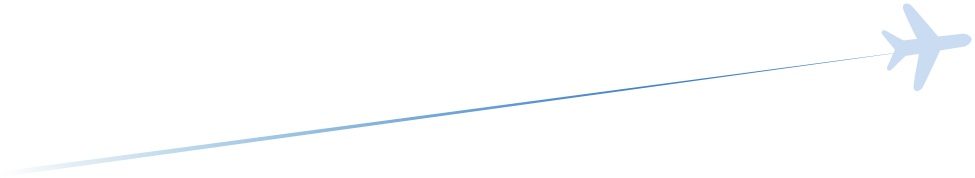

D3.8 Mapping ATM research concepts, past and future - including the Engage wiki implementation

Deliverable ID:	**3.8**
Dissemination Level:	**PU**
Project Acronym:	**Engage**
Grant:	**783287**
Call:	**H2020-SESAR-2016-2**
Topic:	**SESAR-ER3-01-2016 Knowledge Transfer Network**
Consortium Coordinator:	**University of Westminster**
Edition date:	**16 February 2021**
Edition:	**01.01.00**
Template Edition:	**02.00.02**

Founding Members
EUROPEAN UNION EUROCONTROL
SESAR JOINT UNDERTAKING

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Graham Tanner / University of Westminster	Consortium member	16 February 2021
Pablo Hernández / Innaxis	Consortium member	16 February 2021
Junzi Sun / Technische Universiteit Delft	Consortium member	16 February 2021
Inés Gómez / Innaxis	Consortium member	16 February 2021
Tatjana Bolić / University of Westminster	Consortium member	16 February 2021
Paula López / Innaxis	Consortium member	16 February 2021
Andrew Cook / University of Westminster	Project coordinator	16 February 2021

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Andrew Cook / University of Westminster	Project coordinator	16 February 2021
Graham Tanner / University of Westminster	Consortium member	16 February 2021

Approved for submission to the SJU By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
Andrew Cook / University of Westminster	Project coordinator	16 February 2021

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
N/A		

Document History

Edition	Date	Status	Author	Justification
01.00.00	24 December 2020	Release	Engage Consortium	New document for review by the SJU

01.01.00	16 February 2021	Release	Engage Consortium	Incorporating SJU comments

Copyright Statement

Engage

THE SESAR KNOWLEDGE TRANSFER NETWORK

This deliverable is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 783287 under European Union's Horizon 2020 research and innovation programme.

Abstract

This report presents the wiki produced by Engage, SESAR's Knowledge Transfer Network. It summarises the key development processes, status and planning for the wiki, which has been built over the past two years, with increased activity in 2020 to resolve underlying data provision issues. *Inter alia*, the wiki hosts the first interactive research map of European ATM, an ATM concepts roadmap, the first consolidated listing of European university programmes, and sets out progress towards a new, one-stop (data) repository for the research community.

Table of Contents

List of Tables

List of Figures

1 Introduction

The Engage wiki was launched at the SESAR Innovation Days' closing session, on 10 December 2020. User registration was opened on the same day. Any researcher or tutor may apply for a (free) authorised account, to become an active member of SESAR's Knowledge Transfer Network community.

Live link to wiki: https://wikiengagektn.com/

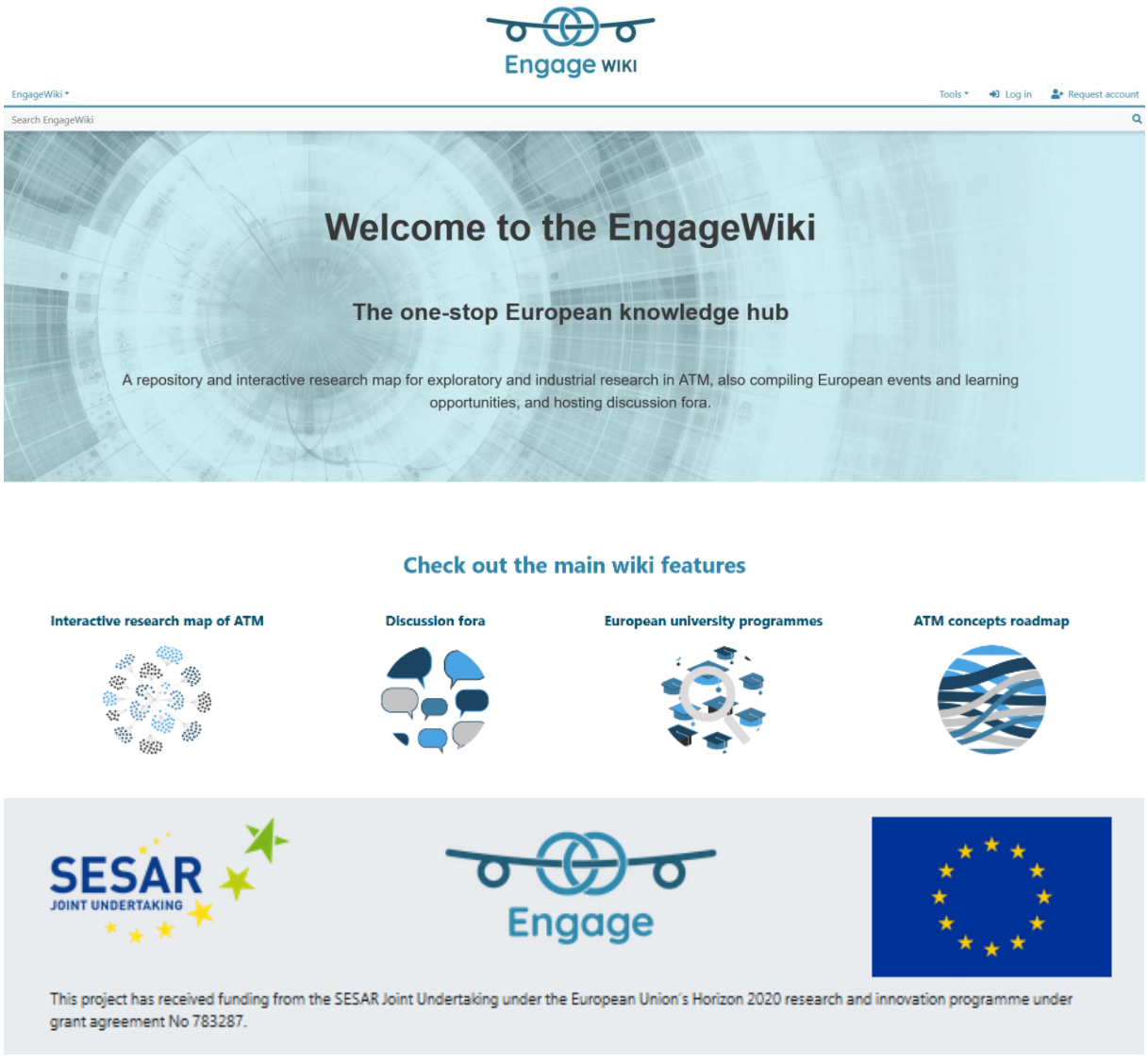

Figure 1: Wiki landing page image extracts

This report summarises the key development processes, status and planning for the wiki, which has been conducted over the past two years, with increased activity in 2020 to resolve underlying data provision issues, many of which have been resolved.

Table 1 summarises the key content features, current status, and the section of the report (where appropriate) that details the development to date. This is an interim report, with a final edition expected in September 2021.

Table 1: Summary of wiki content and current status

Wiki activity / page name	Short description	Current status
Interactive research map	Interactive research map visualisation where users can explore the results of a bottom-up clustering from unsupervised machine learning applied to SESAR 1 and SESAR 2020 projects and papers	**Active** in the wiki; detailed in this report in Section 3
ATM concepts roadmap	Interactive roadmap that shows how previous (SESAR) research connects with the flagship activities of the 2020 Strategic Research and Innovation Agenda, and identifies future challenges (with a dedicated users' discussion forum)	**Active** in the wiki; detailed in this report in Section 4
European university programmes	Interactive database of undergraduate (UG) and postgraduate (PG) programmes offered in Europe; features UG courses related to air transport engineering and aviation management and PG courses that perform ATM-related research; user-updateable	**Active** in the wiki; detailed in this report in Section 5.1.4
Discussion fora	Discussion fora for common interest research communities. Open to all registered users	**Enabled** throughout wiki: see for examples: Figure 2, Table 7, and Table 12.
Research repository	This remains a core objective of Engage, i.e. to establish a one-stop, go-to source for information: a single European point of entry for ATM knowledge. With improved search functionality, researchers may turn to the hub as an accessible meta-source of research data	Wiki page is **set up** ready at: https://wikiengagektn.com/EngageWiki:Research_repository This functionality is under development: see Section 2.1.
PhD funding opportunities	Open PhD funding opportunities. Open to registered users to add new opportunities (Content will be fully migrated from Engage website.)	**Active** in the wiki: https://wikiengagektn.com/EngageWiki:PhD_funding_opportunities
Jobs and internships	Vacant job and internship positions. Open to registered users to add new vacancies (Content will be fully migrated from Engage website.)	**Active** in the wiki: https://wikiengagektn.com/EngageWiki:Jobs_and_internships
Projects repository	WP-E summaries and materials that are typically not available elsewhere (e.g. SESAR JU or CORDIS)	**Active** in the wiki; detailed in this report in Section 5.1.5

Figure 2: Example of discussion forum (for the ATM concepts roadmap)

