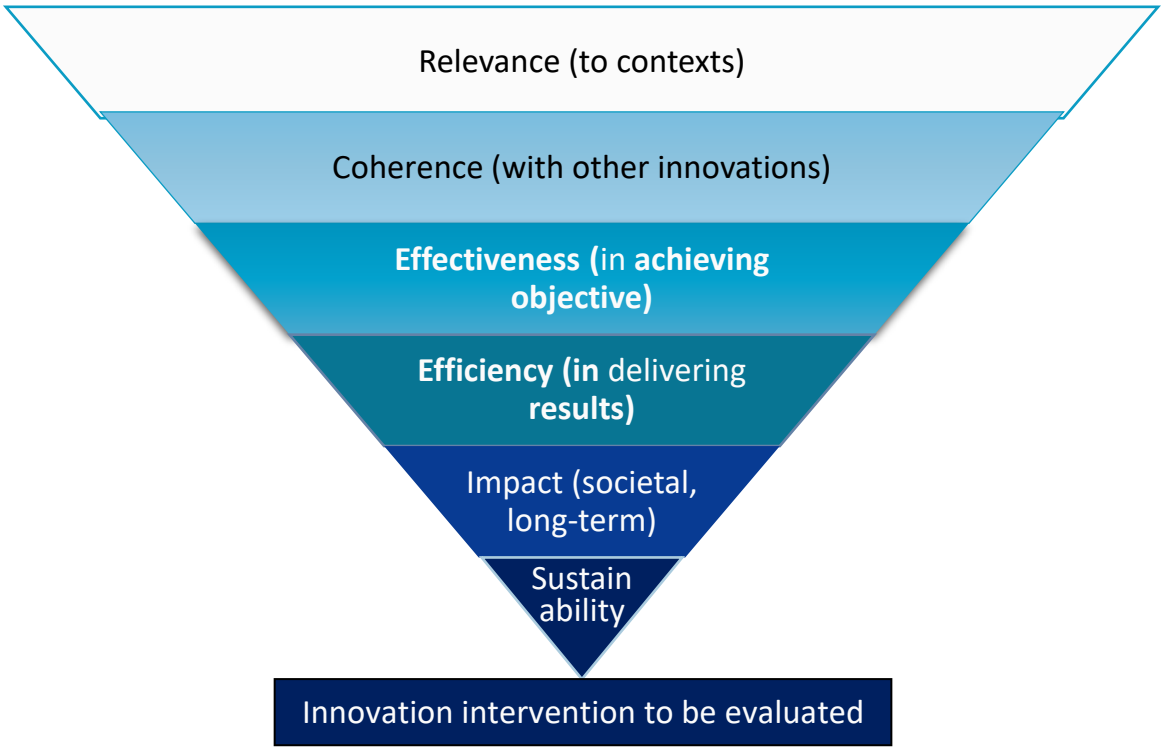
- one should first assess the most fundamental questions when assessing innovations for a particular context, such as whether it may actually have the potential to serve population needs and preferences, and whether it fits within the target context

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

- the evaluation criteria become increasingly difficult and resource intensive to address, so that innovations not passing through an earlier step should not absorb analytic resources at a later stage of the assessment process.

Report D.T2.2.3 illustrates this process based on the OECD DAC evaluation criteria adapted to AHA decision making in much more detail, and links it with multi criteria decision analytic methods to be applied at different stages of the resulting assessment framework. The assessment process starts with relevance and coherence, followed by effectiveness and efficiency, and finally arriving at impact and sustainability (Figure 5) The remainder of this section describes these six criteria in terms of their definitions and purpose, and briefly discusses how they provide a useful and transparent structure for cross-sectoral AHA decision making.

Figure 5: Funnel of evaluation criteria



Sources: Own drawing based on OECD 2002 & OECD 2019.

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

2.1 THE EVALUATION CRITERION "RELEVANCE"

The evaluation criterion 'relevance' is concerned with contextual factors as they relate to the target group of beneficiaries. According to OECD DAC, *relevance* means:

> *"The extent to which the intervention objectives and design respond to beneficiaries', global, country, and partner/institution needs, policies and priorities, and continue to do so if circumstances change."*

According to the OECD, relevance relates to the degree to which the intervention is sensitive to the economic, environmental, equity, social, political economy, and capacity conditions. Please note that relevance also refers to national, regional and local government, civil society organizations, private entities and international bodies. The overall relevance question:

> *Is the intervention doing the right thing to address the needs and preferences of a target population in a particular context?*

Relevance therefore refers, among others, to contextual indicators about demographics and socio-structural data. For example, relevance-related indicators are current population according to age groups, family status, and household size (see, Bertelsmann's social planning indicators). The target group (e.g., persons above 65 years of age or other) and their needs and priorities are the central focus of this criterion, but note that a life course-perspective is shifting increasingly into the focus of AHA decision-makers.