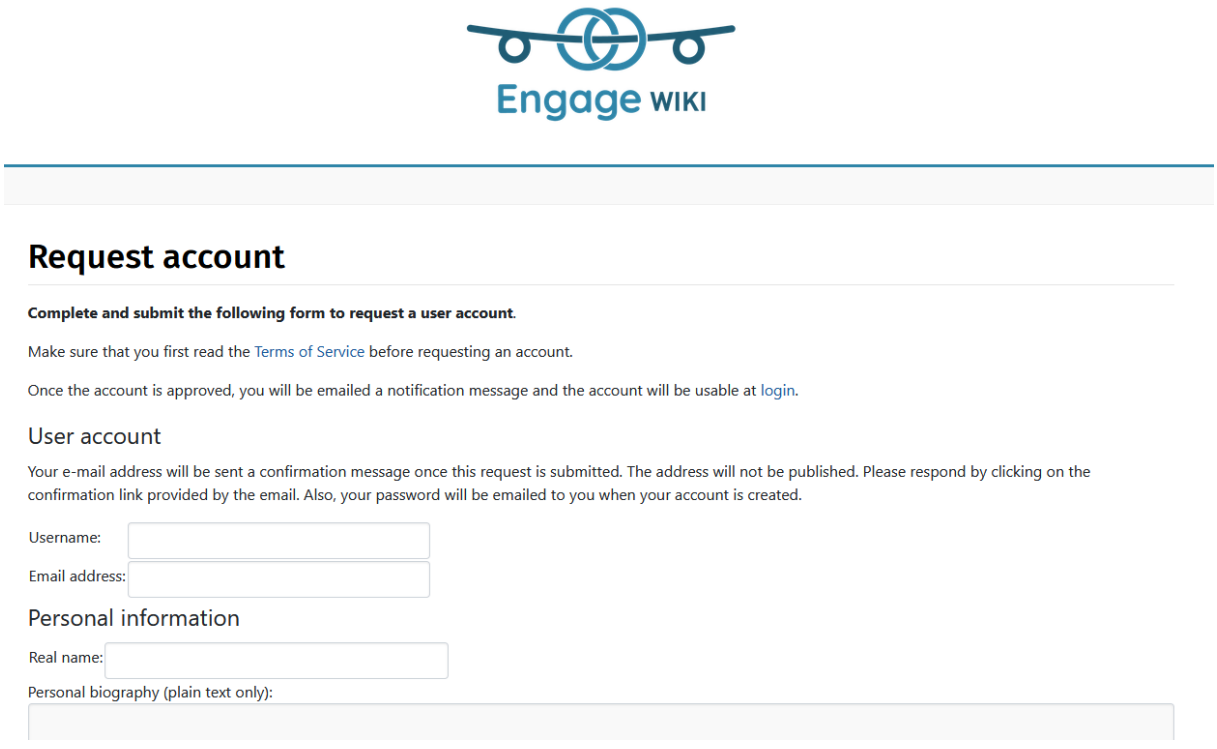

Figure 3: Requesting an account

2 Data sources and background

2.1 Data sources – at launch of wiki

This section outlines the sourcing of data required for the Engage interactive research map, ATM concepts roadmap and research repository. This activity started in May 2019, building on earlier work which explored opportunities for innovative ATM research (as reported in Deliverable 3.5 [24]). The initial task was to identify potentially available material at a high level across exploratory and industrial research (ER/IR) under the supervision of the SJU, from the SESAR 1 and SESAR 2020 programmes:

- SESAR 1 projects and activities 2008-2016 (see Figure 4);
- SESAR 2020 projects and activities 2015-2024 (see Figure 5).

This process was assisted by a series of meetings with the SJU to determine the availability and usage of SESAR deliverables by Engage. A phased approach was agreed in which SESAR 1 deliverables would be initially sourced from the SJU (all SESAR 1 projects closed), to be followed by SESAR 2020 material from CORDIS.

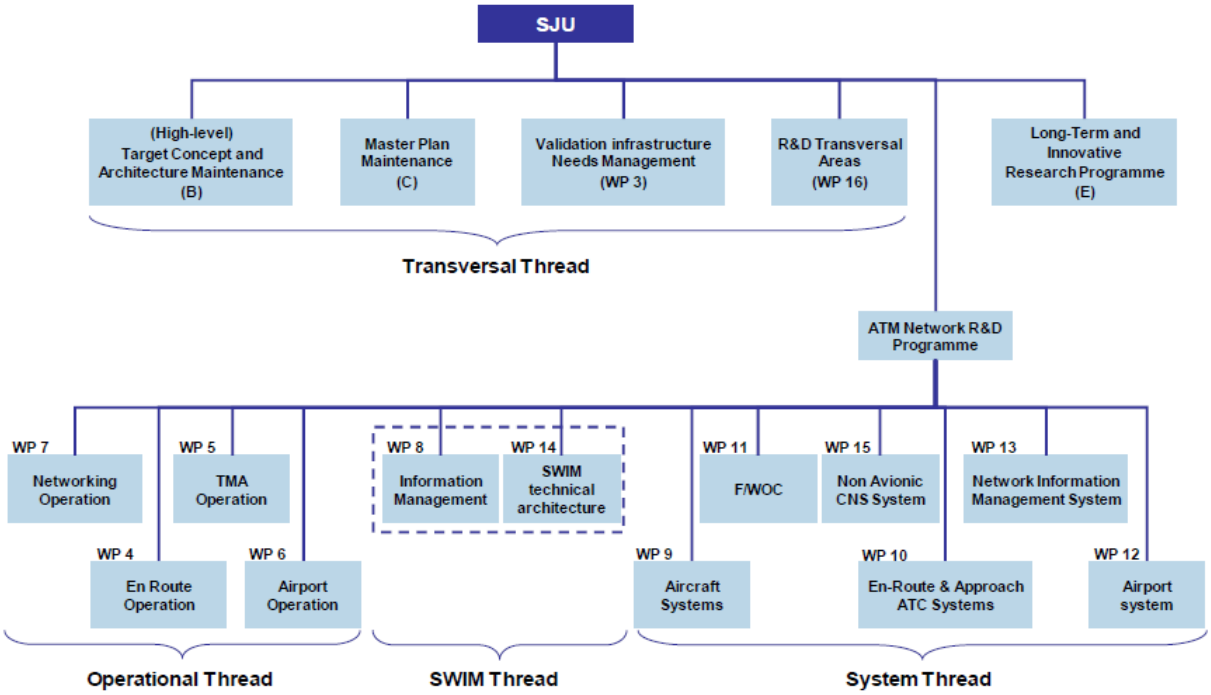

Figure 4: WP activities of the SESAR 1 Programme 2008-2016

Source: [1] (Figure 3)

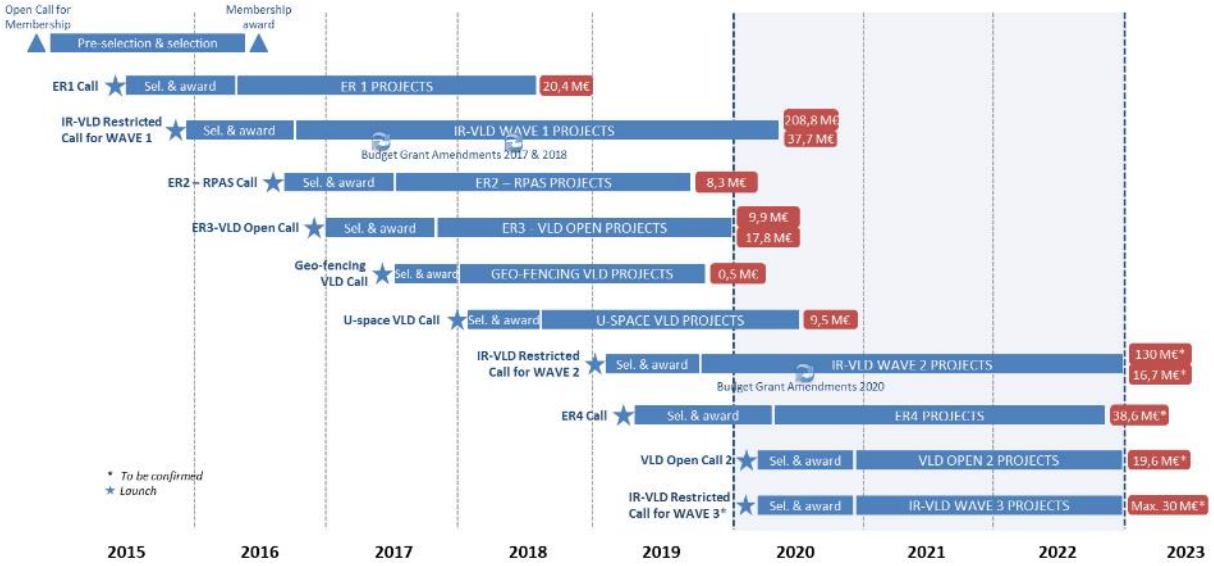

Figure 5: Call activities of the SESAR 2020 Programme 2015-2022, as of 2020

Source: [2] (Figure 10)

Table 2 lists the identified Calls, although suitable material was not available for all of these (or the topics were of less critical importance) for the wiki:

- SESAR 1 WP-E: no proposal was awarded funding in response to the 2nd Call for research networks;
- SESAR 1 IR: many deliverables remain confidential;
- SESAR 2020 VLD Geofencing and U-space: considered out of scope;
- SESAR 2020 IR Waves 2, 3 and ER4: are still at an early stage with limited material published so far.

Table 2: Identified SESAR 1 and SESAR 2020 Calls

SESAR programme	Call	Call ID	Project start-end dates
SESAR 1	WP-E	10-220210: 1st Call for research networks 10-220718: 2nd Call for research networks (economics and performance) 10-220719: 1st Call for WP-E research projects 12-120610: 2nd Call for WP-E research projects	2011-2016
	IR	BAFO1; BAFO2; BAFO3 (Best and Final Offer)	2011-2016
	IR (Demo) IR (AIRE III)	SJU/LC/0070: Integrated Flight Trials and Demonstration Activities	2012-2014

	IR (RPAS)	SJU/LC/0087: Integrated Remotely Piloted Air System Demonstration Activities	2013-2016
	IR (LSD)	SJU/LC/0102: Large Scale Demonstration Activities	2014-2016
SESAR 2020	ER1	H2020-SESAR-2015-1: 1st Call for research projects	2016-2018
	IR Wave 1	H2020-SESAR-2015-2: IR-VLD Wave 1	2016-2019
	ER2	H2020-SESAR-2016-1: RPAS Exploratory Research Call	2017-2019
	ER3	H2020-SESAR-2016-2: Exploratory Research and Very Large Scale Demonstrations Open Call	2017-2019
	VLD Geofencing	SESAR-2017-1 VLD Geofencing	2018-2019
	U-space	CEF-SESAR-2018-1: U-space	2018-2020
	IR Wave 2	H2020-SESAR-2019-1: IR-VLD Wave 2	2019-2022
	ER4	H2020-SESAR-2019-2: Exploratory Research 4	2020-2022
	VLD Open 2	H2020-SESAR-2020-1: 2nd Call for Very Large Scale Demonstration	2020-2022
	IR Wave 3	H2020-SESAR-2020-2: IR-VLD Wave 3	2020-2022

2.1.1 SESAR 1 deliverables

Following the resolution of legal difficulties that delayed access to SESAR documentation, two batches of SESAR 1 deliverables were made available to Engage from various SESAR libraries. These have been complemented with a small number of public deliverables and papers (<20) retrieved from the ComplexWorld project wiki (see Section 5.1.5). Before being used by the research map and ATM concepts roadmap, it was necessary to perform a series of file management processes:

- determine the WBS number of each deliverable (note: it was not possible to systematically derive this); this enabled each file to be matched with a project;
- detect deliverables duplicated between the various SESAR libraries (note: in some cases a deliverable may consist of multiple files);
- assign metadata per file, such as deliverable number, title and authorship.

(Further data management and cleaning activities specific to the research mapping are explained in Section 3.2.) The file metadata were collated in a simple database, and for traceability, the original .doc/.docx/.pdf file names were retained. Project metadata were also collated in a separate database, recording the funding Call, project theme, consortium partners, total budget (with co-funding), start/end dates and TRL. This was a largely manual task and in many instances the data are not publicly available.

Whilst no further SESAR 1 deliverables are expected to be sourced, additional work is required to further populate both project and document metadata.

2.1.2 SESAR 2020 deliverables

All public SESAR 2020 reports are available from CORDIS [26], the EC's repository of results from projects funded by EU framework programmes. Note that for IR projects in CORDIS, public reports include deliverables, SESAR Solution data packs and demonstration reports (the latter relating to VLD projects).

Whilst deliverables sourced from CORDIS cannot be republished without permission from the corresponding projects, as public documents, they can be analysed for use by the research map and ATM concepts roadmap tasks – priority has been given to sourcing final project results and demonstration reports. Project metadata are exportable from CORDIS, enabling the required fields to be fully populated for the 80 SESAR 2020 projects included in the first public version of the wiki (ER1-3 and IR Wave 1).

The SESAR 2020 file management processes were as follows:

- the H2020 grant agreement number was used as the unique project identifier in place of the WBS number;
- check each deliverable to ensure the downloaded file matched the CORDIS description (note: as with SESAR 1, in some cases a deliverable may consist of multiple files; some CORDIS-listed files were missing);
- determine the relevance of files listed in CORDIS, e.g. videos were unsuitable for research map or ATM concepts roadmap analysis;
- assign metadata per file, such as deliverable number, title and authorship.

The remaining SESAR 2020 deliverables need to be sourced for ER1-3 and IR Wave 1 projects, and corresponding document metadata completed (primarily deliverable authorship). Project metadata and deliverables also need to be sourced for ER4, VLD Open 2 and IR Waves 2-3; note this will be an iterative approach as new material continues to be published by on-going projects.

2.1.3 SIDs papers

In parallel to the sourcing of SESAR deliverables, conference papers presented at the SESAR Innovation Days were collated with the assistance of EUROCONTROL. The complete set of 2011-2019 papers were available for use with the first public version of the wiki (see Table 3).

The SIDs paper file management processes were as follows:

- papers were checked and matched with SESAR projects in order to assign the WBS number or H2020 grant agreement number; note:
    - accepted papers could be associated with non-SESAR projects;
    - accepted papers could be associated with multiple SESAR projects (i.e. joint authorship);
    - a SESAR project could have multiple papers accepted at a conference;
    - there may be papers for which the link with a SESAR project has not been identified.
- assign metadata per paper, such as title, conference track and authorship.

Papers from SIDs 2020 (and future conferences) need to be sourced, matched with SESAR projects and metadata completed.

Table 3: SIDs papers sourced and matched with SESAR projects

SIDs	Total papers	Papers matched with SESAR projects
2011	28	17 WP-E
2012	27	14 WP-E; 1 IR
2013	28	13 WP-E; 1 IR
2014	30	19 WP-E; 2 IR
2015	28	17 WP-E; 1 IR
2016	32	3 WP-E; 2 IR; 6 ER1
2017	35	20 ER1
2018	34	9 ER1; 1 ER2; 1 ER3
2019	38	1 ER1; 4 ER3; 6 IR Wave 1

2.1.4 Use case survey

A use case survey based on a standalone version of the interactive research mapping tool was conducted in August 2020. An invitation to test the tool was sent to a range of users, to capture both expert and novice views from academia and industry (users included Engage industry partners, PhD candidates and others outside the KTN). User feedback was treated anonymously, with all comments carefully considered, however the following non-attributable points were prioritised for implementation in the interactive research map in time for the wiki launch:

- add cluster names *[implemented];*
- add a link to reveal cluster information/groupings/keywords *[implemented];*
- add simple instructions *[implemented];*
- add report name(s) to output box *[implemented];*
- populate metadata *[partially populated; still to be completed];*
- remove personal names from keywords *[names removed].*

After discussion within the development team, some suggestions were not thought to be suitable for implementation, although other ideas will be considered for inclusion in the next version (note: some rely on the availability of more comprehensive metadata). These include:

- how will documents be handled when the 'Reports' button on the output box is available?
    - *Engage: this is subject to permission being granted to republish deliverables*
- display different project nodes to signify maturity, budget or other information
    - *Engage: to be considered, though may overly complicate displayed information*

- add new search filters such as partners/budget/theme/etc.
    - *Engage: additional filters such as budget (e.g. <€1m, €1m-€5m, etc.) and Call (e.g. ER1, ER2, IR Wave 1, etc. or ER fundamental, ER application-oriented, etc.) are being considered, however these rely on complete metadata*
- process other non-SESAR documents
    - *Engage: to be considered, though depends on availability of non-SESAR material and corresponding metadata*
- highlight nodes that match the search filter results
    - *Engage: to be considered, though may overly complicate displayed information*
- increase project timeline granularity to include month and year
    - *Engage: to be considered, however relies on more detailed metadata*
- add 'buzzword' keywords to the project metadata, e.g. digitalisation, AI
    - *Engage: to be considered, though relies on new keywords for each project that will most likely need to be manually derived*

2.1.5 Summary of data used in the first public version of the wiki

Table 4 summarises the coverage of SESAR projects at the time of launch. Deliverables from 338 (of 374) SESAR projects were analysed for use by the interactive research map and ATM concepts roadmap tasks (both described later in this report), mainly from SESAR 1. For some projects, SIDs papers were also included.

Table 4: Project summary at wiki launch 2020

SESAR programme	*with project metadata*	SESAR projects with deliverables*	with SIDs papers
SESAR 1	*294*	293	90
SESAR 2020	*80*	45	48
Total	*374*	338	138

* At least one project deliverable or project material.

As highlighted in the preceding sub-sections, there were a number of gaps in the data that could not be filled in time for the wiki launch.

- SESAR 1 project metadata: budgets, start/end dates, consortium partners, project theme and TRL are largely missing

- *generally unavailable and a time-consuming process checking each deliverable*

- SESAR 1 and SESAR 2020 deliverable metadata: authors missing
  - *manual processing of metadata are required*
- SESAR 2020 deliverables/data packs: ER4, VLD Open 2 and IR Waves 2-3 still to be sourced
  - *although available from CORDIS, separately downloading each file or archive is a time-consuming process*
- Republishing SESAR 1 deliverables on the wiki: permission required
  - *Permission is expected to be granted in the future*

Table 5 consolidates the missing data, highlighting the consequence of not resolving the issues relating to the research map in time for the next version of the wiki.