Relevance is also concerned with the determinants of AHA, which includes for example the participation of older persons in social, economic, cultural, spiritual, and civic affairs, as laid out by the WHO Active Ageing Framework (cf., WHO, 2002). This would

include for example educational attainment (AAI, AFEE), Lifelong learning (AFEE, AAI) and positive social attitude toward older people (AFCI, AFEE). However, the needs and preferences of the target population may generally span all indicator domains, such as health and care; security and safety; or mobility and transport etc. It is important to note, in this context, that certain indicators cannot just be linked to the relevance criterion per se, but they may also be important for other evaluation criteria. For example, lifelong learning may not only be part of the intervention context, but also resemble a positive effect of an innovative intervention. Life expectancy may not only express a particular population need in the context of socio-structural data selection, but also be used to assess innovations' effectiveness.

In conclusion, in order to assess innovations' relevance, we need to understand what it is that the target population in a particular setting *'needs'* and *'wants'*. Deliverable D.T2.2.3 will further address this issue and embed the AHA innovation relevance assessment within the methodological framework of MCDA, where population needs and preferences are expressed as quantitative weights between a set of relevant indicators for AHA innovation assessment.

2.2 THE EVALUATION CRITERION "COHERENCE"

The evaluation criterion of coherence is generally concerned with how the innovative intervention compares to other interventions in a sector, across sectors, within a country and across countries such as the Alpine Region. According to OECD DAC, *coherence* means:

> *"The compatibility of the intervention with other interventions in a country, sector or institution."*

Coherence refers to both internal and external coherence: On one hand, it addresses synergies and inter-linkages between the intervention and other interventions carried out by the same government, or institution (including legally binding international norms). On the other hand, external consistency refers of the intervention of other actors in the same context. The aim of the criterion of coherence is to assess whether the intervention brings added value without duplicating existing efforts.

The OECD argues that the need of including this criterion is to raise awareness that the lack of coherence can lead to duplication of efforts and undermine overall progress. Interventions should be evaluated in the broader context and not only from an intervention-, or institution-centric perspective.

The overall coherence question is:

How well does the intervention fit?

For the purposes of ASTAHG, however, we take a broader perspective on innovations' coherence, transcending beyond the question of whether the innovation sits well within the existing mix of AHA policies, initiatives and innovations in a particular setting. More precisely, to answer the questions of how well the innovation fits within a certain context, we also need to understand the wider ecosystem to which the innovation should transfer into. This includes a detailed assessment of the maturity (i.e. "readiness") of a particular context. Likewise, we need to understand factors that determine the cost and outcomes of an innovation in a particular setting, and whether these factors constitute barriers to the geographic transfer of this innovation. Deliverable D.T2.2.3 will elaborate further on this issue.

2.3 THE EVALUATION CRITERION "EFFECIVENESS"

The evaluation criterion of effectiveness is concerned with outcomes as they relate to the individual or aggregate level of beneficiaries. According to OECD DAC definition, *effectiveness* means:

> "*The extent to which the intervention achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups*."

For several decades, efforts have been made to improve the effectiveness of interventions in international development across the world. In 2005, international organisations (e.g. United Nations, World Bank, OECD) and over 150 countries adopted the *Paris Declaration on Aid Effectiveness*, outlining five key principles: Ownership, alignment, harmonization, results, mutual accountability. These principles therefore contribute to measuring effectiveness, also in the context of policy initiatives and programmes.

The overall effectiveness question is:

> *Is the intervention achieving its objectives, and to which degree does it do so?*

Hence, effectiveness refers to the outcomes of AHA innovations. For instance, the Global AgeWatch Index (GAWI) provides a multidimensional index for quality of life

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

and wellbeing of older people. Going in a similar direction, the Bertelsmann's social planning indicator areas suggests indicators such as:[2]

- Percentage of persons with good functional health[3]
- Percentage of older persons with long-term care needs
- Percentage of older persons with (physical and cognitive) disabilities
- Quality of Life (perceived level of Quality of Life) of older persons (e.g., using standardized, validated questionnaires such as EQ-5D or WHOQOL-OLD).

Comparing indicators from Report D.T2.2.1, quality of life is a likely effectiveness indicator (see, AFEE, AFCI) for AHA frameworks. Social connectedness indicators (e.g., whether persons meet and can rely on other persons, see AAI, AFEE, GAWI), engagement in socio-cultural activity (e.g., AFCI, AFEE) could also serve as effectiveness indicators. Please note that some of these indicators could also fall under the relevance criterion, depending on the context or intended outcome of the intervention.

2.4 THE EVALUATION CRITERION "EFFICIENCY"

The evaluation criterion of efficiency is concerned with outputs or outcomes as they are put into relation to all inputs including costs of beneficiaries and of public goods. According to OECD DAC, *efficiency* means:

[2] Sozialplanung für Senioren. Planungshilfe für ein seniorenpolitisches Konzept. Online available (30.06.2020): https://www.sozialplanung-senioren.de/die-themenfelder/indikatoren-themenfeld-5-gesundheit-und-pflege/51-gesundheitsfoerderung-und-praevention/index.html

[3] E.g. German Survey of Old Age

"The extent to which the intervention delivers, or it is likely to deliver, results in an economic and timely way."

Here, "economic" means the cost-effective conversion of inputs related to not only outputs, but also to outcomes and impacts as part of the entire results chain (see also the OECD DAC glossary, 2002; p. 21 on efficiency, referring to results). Inputs could be funds, expertise, time and other resources and are usually measured in monetary units.

The overall efficiency question is:

How well are resources being used?

In this context, we distinguish between *'technical efficiency'* and *'allocative efficiency'*. Technical efficiency refers to the question of whether a certain outcome is achieved with the minimal amount of inputs. Allocative efficiency, on the other hand, refers to the question of whether the existing mix of AHA policies, initiatives and innovations provides the best value for money in a given context. Hence, in order to prioritise AHA-innovations which typically span across traditional silos of policy making, an assessment framework needs to address both technical and allocative efficiency. Addressing allocative efficiency, however, puts the choice of appropriate indicators to measure innovations' effectiveness back into the focus, as respective indicators must be comparable across innovations. In addition, decision makers require, at least in principle, an understanding of the shadow-price of displaced interventions in order to base decisions on the cost-effectiveness criterion, but this is particularly challenging in the context of cross-sectoral decision making. Deliverable D.T2.2.3 elaborates further on these issues.

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

2.5 THE EVALUATION CRITERION "IMPACT"

The evaluation criterion of impact relates to the aggregate level of beneficiaries. Long-term societal and regional results are reflected in the impact criterion. According to OECD DAC, *impact* means:

> *"the extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects"*

Impact refers to the social, environmental and economic effects, which are considered broader than those covered by the effectiveness criterion. Furthermore, an impact evaluation examines the indirect, secondary and potential consequences of an intervention.

The overall impact question is:

> *What difference does the intervention make, which side effects or external effects may occur, and what are the societal consequences of the intervention in the longer run?*

First and foremost, as indicators for assessing innovations' effectiveness must be identified and agreed upon, so must dimensions of impact. The key questions are what constitute ***'higher level effects'*** of an activity in the context of AHA, and for whom. A convenient starting point for commonly agreed criteria in the realm of AHA is the EIP on AHA's headline target (increase of healthy life expectancy by 2 years until 2020) and its triple-win, i.e. improved quality of life, sustainability of health and care systems, as well as innovation and growth. Though this would certainly be in line with many of the

ASTAHG objectives, there may also be other important components of innovations' impact for the purposes of ASTAHG (and perhaps beyond). Also, in addition to those metrics identified in D.T2.2.1, other important impact dimensions should perhaps be considered, such as distributional effects and equity (e.g. between genders, age groups, socio-economic subgroups, etc.), impact on ageism and stereotypes against older people, or moving towards and an age-inclusive society, amongst others (cf. WHO, 2015). Another sample impact indicator from Report D.T2.2.1 is the "ability to age in place" (AFEE).

2.6 THE EVALUATION CRITERION "SUSTAINABILITY"

The evaluation criterion of sustainability is concerned which whether intervention results last beyond project completion. This could also include whether interventions were taken up by other entities and distributed to a wider audience. According to OECD DAC, *sustainability* means:

"The extent to which the net benefits of the intervention continue, or are likely to continue"

This criterion addresses the financial, economic, social, environmental and institutional capacities needed to sustain intervention effects over time. It asks about the likelihood that these effects will continue over the medium and long term.

The overall sustainability question is:

Will the benefits last?

This is linked to the question about what systems need to be in place to ensure the sustainability of the intervention. Deliverable D.T2.2.3 elaborates further on the sustainability criterion and links the methodological framework of MCDA for cross-sectoral budget allocation for AHA innovations.

3 THEORY-OF-CHANGE MODELS

To establish innovation evaluation metrics for the purposes of the Governance Assessment Methodology (D.T2.2.3), it is important to obtain an understanding of what the innovative intervention under assessment is actually about. **Theory of change (TOC)** models can help in this context. The idea is also referred to as logical framework approach, logic model or result chain (cf. Frechtling, 2007). In the following section, the elements of theory-of-change models are described and how they are related to the evaluation criteria. Subsequently, we discuss how to develop such theories of change.

3.1 ELEMENTS OF THEORY-OF-CHANGE MODELS

The underlying idea of a TOC is to model an intervention logic from programme inputs to program impact. Indicators are considered an important part of such a model because they measure whether the elements of an intervention are being achieved. The model makes explicit the assumptions of an intervention by identifying the main channels through which an innovative intervention is expected to move from inputs to impact (Feinstein & Beck, 2006).

One could think about the idea behind a TOC like being a strategic game with multiple players—that is, as a game master it would be important to be able to see beyond the immediate actions, to project consequences, and ultimately to project impacts. In general, the input-output link is the strongest as it is the easiest to measure, but the

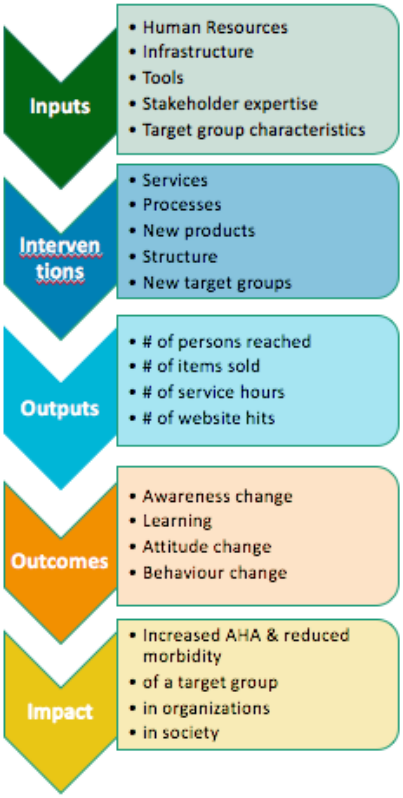
This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

link tends to become weaker when it gets to outcomes and impact. There is a danger that outcomes and impacts are then excluded from the results chain (see Kahlert 2013).