Table 5: Consolidated list of the outstanding data issues

Missing	Consequence
The names of authors have been removed from the project information output box used in the research map, as requested in feedback received during development	When the user clicks on the corresponding node representing a particular project, they are currently unable to see any information regarding who to contact for further information on the project; this has been flagged in lessons learned for SESAR 3 (in Section 6)
Metadata for SESAR 1 projects (e.g. partners, dates, budgets) are largely missing due to their unavailability from public sources or the documents made available to Engage	Although functionality is unaffected, many of the IR results presented to the user contain gaps and 'n/a' in the fields, thus affecting usability
SESAR 1 content for the research repository	The repository cannot be fully implemented until permission is granted, allowing SESAR 1 deliverables to be republished in the wiki; SESAR 2020 deliverables will be initially reachable via links to CORDIS (see also Section 7)

2.2 Data sources – next steps

The following data updates are required for analyses by the research map and ATM concepts roadmap in time for the next version of the wiki:

- subject to the availability of the required SESAR 1 metadata, missing content will be added to help describe projects and deliverables;
- the remaining SESAR 2020 deliverables will be sourced, and metadata prepared;
- complete metadata are required for fields being considered as new filters for the research map (e.g. project budget);
- papers presented at the tenth SIDs (December 2020) will be sourced, and metadata prepared.

3 Interactive research map

3.1 Motivation and overview

In the SESAR Master Plan (2020 edition, [3]), SESAR research is grouped according to four key features: optimised ATM network services; high-performing airport operations; advanced air traffic services; and, enabling aviation infrastructure. Commonalities across these domains may be mapped through the Essential Operational Changes. This picture is further extended by the presence of ATM Technology Changes, Operational Improvement Steps and Enablers, and ESSIP objectives. This creates a necessarily complex landscape for mapping between research activities in technical contexts, such as EATMA.

However, at the research and thematic exploration level, the lack of simple and mappable access to specific deliverables by those engaged in exploratory research (ER) has long-since been cited as a barrier for improving awareness of industrial research (IR) and for mutually building links between the two domains. One approach to bridge this gap is to provide better connections and links between such research projects.

This Engage mapping activity is focused on developing a set of tools and methods to explore the valuable information and results that are produced through SESAR, particularly through building interactive maps that are easy to use and embrace both ER and IR activities. We here present a bottom-up approach to map and classify research projects. By gathering information in the form of keywords from project documentation, and by subsequently applying clustering algorithms, it is possible to define new clusters (themes) of homogeneous projects.

3.2 Research mapping process – building the clusters

3.2.1 Overview of the research mapping flow

The process of knowledge mapping based on SESAR ER and IR outputs described in Section 2.1 comprises several key steps. In Figure 6, the entire flow is illustrated. Four major steps are defined:

1. pre-processing;
2. natural language processing (NLP);
3. keywords analysis and statistics;
4. machine learning.

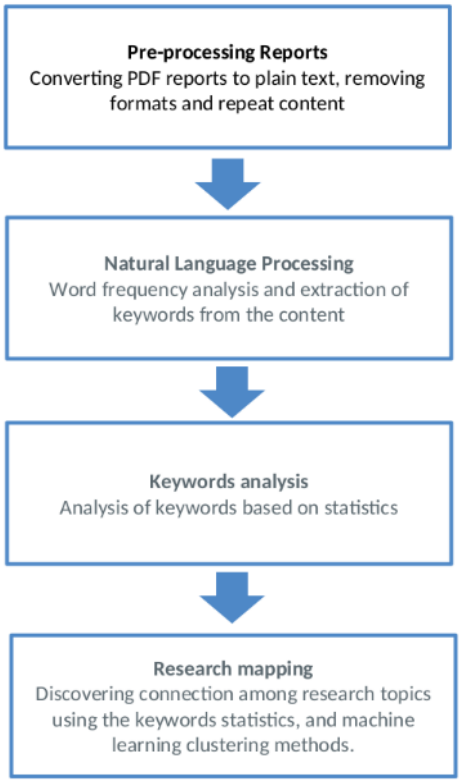

Figure 6: Research mapping flow

In the following, the details of the steps are discussed. Examples and results are shown in each section.

3.2.2 Pre-processing SESAR outputs

Almost all SESAR project deliverables and research papers are published in PDF or Word format. The projects use the project report templates supplied by the SESAR JU. This means that these documents share similar structures and provides the potential of being processed by customised software.

Before any computer algorithm can be applied to mining the text of these reports, it is important to convert them into plain text, so they can be processed by software tools.

There are several challenges while dealing with these PDF documents. First of all, there are fixed contents that can appear in all documents, such as header, footer and front pages. These elements have to be removed, which often requires manual effort. Secondly, there are often different font formats included in these documents: the developed tools have to be able to remove the formatting information, and keep only the text. Lastly, even though most of the content is presented in text format, there could be potential text information in images. This may require additional pre-processing involving image recognition, which is currently out of scope.

3.2.3 Common information removal

For all project reports, pages containing general information on the project, such as authoring and approval information and document history, are removed through a manual process (although see later comment on automation).

Then, the common elements of the documents are removed, which include the header and footer sections. Figure 7 shows an example of cropping a PDF report, leaving only the content area, to illustrate the extraction of the content area of a SESAR industrial research report.

On the left-hand side is a page from an original public SESAR industrial research report, where the header and footer information of the document is visible. The right-hand side shows the area where the main content is located. All contents outside of the red box are removed using the Adobe Acrobat tool. The process can be semi-automated by first manually marking one page in the document. Then, the entire document will be processed using this pre-defined content setting. Furthermore, if all documents are generated using the same SESAR template, a script could be made to fully automate the process for all reports. A similar approach was followed for the SIDs papers (see Section 2.1.3).

Figure 7: Example of cropping the PDF report, leaving only the content area

3.2.4 Text extraction

Next, the plain text needs to be extracted from the PDF documents generated by the previous step.

Several different tools were evaluated for the extraction of text. For example, Adobe Acrobat, web-based PDF conversion tools, as well as the standalone open-source pdf-to-text tools.

A customised tool was finally built based on the open-source package library in Python, *pdftotext*, to accomplish this task. Besides its benefit of being open-source, the choice of this Python library allows easy integration of PDF pre-processing into the mapping framework.

The text extraction tool removes all the formatting elements contains in any PDF document and leaves only the text content. In Figure 8, further pre-processing of the page from the previous example is illustrated. The plain text output is shown.

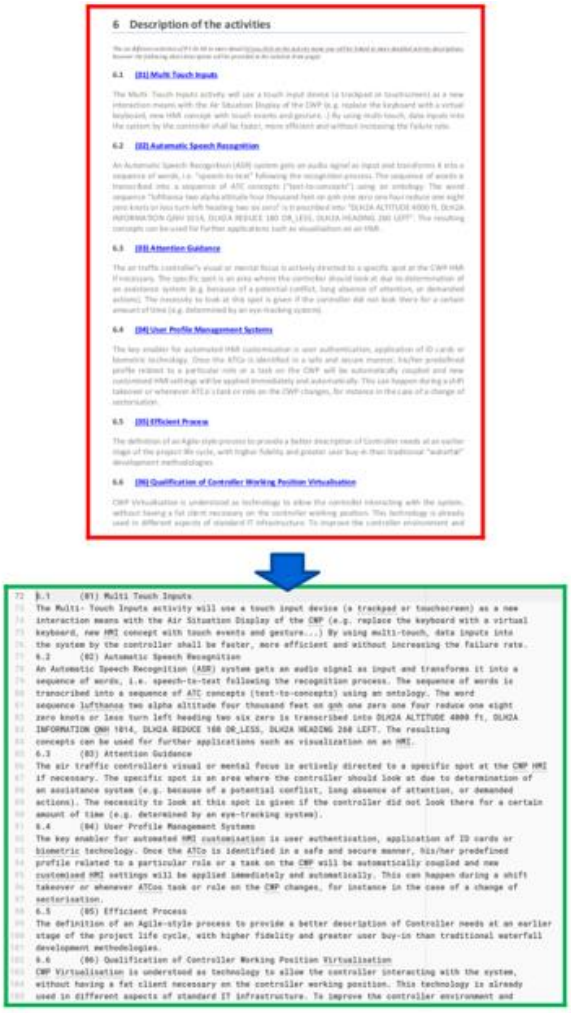

Figure 8: An example conversion of PDF page into plain text format

3.2.5 Natural language processing

Natural language processing (NLP) is a powerful text-mining approach for analysing text content with modern computing power. One of the fundamental uses of NLP is to extract key information from a large amount of text.

A customised NLP tool was developed by Engage. It uses the plain text generated from each report as input. Then, it is designed to filter out information that is not relevant to research (e.g. common verbs, punctuation, common nouns, etc.). Only information that could potentially be related to the research is kept. Finally, using statistical methods, the keywords of each report can be automatically generated.

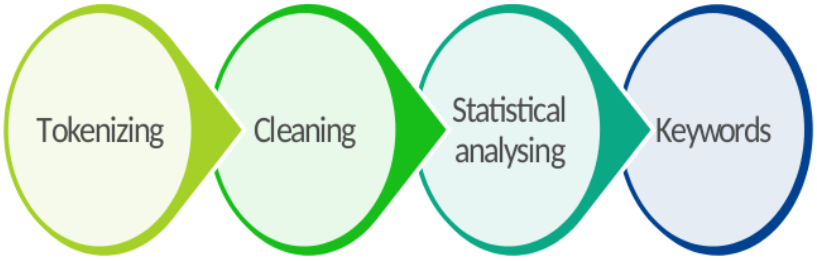

Figure 9: Three key process of the NLP process for information extraction

The key steps of the NLP approach are illustrated above, which are: tokenising, cleaning, and statistical analysing. Keywords representing the core information of each report are generated. To accomplish this, the open-source, natural language toolkit (NLTK) Python library was used. Based on this library, a customised tool was developed to perform the task of mining the reports.