The TOC model is a systematic, visual way to present the assumed relationships among resources (inputs), the planned intervention (activity), and expected results. The goal is to create a results chain, based on explicit assumptions to be tested in the evaluation. Although the model shows a linear flow from inputs to impact, feedback loops exist. This means that outcomes can influence resources and vice versa.

- **Inputs** refer to indicators that measure the resources that go into the intervention (e.g. quantity and qualification of personnel, materials, funds, beneficiary characteristics). Different types of cost may be used in an economic evaluation, also depending on the perspective the evaluation takes, e.g. perspective of the provider, the healthcare system, or the society (e.g. Drummond et al., 2005).
- **Interventions** refer to the various projects, programmes and products that are carried out.
- **Outputs** refer to indicators that measure the immediate outputs directly resulting from the activities (e.g. number and rates of participants, number and duration of services).
- **Outcomes** refer to programme effects with respect to beneficiaries and affected persons (e.g. older persons).
- **Impact** is about the societal, long-term impact of an intervention, including both intended and unintended consequences.

Figure 6: Results chain from input to impact of an intervention



Source: own drawing based on Frechtling, 2007.

Consider the following example of a psycho-social intervention for older persons. Inputs (financial and human resources, e.g., infrastructure, technological devices, psychotherapists), are expected to produce certain outputs (e.g., participation in psycho-social activity), which are expected to lead to certain outcomes (e.g., client satisfaction or increased health related quality of life), and which in turn would generate long-term impacts (e.g., improved wellbeing and healthy life expectancy of the target group).

There are several assumptions that would need to hold in order for the impact chain to occur. For example, a supply of appropriate therapists needs to be available, users

would be interested in the intervention. The links among level of output, outcomes and impacts are generally hypothetical.

Adapting the model by Edison et al (2013), which distinguishes between innovation metrics related to determinants, inputs, outputs, performance, and activities, we anchor and correlate the innovation measurement metrics in each element of the theory-of-change/results chain model for evaluating an innovative intervention. Please note that efficiency relates not only to outputs, but could also refer to outcomes and impact.

3.2 HOW TO DEVELOP THEORY-OF-CHANGE MODELS

Leeuw and Vaessen (2009) provide guidance on how to develop theories for carefully linking interventions to outcomes. Evaluators refer to "*black box*" problems when interventions are primarily viewed as producing certain effects without knowing why these effects occur. Ideally, an evaluation would make explicit the theories behind an intervention by investigating those—which theory-based evaluation approaches are able to offer. There is also an argument that mechanisms of change are more generalizable than concrete interventions adapted to a particular context, so that evaluators should make explicit theories of change in order to produce more generalizable assessments. This is particularly relevant in the context of evaluating complex interventions, where innovation effects may critically depend on many aspects, which is also typically the case in the realm of AHA.

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

Leeuw and Vaessen (2009) suggest the following steps for developing a theory of change model:

1. **Seeing interventions as theories:** First, whenever a new intervention is introduced, certain expectations exist that it will ameliorate a problem or achieve certain outcomes. Second, expectations involve a set of assumptions about how and why intervention activities and resources will bring about positive change. However, the underlying theory of an intervention is usually not made explicit. Theory-based evaluations are able to open the black box and shed light on the social, behavioural and institutional mechanisms.

2. **Articulating intervention theories on impact:** Theories of change can be developed and illustrated in many ways—a graphic display via logic models or results chains being examples. The theory of change can provide an overall framework for understanding the black box from input to impact. Starting point could be an intervention's already existing logical framework, expectations of policy and programme stakeholders, written evidence of past experiences of comparable interventions, and literature on change processes. The evaluator would compare and contrast the gathered assumptions and then develop an overarching intervention theory.

3. **Testing intervention theories on impact:** Evaluators need to test the quality of the assumptions on how an intervention is expected to lead to certain outcomes and impact. In an iterative process, a credible and reliable "causal story" is developed and refined, for example through causal contribution analysis (see Mayne, 2001), where the intervention is considered as one of the causes of the observed change. At the same time, alternative and competing explanations are investigated and ideally eliminated. For example, implementation failure might

be such an explanation. The challenge of this approach is the difficulty to attribute change, i.e. that it is challenging to estimate the magnitude and the extent to which the observed changes of indicator target values take place.

4 QUALITY CRITERIA FOR SELECTING INDICATORS

AHA Innovation evaluation metrics feed into the assessment of new and innovative policies, initiatives and innovations in a respective target setting. For this matter, it is not possible within the remit of this report to suggest a concrete, comprehensive and universally applicable list of indicators. The aim is rather to provide a framework as well as theories and criteria to choose, from a long list of potential indicators (D.T2.2.1) those that may best support innovation assessment and AHA decision making in a respective context. This way, the assessment framework developed within ASTAHG becomes flexible enough to be applicable to different settings and to provide results relevant for decision makers in various contexts. In the following section, we discuss the general lack of existing metrics for innovation assessment. In Section 4.2, we develop quality criteria to be used for selecting technically sound and policy-relevant indicators.

4.1 THE LACK OF METRICS FOR INNOVATION MEASUREMENT

The report distinguishes between indigenous innovation, i.e. innovation within a sector, and transfer innovation, i.e., an intervention that is being transferred from one context to another and therefore has innovative character in the new setting. The latter understanding is especially relevant for the ASTAHG project because multiple AHA sectors and geographic regions are involved. A central question underlying this report is what are the criteria that determine whether an innovation is preferable to

status-quo interventions. Several aspects of an innovative character can be distinguished:

- Development of a new substance, technology, process, data application, service
- Reduction of effort, cost, resources, data size, complexity, overheads
- Transfer to a new target group, sector, discipline, or process
- Novel combination of existing interventions

Edison et al. (2013) carried out a systemic literature review as well as surveyed and interviewed stakeholders to identify a comprehensive definition of innovation and to examine existing innovation evaluation measures. They then developed metrics for evaluating determinants, inputs, outputs and performance through aggregation and categorization. Based on such metrics, the authors constructed a conceptual model of key measurable elements of the innovation. The authors found a large number of metrics to measure innovation suggested in the literature and organized them into key aspects of innovation measurement as these relate to determinants, inputs, outputs, performance, and activities. They identified several challenges of innovation assessment based on the literature review:

1. A general lack of recognition of the importance of innovation assessment
2. A missing consistent definition of innovation
3. A lack of metrics for innovation assessment (which hinders organizations from pursuing innovation assessment)
4. Limited existence of guidelines and frameworks
5. Cost associated with innovation assessment

Edison et al (2013) explained the lack of metrics for innovation measurement in three ways:

- Lack of awareness of appropriate metrics. No studies existed that aggregated the existing metrics. This fact could explain the interviewees' ignorance about the existence of appropriate metrics.
- Lack of validation of metrics: The minority of metrics found in the literature were subjected to validation
- Interpretation: Because innovation can be defined differently, the values of the metrics are difficult to interpret. As a result, organizations often measure merely the revenue generated

Within the context of ASTAHG, we aim to address the challenges listed above. Deliverable D.T2.2.1 provides a long list of metrics, this report (Deliverable D.T2.2.2) provides theories and criteria for indicator selection in a particular decision context, and Deliverable D.T2.2.3 provides an assessment framework to support cross-sectorial AHA decision making.

4.2 SELECTING QUALITY CRITERIA FOR INDICATORS

Generally speaking, indicators ought to be selected using a systematic, carefully considered approach. Ideally, indicators are based on routinely collected data and existing data bases, which would minimize the cost burden of data collection. The guide "measuring the age-friendliness of cities" (WHO, 2015) provides an overall framework for defining indicator sets and measurement strategies. It aligns indicator sets with the theory-of-change model, where indicators range from input to output, outcome and impact indicators. Equity indicators are added that cross-cut the theory-of-change framework. This approach is very much aligned with the framework

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

developed in this report and Deliverable D.T2.2.3, where the theory-of-change model is embedded within the OEDC DAC criteria *relevance, coherence, effectiveness, efficiency, impact* and *sustainability* and operationalised within a multi criteria decision analytic framework.

Furthermore, the 2015 WHO guide also suggests key criteria for selecting indicators, which are provided in Table 1.

Table 1 : Quality criteria for selecting indicators (WHO 2015)

Quality criteria	Question for selecting an indicator
Measurable	Is the indicator actually measurable or observable?
Valid	Is the indicator measuring what it is supposed to measure?
Replicable	Can the indicator be collected in a standard way across time or across different contexts?
Sensitive to change	Will variations in the indicator be observable over time on account of specific actions?
Disaggregation possible	Can the indicator be disaggregated by gender, age group etc.?
Aligns with local goals and targets	Does the indicator link to a broader local agenda?
Can be linked to action	Does the indicator provide an understanding of the various actions that might need to be undertaken?
Within local influence	Does the local government or community have the mandate or authority to act on this indicator?
Easy to collect	Are the data required to produce the indicator easy to collect in a timely manner?
Socially acceptable	Is the collection of this information acceptable to the communities and individuals concerned?

Source: Own drawing based on WHO (2015)

Apart from these 10 key criteria for the various types of indicators, many international organizations and policy entities adhere to the widely used SMART criteria (e.g., UNDP 2009) for deciding on which indicators and variable to include (Table 2).

Table 2 : SMART indicators

Quality criteria	Description
Specific	Describing specific future conditions, ideally a change in individual or institutional behaviour, or in quality of life (depending on the level of outcomes)
Measurable	Using quantitative and qualitative measures to make it possible whether indicator was achieved or not
Achievable	Being in the capacity of the innovators to achieve
Relevant	Contributing to selected priorities of beneficiaries or national framework
Time-bound	Setting an expected date of accomplishment

Source: Own drawing based on UNDP, 2009.