3.2.6 Tokenising

In natural language processing, 'tokenisation' refers to the process of dividing the contents into pieces, which are called 'tokens'. They are often loosely referred to as 'words' or specific 'terms'. Several tokenisation techniques were tested, for example:

- **Treebank word tokeniser:** This is an efficient and fast method for tokenising natural languages. However, it has limited ability to handle combinations of words.
- **Multi-word expression tokeniser:** This method takes a string that has already been separated into tokens and retokenises it. It allows the merging of different words into a single token. This can be used as an extension to the previous method. However, it requires manual definition of the combination of these words, or requires training from the tokens.
- **Punkt sentence tokeniser:** On some occasions, it is also desired to first extract sentences before filtering out tags that are not an essential part of the content. The Punkt sentence tokeniser method can be used to accomplish such a task. However, it requires a machine learning model to be trained using a large quantity of text in the same language.

Once the text is tokenised using the previously defined tokeniser, they are represented by an array of tokens, which can be handled by a computer efficiently. In Figure 10, the tokenised content from the example text in the previous section is shown.

```
['6.1', '(', '01', ')', 'Multi', 'Touch', 'Inputs', 'The', 'Multi-', 'Touch', 'Inputs', 'activity', 'will', 'use', 'a',
'touch', 'input', 'device', '(', 'a', 'trackpad', 'or', 'touchscreen', ')', 'as', 'a', 'new', 'interaction', 'means',
'with', 'the', 'Air', 'Situation', 'Display', 'of', 'the', 'CWP', '(', 'e.g', '.', 'replace', 'the', 'keyboard', 'with',
'a', 'virtual', 'keyboard', '.', 'new', 'HMI', 'concept', 'with', 'touch', 'events', 'and', 'gesture', '...', ')', 'By',
'using', 'multi-touch', ',', 'data', 'inputs', 'into', 'the', 'system', 'by', 'the', 'controller', 'shall', 'be', 'faster',
',', 'more', 'efficient', 'and', 'without', 'increasing', 'the', 'failure', 'rate', '.', '6.2', '(', '02', ')', 'Automatic',
'Speech', 'Recognition', 'An', 'Automatic', 'Speech', 'Recognition', '(', 'ASR', ')', 'system', 'gets', 'an', 'audio',
'signal', 'as', 'input', 'and', 'transforms', 'it', 'into', 'a', 'sequence', 'of', 'words', ',', 'i.e', '.', 'speech-to-
text', 'following', 'the', 'recognition', 'process', '.', 'The', 'sequence', 'of', 'words', 'is', 'transcribed', 'into',
'a', 'sequence', 'of', 'ATC', 'concepts', '(', 'text-to-concepts', ')', 'using', 'an', 'ontology', '.', 'The', 'word',
'sequence', 'lufthansa', 'two', 'alpha', 'altitude', 'four', 'thousand', 'feet', 'on', 'qnh', 'one', 'zero', 'one', 'four',
'reduce', 'one', 'eight', 'zero', 'knots', 'or', 'less', 'turn', 'left', 'heading', 'two', 'six', 'zero', 'is',
'transcribed', 'into', 'DLH2A', 'ALTITUDE', '4000', 'ft', ',', 'DLH2A', 'INFORMATION', 'QNH', '1014', ',', 'DLH2A',
'REDUCE', '180', 'OR_LESS', ',', 'DLH2A', 'HEADING', '260', 'LEFT', '.', 'The', 'resulting', 'concepts', 'can', 'be',
'used', 'for', 'further', 'applications', 'such', 'as', 'visualization', 'on', 'an', 'HMI', '.', '6.3', '(', '03', ')',
'Attention', 'Guidance', 'The', 'air', 'traffic', 'controllers', 'visual', 'or', 'mental', 'focus', 'is', 'actively',
'directed', 'to', 'a', 'specific', 'spot', 'at', 'the', 'CWP', 'HMI', 'if', 'necessary', '.', 'The', 'specific', 'spot',
'is', 'an', 'area', 'where', 'the', 'controller', 'should', 'look', 'at', 'due', 'to', 'determination', 'of', 'an',
'assistance', 'system', '(', 'e.g', '.', 'because', 'of', 'a', 'potential', 'conflict', ',', 'long', 'absence', 'of',
'attention', ',', 'or', 'demanded', 'actions', ')', '.', 'The', 'necessity', 'to', 'look', 'at', 'this', 'spot', 'is',
'given', 'if', 'the', 'controller', 'did', 'not', 'look', 'there', 'for', 'a', 'certain', 'amount', 'of', 'time', '(',
'e.g', '.', 'determined', 'by', 'an', 'eye-tracking', 'system', ')', '.', '6.4', '(', '04', ')', 'User', 'Profile',
'Management', 'Systems', 'The', 'key', 'enabler', 'for', 'automated', 'HMI', 'customisation', 'is', 'user',
'authentication', ',', 'application', 'of', 'ID', 'cards', 'or', 'biometric', 'technology', '.', 'Once', 'the', 'ATCo',
'is', 'identified', 'in', 'a', 'safe', 'and', 'secure', 'manner', ',', 'his/her', 'predefined', 'profile', 'related', 'to',
'a', 'particular', 'role', 'or', 'a', 'task', 'on', 'the', 'CWP', 'will', 'be', 'automatically', 'coupled', 'and', 'new',
'customised', 'HMI', 'settings', 'will', 'be', 'applied', 'immediately', 'and', 'automatically', '.', 'This', 'can',
'happen', 'during', 'a', 'shift', 'takeover', 'or', 'whenever', 'ATCos', 'task', 'or', 'role', 'on', 'the', 'CWP',
'changes', ',', 'for', 'instance', 'in', 'the', 'case', 'of', 'a', 'change', 'of', 'sectorisation', '.']
```

Figure 10: An example of tokenised text

Observe that not all tokens are useful information for the knowledge mining process. For example, punctuation and common English words need to be filtered out in a further processing step.

3.2.7 Cleaning

The cleaning process is designed to continue the previous step to filter out information that is not related to the purpose of knowledge mapping. The following steps are taken to produce a clean list of tokens:

- Removing punctuation (for example: “.”, “,”, “:”, “(”, “)”, etc.)
- Removing tokens with lengths less than three characters (for example: “of”, “on”, “a”, “an”, etc.)
- Removing common English words (e.g.: “the”, “and”, “with”, etc.)
    - This module is designed based on the English language *stopwords* library from Python
- Converting words in plural form to singular form (for example.: inputs -> input, technologies -> technology, controllers -> controller)
    - This model is designed based on the *WordNet* English language lexical database, which is part of NLTK
- Removing words that are not a noun (such as adjectives and adverbs)

- This task is performed by employ the position tag identification using NLTK

- Removing very common words for SESAR research and publications (for example: “ATM”, “figure”, “aircraft”, etc.)
  - This is based on a customised set of stop words constructed by the team

- Generating common bigrams tokens from the single word tokens (for example: “trajectory” + “prediction” -> “trajectory_prediction”)
  - This task is accomplished by using the bigram generator using NLTK. Then the most common bigrams are used for the input of multi-word tokeniser to produce these combined tokens

After all previous measures are taken, a cleaned list of tokens can be generated. For example, in Figure 11, the final list of tokens from the previous example is shown.

```
['multi', 'touch', 'input', 'activity', 'input', 'device', 'trackpad', 'touchscreen',
'interaction', 'mean', 'situation', 'display', 'cwp', 'keyboard', 'keyboard', 'hmi', 'concept',
'event', 'gesture', 'data', 'input', 'system', 'controller', 'failure', 'rate', 'speech',
'recognition', 'speech', 'recognition', 'asr', 'system', 'signal', 'input', 'sequence', 'word',
'recognition', 'sequence', 'word', 'sequence', 'concept', 'text-to-concepts', 'ontology',
'word', 'sequence', 'foot', 'qnh', 'knot', 'turn', 'zero', 'altitude', 'information', 'qnh',
'or_less', 'dlh2a', 'left', 'concept', 'application', 'visualization', 'hmi', 'attention',
'guidance', 'controller', 'focus', 'spot', 'cwp', 'hmi', 'spot', 'area', 'controller',
'determination', 'assistance', 'system', 'conflict', 'absence', 'attention', 'action',
'necessity', 'spot', 'controller', 'amount', 'system', 'user', 'profile', 'management',
'system', 'enabler', 'hmi', 'customisation', 'authentication', 'application', 'card',
'technology', 'atco', 'manner', 'his/her', 'role', 'task', 'cwp', 'hmi', 'setting', 'shift',
'takeover', 'task', 'role', 'cwp', 'change', 'instance', 'case', 'change', 'sectorisation']
```

Figure 11: An example of cleaned tokens

Most of the remaining tokens are related to research content. It is already possible to see that some tokens appear more often than others from one page of content in a SESAR research deliverable.

3.2.8 Additional cleaning of personal content

After the common NLP process, names (from authors or references) can sometimes be identified as tokens. In rare cases, when a name appears multiple times, it can end up in the list of keywords (in the following step). In order to prevent sensitive information appearing in the research map, an additional cleaning step was introduced to remove any names in the results.

This process consists of both automatic and manual efforts. Common names are identified by a standard NLP library: Stanford Named Entity Recogniser (*StanfordNER*). Uncommon names are spotted and removed manually.

Often, names are mistaken as keywords when lists of references are placed among the deliverable content, rather than the end of the document. It is recommended that in the future SESAR deliverable, the reference lists should be placed at the end of the documents.