Furthermore, Angelis and Kanavos (2016) suggest key properties for selecting variables for MCDA in Health Technology Assessment (HTA). These properties encompass the following characteristics:

- **unambiguous** (a clear relationship between the indicators and what they describe)
- **comprehensive** (cover the full range of consequences)
- **direct** (describe the consequences of options as directly as possible)
- **operational** (collectable information)
- **understandable** (understood and communicated across decision makers and other stakeholders)

For the purposes of this project, we decided to compare and merge the above criteria sets. We also differentiated between a '***technical realm'*** and a '***policy realm'*** of desirable criteria for AHA innovation metrics. By merging identical and / or similar criteria, we arrived at 7 technical and 7 policy criteria (see Table 3).

Table 3: ASTAHG criteria for selecting indicators

	Quality criteria for indicator selection	Sources
Technical indicator criteria	Measurable	WHO, SMART
	Specific	SMART, Angelis & Kanavos
	Valid	WHO
	Replicable	WHO
	Sensitive to change	WHO
	Disaggregation possible	WHO
	Easy to collect	WHO, Angelis & Kanavos
Policy indicator criteria	Achievable	SMART
	Aligns with local goals and targets	WHO
	Can be linked to action	WHO
	Within local influence	WHO
	Socially acceptable	WHO
	Time-bound	SMART
	Understandable	Angelis & Kanavos

Source: own drawing

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

The indicator criteria from Table 3 could be used within a checklist for selecting appropriate AHA innovation evaluation metrics from the long list of indicators provided in Deliverable D.T2.2.1. Tables 4 and 5 provide examples of how such a checklist could be applied to a respective target indicator.

Table 4 : Example for applying indicator criteria to "health-related quality of life"

	Quality criteria for indicator selection	Check	Comments
Technical indicator criteria	Measurable	☑	
	Specific	☑	
	Valid	☑	
	Replicable	☑	
	Sensitive to change		depending on the measurement instrument and the respective innovation under assessment. Consider for instance potential ceiling effects, etc.
	Disaggregation possible	☑	E.g. by population subgroups
	Easy to collect		Usually not collected routinely
Policy indicator criteria	Achievable	☑	
	Aligns with local goals & targets	☑	
	Can be linked to action	☑	
	Within local influence	☑	
	Socially acceptable	☑	
	Time-bound	☑	
	Understandable	☑	

Source: Own drawing

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

Table 5 : Example for applying indicator criteria to "healthy life expectancy at birth"

	Quality criteria for indicator selection	Check	Comments
Technical indicator criteria	Measurable	☑	
	Specific	☑	
	Valid	☑	
	Replicable	☑	
	Sensitive to change		Change might be small due to general deterioration of physical health
	Disaggregation possible	☑	E.g. by population subgroups
	Easy to collect	☑	
Policy indicator criteria	Achievable	☑	
	Aligns with local goals & targets	☑	
	Can be linked to action		Potentially difficult to link to a specific action
	Within local influence	☑	
	Socially acceptable	☑	
	Time-bound		Expected changes may only be observed at a later stage
	Understandable	☑	

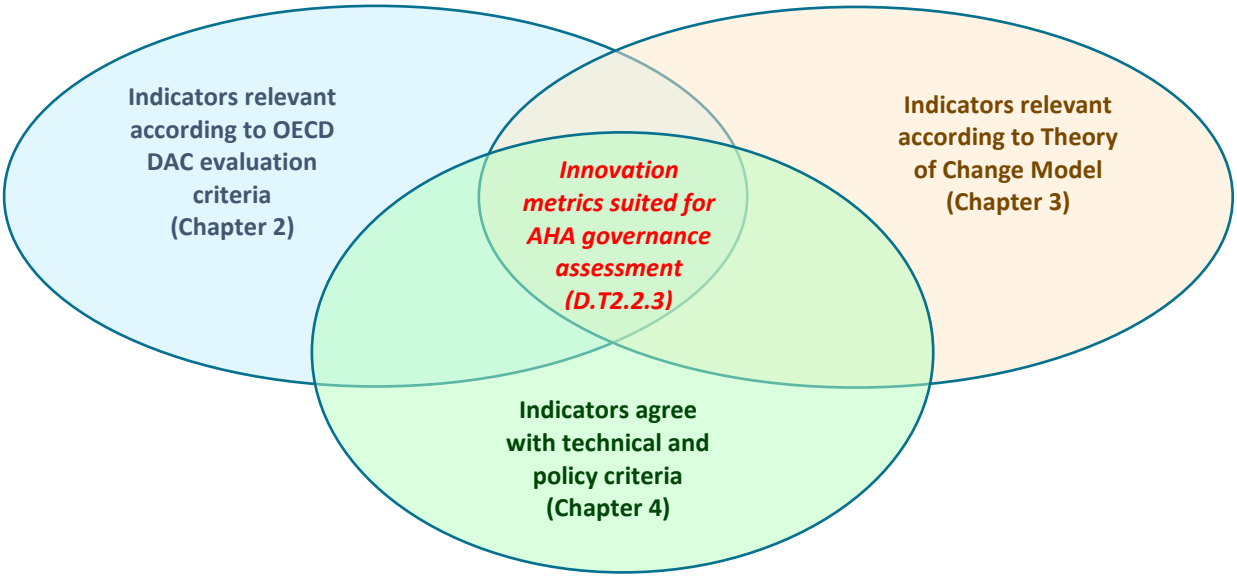
Source: Own drawing

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

5 DISCUSSION AND CONCLUSION

As mentioned before, AHA Innovation evaluation metrics feed into the assessment of new and innovative policies, initiatives and innovations in a respective target setting. For this matter, it is not possible within the remit of this report to suggest a concrete list of metrics. The aim was rather to provide a framework as well as theories and criteria to choose, from a long list of potential indicators (D.T2.2.1), those that may best support innovation assessment and AHA decision making in a respective context. This way, the assessment framework developed within ASTAHG becomes flexible enough to be applicable to different settings and to provide results relevant for decision makers in various contexts. Figure 7 displays a Venn-diagram that may be helpful for users of the ASTAHG governance assessment methodology to apply the theories, methods and criteria reported here in order to select and develop relevant AHA innovation evaluation metrics from the long list of indicators reported in D.T2.2.1.

Figure 7: Choice of AHA-innovation evaluation metrics



Source: Own drawing

The choice of AHA innovation evaluation metrics should be based on the abovementioned frameworks, theories and criteria, but it needs to reflect local circumstances and therefore must take place on a local level, including respective stakeholders from all AHA sectors affected and involved in AHA decision-making. The steps that should be taken to define a concrete list of AHA innovation evaluation metrics for AHA governance assessment in a certain context may involve:

- **Step 1:** Use the OECD DAC evaluation criteria (relevance, coherence, effectiveness, efficiency, impact sustainability) as dimensions of AHA innovation evaluation metrics
- **Step 2:** Combine the evaluation criteria with the results chain (input, activities, output, outcomes, impact using theories of change
- **Step 3:** Determine the available indicators for each step of the results chain, based on the long-list of AHA impact evaluation metrics (D.T.2.2.1)
- **Step 4:** Determine the quality dimensions for each indicator along desirable technical and policy criteria for AHA innovation evaluation metrics
- **Step5:** Confirm the respective short-list of indicators for AHA governance assessment (Deliverable D.T2.2.3)

This report therefore bridges between the long-list of AHA impact evaluation metrics presented in Report D.T2.2.1, and the AHA governance assessment methodology presented in Report D.T2.2.3. The latter report aims to provide a comprehensive assessment framework for innovation to support AHA decision-making in a multisectoral context. It is based upon the OECD DAC evaluation criteria and provides a stepwise approach where AHA innovations funnel through until only those remain that are deemed relevant, coherent, effective, efficient, provide impact and are financially sustainable in a particular setting.

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

The core of the assessment framework is a Multi Criteria Decision Analytic approach that essentially incorporates:

- relevant indicators drawn from the long list provided in D.T2.2.1 and selected along the theories and indicator properties explained in this report
- Relevance weights for indicators based on AHA stakeholder needs and preferences

Ultimately, this will enable comparative assessment of AHA innovations with multiple and diverse outcomes and help local decision-makers identifying the most beneficial technologies available for their respective geographic settings.

6 REFERENCES

Angelis, A. & Kanavos, P. (2016) Value-Based Assessment of New Medical Technologies: Towards a Robust Methodological Framework for the Application of Multiple Criteria Decision Analysis in the Context of Health Technology Assessment, PharmacoEconomics Vol. 34, pp. 435–446.

ASTAHG (2018): *project proposal.*

Bertelsmann Stiftung (2020). Sozialplanung für Senioren. Online Available (15.04.2020): https://sozialplanung-senioren.de/index.html

Boehler, C., de Graaf, G., Steuten, L., Yang, Y. & Abadie, F. (2015). Development of a web-based framework for the assessment of health and economic outcomes of the European Innovation Partnership on Active and Healthy Ageing (EIP on AHA). *BMC Medical Informatics and Decision Making*, 15(suppl. 3): S4

Boehler, C. & Abadie, F. (2015). *Monitoring and Assessment Framework for the European Innovation Partnership on Active and Healthy Ageing (MAFEIP). Conceptual description of the monitoring framework for the EIP on AHA.* Seville, Spain: European Commission, Joint Research Centre (JRC), Institute for Prospective Technological Studies (IPTS), JRC96205.

Boehler, C. & Abadie, F. (2016). *Monitoring and Assessment Framework for the European Innovation Partnership on Active and Healthy Ageing (MAFEIP). Gap analysis.* Seville, Spain: European Commission, Joint Research Centre (JRC), Institute for Prospective Technological Studies (IPTS).

Carayannis, E.G. & Campbell, D.F.J. (2009) *"Mode 3" and "quadruple helix": toward a 21st century fractal innovation ecosystem.* International Journal of Technology Management, Vol. 46(3/4), pp. 201–234.

Chianca, T. (2008) The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement, Journal of MultiDisciplinary Evaluation, Vol. 5, No. 9.

Drummond M. F. et al. (2005) Methods for the Economic Evaluation of Healthcare Programmes (3rd edt), Oxford University Press, Oxford, UK.

This project is co-financed by the European Regional Development Fund through the Interreg Alpine Space programme.

Edison, H.; Ali, N.B.; Torkar, R. (2013) Towards innovation measurement in the software industry). *Journal of Systems and Software*. 86 (5): 1390–1407.
Feinstein, O. & Beck, T. (2006). Evaluation of development interventions and humanitarian actions. In: The Sage Handbook of Evaluation ed. By I. Shaw et al., pp. 536-558.