3.2.9 Statistical analysis

The previous example is based only on one page extracted from a deliverable. By combining tokens from all the pages, it is possible to construct a more accurate landscape of the research project. One of the most efficient ways to analyse these combined tokens is to use statistics. Frequency analysis is employed to extract the most important information.

In this way, it is also possible to visualise the results of the entire NLP approach. In Figure 12, the most common tokens and their frequency counts are shown. From this visualisation, one can instantly recognise the most common research elements indicated in the report.

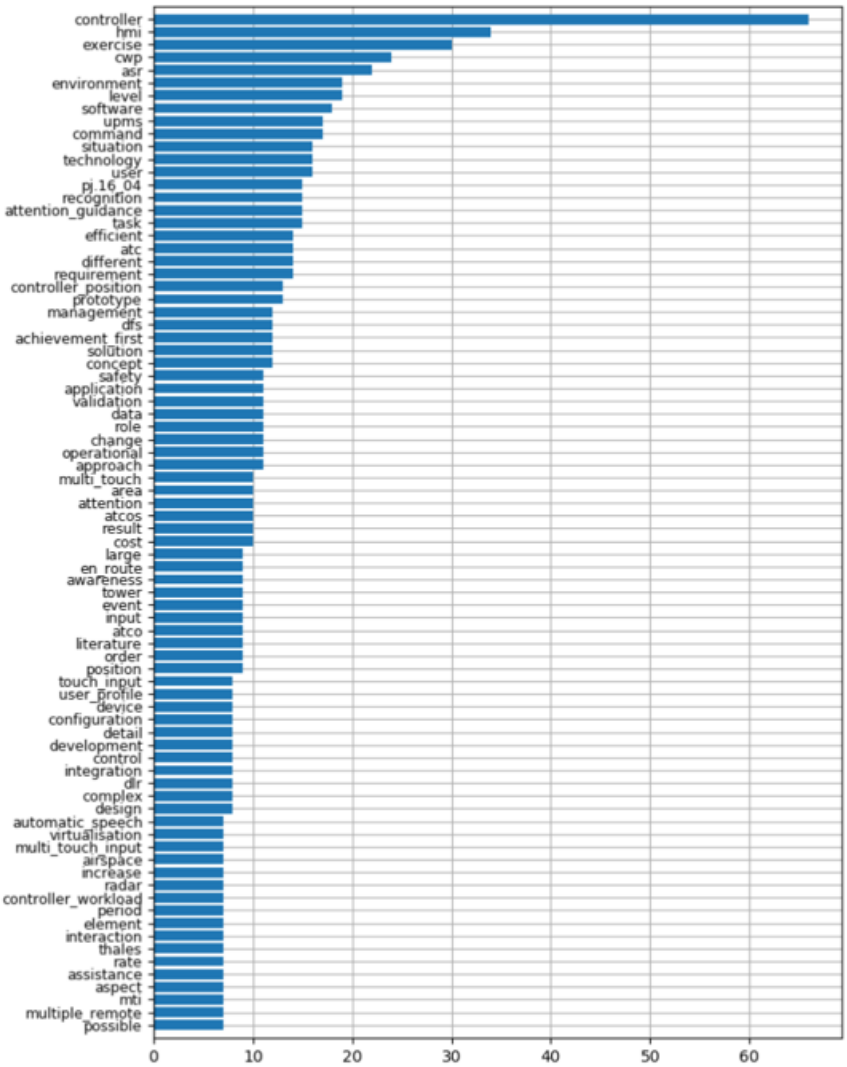

Figure 12: Final statistical analysis of the tokens in the example report

Most common 80 tokens and their frequency counts are shown

3.2.10 Keyword identification

Based on the previously generated statistics, the top most common tokens are extracted and used as the keywords of each output. In addition to the common tokens, normalised weights are calculated, according to the frequency of the term (i.e.: all weights add up to 100).

controller:6.10	hmi:3.14	exercise:2.77	cwp:2.22
asr:2.03	environment:1.76	level:1.76	software:1.66
upms:1.57	command:1.57	situation:1.48	technology:1.48
user:1.48	pj.16_04:1.39	recognition:1.39	attention_guidance:1.39
task:1.39	efficient:1.29	atc:1.29	different:1.29
requirement:1.29	controller_position:1.20	prototype:1.20	management:1.11
dfs:1.11	achievement_first:1.11	solution:1.11	concept:1.11
safety:1.02	application:1.02	validation:1.02	data:1.02
role:1.02	change:1.02	operational:1.02	approach:1.02
multi_touch:0.92	area:0.92	attention:0.92	atcos:0.92
result:0.92	cost:0.92	large:0.83	en_route:0.83
awareness:0.83	tower:0.83	event:0.83	input:0.83
atco:0.83	literature:0.83	order:0.83	position:0.83
touch_input:0.74	user_profile:0.74	device:0.74	configuration:0.74
detail:0.74	development:0.74	control:0.74	integration:0.74
dlr:0.74	complex:0.74	design:0.74	automatic_speech:0.65
virtualisation:0.65	multi_touch_input:0.65	airspace:0.65	increase:0.65
radar:0.65	controller_workload:0.65	period:0.65	element:0.65
interaction:0.65	thales:0.65	rate:0.65	assistance:0.65
aspect:0.65	mti:0.65	multiple_remote:0.65	possible:0.65
assurance:0.65	description:0.56	multi:0.56	milan:0.56
label:0.56	tower_environment:0.56	intelligence:0.56	cwpv:0.56
interface:0.56	performance:0.56	virtual:0.56	dlh2a:0.56
traditional:0.56	demonstrate:0.56	usability:0.56	prague_approach:0.56
operational_concept:0.56	ontology:0.56	voice:0.56	part:0.56

Figure 13: Final keyword identified in the example report

In Figure 13, the final list of keywords and their weights from the example SESAR deliverable are listed. For instance, it is easy to infer from the most frequent keywords that the report is focused on the human-machine interface (HMI) and controller working position (CWP) for controllers. With further analysis, it is possible to infer specific techniques addressed by the report, such as multi-touch input, automatic speech, visualisation, attention guidance, etc. A similar approach is applied for the SIDs papers.

3.2.11 Research clustering

The final step is to find and explore the connection between different research activities and outputs using the plain text extracted from all the outputs described in Section 2.1. A multi-dimensional map of all these outputs was generated using an unsupervised machine learning algorithm based on the combination of two key techniques:

- TF-IDF (term frequency–inverse document frequency) information retrieval;
- unsupervised K-means clustering.

Section 3.3 presents the resulting clusters, and Section 3.4 shows the content as live in the wiki.

3.3 Resulting clusters

In total, 14 clusters were identified, based on the similarities in project keywords. For each cluster, a name was selected manually based on expert judgement. The results are shown in Table 6. Note that these clusters may well change when the new outputs summarised in Section 2.2 are analysed. As noted in Table 12, the "large scale demonstrations" and "transversal projects" clusters are special cases.

Table 6: Research cluster names and associated keywords

Cluster No.	Cluster name	Cluster keywords (all in lower case, as in the wiki outputs)
1	ATC systems	requirement, block, status, title, controller, clearance, route, system, function, data, message, hmi, category, taxi, information, runway, compliance, rationale, trajectory, alert
2	RPAS	rpas, pilot, exercise, procedure, demonstration, controller, rpa, simulator, atco, mission, traffic, atc, airspace, operator, communication, intruder, scenario, safety, test, radar
3	Capacity on demand and dynamic airspace	trajectory, solution, process, airspace, constraint, network, validation, information, woc, step, exercise, dcb, system, mission, efpl, plan, data, maturity, concept, delay
4	CNS systems	surveillance, system, technology, navigation, test, gnss, requirement, data, ground, solution, performance, verification, application, message, service, communication, receiver, gps, activity, mode
5	Safety, resilience and automation	sector, airspace, controller, trajectory, conflict, system, tool, complexity, solution, scenario, traffic, route, modelling, data, automation, concept, uncertainty, level, operator, plan
6	U-space and urban air mobility	drone, service, mission, system, data, operator, information, communication, pilot, airspace, security, capability, scenario, access, navigation, risk, technology, position, area, requirement
7	**Transversal projects**	programme, activity, process, plan, contribution, framework, validation, development, platform, support, reference, material, system, step, edition, guidance, tool, concept, solution, member
8	AMAN, DMAN	sequence, aman, cta, horizon, controller, arrival, requirement, sector, tma, validation, departure, step, system, traffic, information, flow, solution, constraint, airport, concept
9	Airport operations	airport, aop, solution, validation, requirement, runway, gaming, capacity, plan, monitoring, performance, information, process, status, stakeholder, departure, alert, data, indicator, service
10	Optimisation and complexity techniques	airline, passenger, delay, cost, modelling, airport, network, capacity, ansps, mechanism, value, price, scenario, number, research, data, slot, airspace, trajectory, indicator
11	TMAs and separation	procedure, approach, controller, operator, airport, separation, solution, runway, system, validation, traffic, safety, condition, concept, atco, pilot, exercise, rnp, performance, scenario
12	Meteorology and services	forecast, product, weather, information, met, requirement, airport, observation, sensor, service, probability, temperature, system, data, user, decision, value, parameter, surface, precipitation
13	**Large scale demonstrations**	trial, demonstration, fuel, regulation, exercise, airline, delay, procedure, airport, efficiency, operator, tta, traffic, route, capacity, reduction, arrival, data, assessment, period
14	SWIM	swim, service, supervision, architecture, security, information, infrastructure, registry, specification, verification, requirement, contribution, provision, description, function, governance, technology, system, modelling, message

3.4 Interactive research map – live in the wiki

Live link to wiki:	https://wikiengagektn.com/EngageWiki:Interactive_research_map_of_ATM

In Figure 14, the clustering result is shown, as presented in the wiki. Clusters are represented in different colours. Figure 15 shows the map with a filter applied.

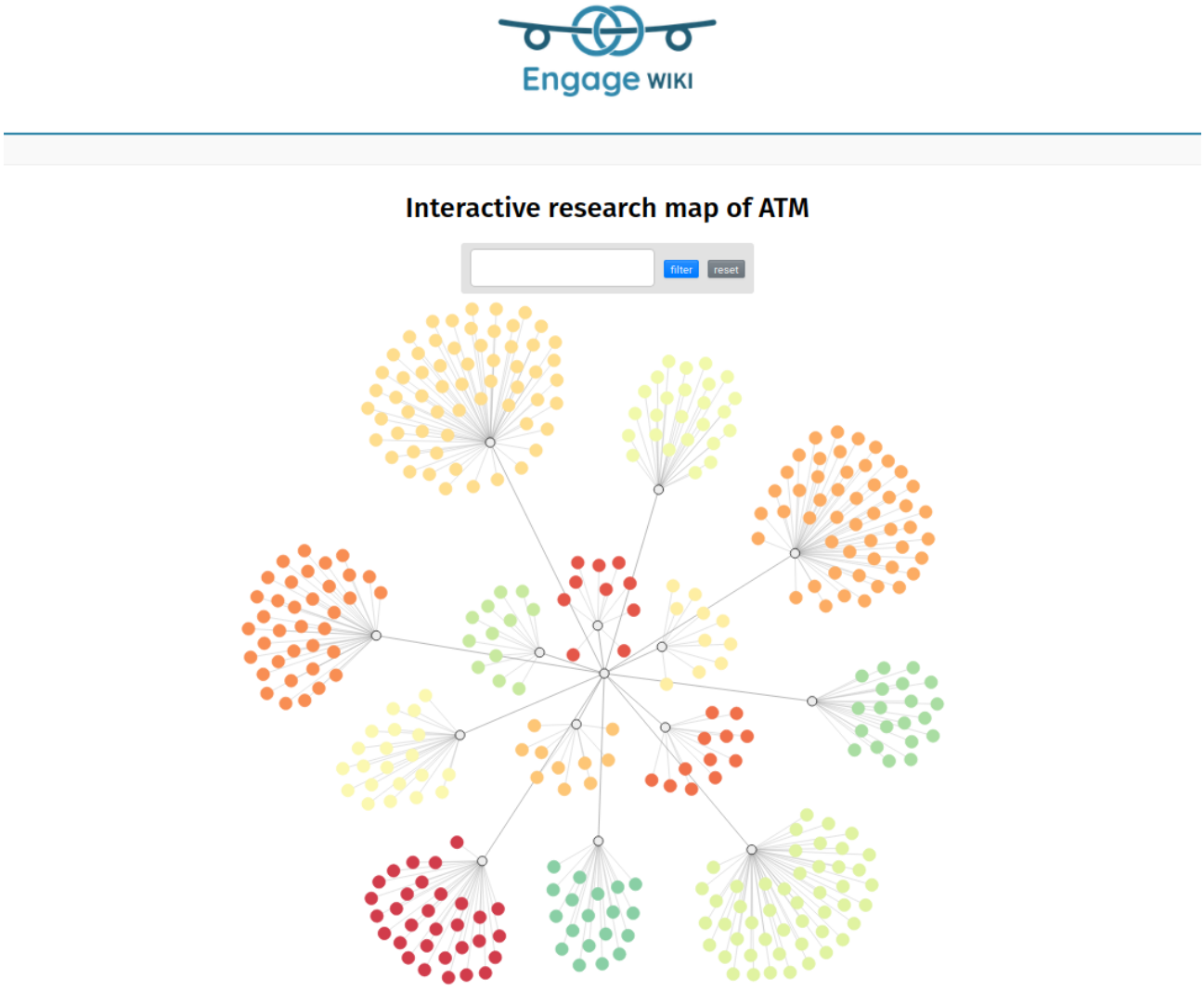

Figure 14: Research mapping clusters – all clusters shown

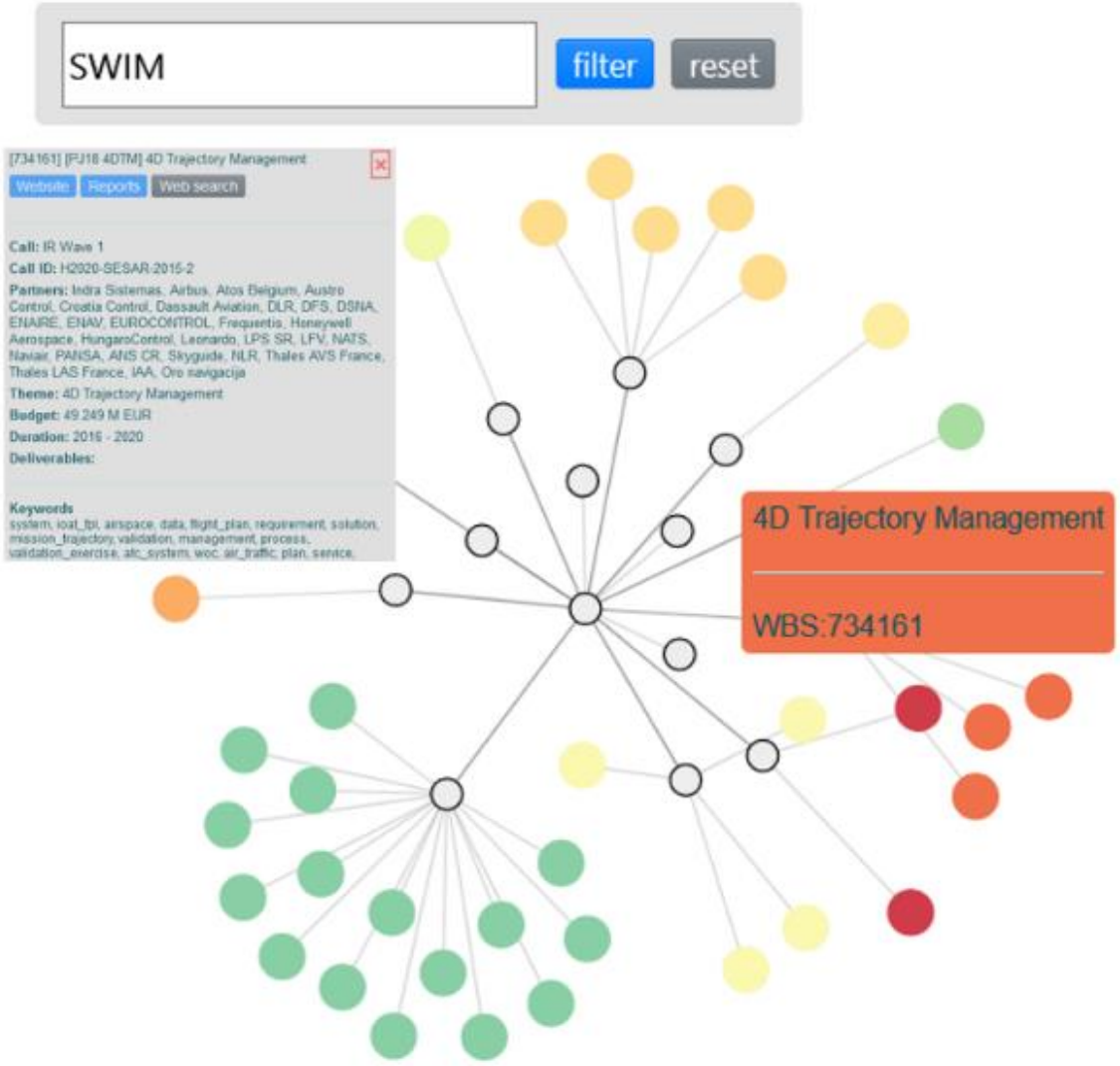

Figure 15: Research mapping clusters – with a filter ("SWIM") applied (illustrative)

Table 7 and Table 8 show the corresponding, explanatory texts in the wiki.

Table 7: Short summary text in the wiki under the interactive research map

Short summary text in the wiki under the interactive research map (links below are non-functional)
Deliverables published by SESAR 1 and SESAR 2020 projects were sourced as an input to this research mapping task. The mapping employs a bottom-up approach for knowledge mining and classifying the research. Natural language processing was applied to extract keywords and frequencies from each project, followed by a machine learning algorithm to generate project clusters based on their similarities. For each project, click on the node to see key information. The tool allows users to filter projects by specific terms (e.g. “SWIM”). We continue to add features suggested by our use case survey, and welcome your feedback. Select ‘Page ... Discussion ...’ to get involved. Click here for a fuller explanatory text.