Forss, K. & Bandstein, S. (2008) Evidence-based Evaluation of Development Cooperation: Possible? Feasible? Desirable? IDS Bulletin, Volume 39 Number 1, Institute of Development Studies.

Frechtling J (2007) Logic Modeling Methods in Program Evaluation, San Francisco, CA: Jossey-Bass.

Kahlert, R. (2013). *Randomized controlled trials to evaluate impact. Their challenges and policy implications in medicine, education and international development.* Texas Digital Library, Online available (30.06.2020): https://repositories.lib.utexas.edu/handle/2152/19463.

Leeuw, F. L. & Vaessen, J. (2009). Impact evaluation and development. NONIE Guidance on Impact Evaluation. Washington, DC: World Bank. Online available (30.06.2020): http://documents.worldbank.org/curated/en/411821468313779505/Impact-evaluations-and-development-NONIE-guidance-on-impact-evaluation

Mayne, J. (2001). Addressing attribution through contribution analysis: using performance measures sensibly, *The Canadian Journal of Program Evaluation*, vol. 16, no. 1, pp. 1–24.

Mihnovits, A. & Zaidi, A. (2015). Global AgeWatch Index 2015. Methodology update. Published by HelpAge International (This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License) Online available (15.04.2020): https://www.helpage.org/global-agewatch/reports/global-agewatch-index-2015-insight-report-summary-and-methodology/

MoU (2019) – *ASTAHG memorandum of understanding for the setting up of the transnational governance board*, ASTAHG deliverable D2.1.2, ANNEX 1.

OECD (1992) Development assistance manual: DAC principles for effective aid. Paris: Organization for Economic Cooperation and Development.

OECD (2002) Evaluation and Aid Effectiveness No. 6 - Glossary of Key Terms in Evaluation and Results Based Management (in English, French and Spanish), OECD Publishing, Paris. Online available (30.06.2020): https://doi.org/10.1787/9789264034921-en-fr.

OECD (2015) *Ageing in Cities*, OECD Publishing, Paris, Online Available (30.06.2020): https://doi.org/10.1787/9789264231160-en

OECD (2019) Updated DAC-evaluation criteria based on a global consultation on the evaluation criteria. Online available (05.03.2020): http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

Rechel, B. et.al. (2013). *Ageing in the European Union. The Lancet.* Health in Europe Series. Vol. 391 (9874), pp. 1312-1322. Online available (30.06.2020): http://www.healthyageing.eu/sites/www.healthyageing.eu/files/resources/Ageing%20in%20the%20EU%20The%20Lancet.pdf

UNECE (2013): Active Ageing Index project. Online available (15.04.2020): https://statswiki.unece.org/display/AAI/Active+Ageing+Index+Home

United Nations Development Programme (2009) Handbook on planning, monitoring and evaluating for development results. New York.: UNDP. Online available (30.06.2020): http://web.undp.org/evaluation/handbook/documents/english/pme-handbook.pdf

WHO (2002). Active Ageing: A Policy Framework. Online available (02.04.2020): https://apps.who.int/iris/bitstream/handle/10665/67215/WHO_NMH_NPH_02.8.pdf?sequence=1

WHO (2007). Global Age-friendly Cities: A Guide. Online available (02.04.2020): https://www.who.int/ageing/publications/Global_age_friendly_cities_Guide_English.pdf

WHO European Observatory on Health Systems and Policies (2012). *Intersectoral governance for health in all policies. Structures, actions and experiences*. McQueen, D.V., Wismar, M., Lin, V., Jones, C.M., Davies M. (Editors). WHO Observatory Studies Series No.26, ISBN: 978 92 890 0281 3. Online available (30.06.2020): http://www.euro.who.int/en/publications/abstracts/intersectoral-governance-for-health-in-all-policies.-structures,-actions-and-experiences-2012

WHO (2013) *The Helsinki Statement on Health in All Policies.* The 8th Global Conference on Health Promotion, Helsinki, Finland, 10-14 June 2013. Online available (30.06.2020): https://www.who.int/healthpromotion/conferences/8gchp/8gchp_helsinki_statement.pdf?ua=1

WHO (2015). Measuring the age-friendliness of cities: a guide to using core indicators. Online available (30.06.2020): https://www.who.int/ageing/publications/measuring-cities-age-friendliness/en/

WHO (2017) *Global strategy and action plan on ageing and health.* Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.

WHO Europe (2018). Age-friendly environments in Europe: Indicators, monitoring and assessments. Regional Office for Europe, Online available (18.05.2020): http://www.euro.who.int/__data/assets/pdf_file/0003/377481/afee-indicators-monitoring-eng.pdf?ua=1

Zaidi, Asghar (2013). Global AgeWatch Index 2013. Purpose, methodology and results. Published by HelpAge International, 2013. Online available (15.04.2020): https://www.helpage.org/global-agewatch/reports/global-agewatch-index-2015-insight-report-summary-and-methodology/

Zaidi, A. et al. (2013). Active Age Index 2012. Concept, Methodology and Final Results. European Centre in Vienna. Online available (15.04.2020): https://www1.unece.org/stat/platform/download/attachments/76287849/Methodology-Paper%20Final.pdf?version=1&modificationDate=1393836329990&api=v2