Table 8: Longer summary text in the wiki complementing the interactive research map

Longer summary text in the wiki complementing the interactive research map (links below are non-functional)
Deliverables published by SESAR 1 and SESAR 2020 projects were sourced as an input to the research mapping task. Whilst our phase 1 data coverage of all potentially available deliverables is not complete, all the main industrial and exploratory research Calls, such as WP-E, ER1-ER3, IR demonstrations, IR RPAS and IR Wave 1 are represented. These have been supplemented with papers from SESAR Innovation Days conferences (2011-2019) and material from the previous ComplexWorld network. Supporting metadata have also been prepared by the Engage team. This is still work in progress and provides a high-level summary of each project. For each project, click on the node to see key information (where available), including the: Call (e.g. “IR Wave 1”), project partners, theme (e.g. “4D trajectory management”), budget, duration, deliverable(s), project keywords, and URL). WBS is the 'Work Breakdown Structure' reference - a unique identifier for activities within the SESAR Programme, from Exploratory Research through to Very Large Scale Demonstrations. By the end of our phase 1 data collection, almost 1500 deliverables have been sourced. The mapping itself employs a bottom-up approach for knowledge mining and classifying these SESAR research project sources. By gathering information in the form of keywords from project documentation and subsequently applying clustering algorithms, we constructed an overview, as a tree, of the projects and their interconnections. The mapping process is based on several key steps. First, the pre-processing step converted pdf and doc files to raw text format. Then, natural language processing (NLP) was applied to extract keywords and frequencies from each project. A customised, unsupervised machine learning algorithm was next applied to generate project clusters based on their similarities. Finally, we created the online research map visualisation tool you now see. This allows the user to explore the research map interactively. The tool also allows users to filter projects by specific terms. In 2021, our phase 2 data sourcing will analyse the remaining public deliverables, primarily produced by SESAR 2020 projects funded through the recent ER4 and IR Waves 2-3 Calls. The missing metadata from older SESAR 1 projects will also be completed (as far as possible), for instance, 70% of these older projects have yet to be populated with the names of consortium partners. Papers from the 2020 SESAR Innovation Days conference, and potentially other relevant conferences, will also be added. After this, the clusters may well be adapted. We continue to add features suggested by our use case survey, and welcome your feedback. Click here to get involved. This mapping also feeds the ATM concepts roadmap.

4 ATM concepts roadmap

4.1 Motivation and overview

The goal of building the ATM concepts roadmap is to map previous research activities in the field and, moreover, to map and identify future research concepts. This is based on the current research status, past coverage and apparent gaps, also taking related strategic research agendas into account. The methodology adopted is described in the following steps:

1. compiling a list of strategic research and innovation documents to analyse; (the complete list of the consulted documents, used to date, can be found in Section 8, [3]-[23]);
2. defining ATM-related research concepts up to 2040, in line with the timeline described in the European ATM Master Plan [3] (see Section 4.2);
3. analysing future research directions from other disciplines in as much as they are, or could be, linked to ATM (see Section 4.4);
4. an *initial* delineation of further, future research focus areas – challenges for 2040 and beyond (see Section 4.5).

4.2 SRIA flagship initiatives

In this first step, ATM and aviation-related strategic documents were reviewed, starting with the 2020 edition of the European ATM Master Plan [3], SESAR's 'airspace architecture study' [20], Flightpath 2050 [21], and the ACARE Clean Aviation SRIA [5]. During the review, the Strategic Research and Innovation Agenda, Digital European Sky [4] (henceforth simply 'the' SRIA) was published, presenting the agenda for the future Integrated ATM partnership, the SESAR successor. The SRIA identifies nine flagship activities/roadmaps for the next multiannual financial framework (MFF) period. Each activity identifies the research, innovation, and in some cases the deployment needs, up to 2040. The Engage consortium adopted the nine flagship activities in its concepts roadmap.

Table 9: Excerpt from SRIA, Digital European Sky [4]

SRIA flagship initiative	Description
Connected and automated ATM	Europe's ATM infrastructure operates with low levels of automation support and data exchange, leading to rigidity, lack of scalability and resilience, and an inability to exploit emerging digital technologies, including in support of new airspace users. The future architecture of the European sky requires increased automation in air traffic control and an infrastructure commensurate with the performance required by each airspace user type and environment, including those in the transition areas between Europe and neighbouring ICAO regions which may have specific regulations and challenges.
Air-ground integration and autonomy	Current ATM systems and technologies are not designed to allow the accommodation or full integration of an increasing number of new forms of mobility and air vehicles which have a high degree of autonomy and use digital means of communication and navigation. The future ATM needs to evolve, exploiting existing technologies as much as possible, and developing new ones in order to increase global ATM performance in terms of capacity, operational efficiency and accommodation of new and/or more autonomous air vehicles, i.e. supporting

	the evolving demand in terms of diversity, complexity from very low-level airspace to high level operations. This progressive move towards autonomous flying, enabled by self-piloting technologies, requires closer integration and advanced means of communication between vehicle and infrastructure capabilities so that the infrastructure can act as a digital twin of the aircraft. Ultimately, manned and unmanned aerial vehicles should operate in a seamless and safe environment using common infrastructure and services supporting a common concept of trajectory-based operations. Future operations should therefore rely on direct interactions between air and ground automation, with the human role focused on strategic decision-making while monitoring automation.
Capacity-on-demand and dynamic airspace	For the last decades, capacity has not been available when and where needed and it has often been available when and where not needed. New airspace users including RPAS/HAO traffic will increase by 2030 and will require an increased level of capacity and its variability. Integrated Air Traffic Management- requires agility and flexibility in providing capacity where and when it is needed, particularly for maximising the use and performance of limited resources, i.e. airspace and ATCOs. It will require the dynamic reconfiguration of resources and new capacity-on-demand services to maintain safe, resilient, smooth and efficient air transport operations while allowing for the optimisation of trajectories even at busy periods.
U-space and urban air mobility	Over the next 10 years, the implementation of this SRIA aims to unlock the potential of the drone economy and enable urban air mobility (UAM) on a wide scale. To that end, a new air traffic management concept for low-altitude operations needs to be put in place to cater safely for the unprecedented complexity and high volume of the operations that are expected. This concept, referred to as U-space, will include new digital services and operational procedures and its development has already started within the SESAR 2020 programme. U-space is expected to provide the means to manage safely and efficiently high-density traffic at low altitudes involving heterogeneous vehicles (small unmanned aerial vehicles, electric vertical take-off and landing – eVTOLs - and conventional manned aircraft), including operations over populated areas and within controlled airspace. U-space will have to integrate seamlessly with the ATM system to ensure safe and fair access to airspace for all airspace users, including UAM flights departing from airports.
Virtualisation and cyber-secure data sharing	The Airspace Architecture Study (AAS) clearly highlighted the lack of flexibility in the sector configuration capabilities at pan-European level. This is caused by the close coupling of ATM service provision to the ATS systems and operational procedures, preventing air traffic from making use of cloud-based data service provision. A more flexible use of external data services, considering data properties and access rights, would allow the infrastructure to be rationalised, reducing the related costs. It will enable data sharing, foster a more dynamic airspace management and ATM service provision, allowing air traffic service units (ATSU) to improve capacity in portions of airspace where traffic demand exceeds the available capacity. It furthermore offers options for the contingency of operations and the resilience of ATM service provision.
Multimodality and passenger experience	Flightpath 2050, Europe's long-term vision document on aviation research, has set the goal that 90% of travellers within Europe should be able to complete their journey, door-to-door (D2D), within 4 hours by 2050. Optimising D2D mobility for people and goods is essential in meeting citizens' expectations for increasingly seamless mobility, where they can rely on the predictability of every planned door-to-door journey and can choose how to optimise it (shortest travel time, least cost, minimal environmental impact, etc.). The role of ATM in the door-to-door chain of a passenger's journey may seem small, but the punctuality of flights, and passengers' perception of flying, is highly dependent on the smooth functioning of the entire journey. This SRIA will, therefore, lead to an improved passenger experience by supporting an integrated transport system.

Aviation Green Deal	The objective of net-zero greenhouse gas emissions by 2050 set by the European Green Deal, in line with the EU's commitment to global climate action under the Paris Agreement, requires accelerating the shift to smarter and more sustainable mobility. This implies the need for aviation to intensify its efforts to reduce emissions, in line with the targets set in Flightpath 2050. To this end, a set of operational measures to improve the fuel efficiency of flights will have to be put in place. At the same time, to ensure sustainable air traffic growth, it is necessary to speed up the modernisation of the air infrastructure to offer more capability and capacity, making it more resilient to future traffic demand and adaptable through more flexible air traffic management procedures and a charging scheme that does not make it interesting to fly unnecessary distance. Furthermore, reducing aircraft noise impacts and improving air quality will remain a priority around airports.
Artificial intelligence (AI) for aviation	ATM decision-support techniques, mostly based on heuristics, present limitations in terms of the technology itself. Hence the performance improvements of the future cannot be achieved using legacy software system approaches. AI is one of the main enablers to overcome the current limitations in the ATM system. A new field of opportunities arises from the general introduction of AI, enabling higher levels of automation and impacting the ATM system in different ways. The FLY AI report provides a set of recommendations and real examples to help the aviation/ATM sector accelerate the uptake of AI. AI can identify patterns in complex real-world data that human and conventional computer assisted analyses struggle to identify, can identify events and can provide support in decision making, even optimisation. Over recent years, developments and applications of AI have shown that it is a key ally in overcoming these present-day limitations, as in other domains. Tomorrow's aviation infrastructure will be more data-intensive and thanks to the application of Machine Learning (ML), deep learning and big data analytics aviation practitioners will be able to design an ATM system that is smarter and safer, by constantly analysing and learning from the ATM ecosystem.
Civil/military interoperability and coordination	The digital transformation of the European ATM network will have an impact on both civil and military aviation and ATM operations. Care must be taken to ensure a sufficient level of civil/military interoperability and coordination, especially concerning trajectory and airspace information exchange, as well as the use of interoperable CNS technologies. Therefore, a joint and cooperative civil-military approach to ATM modernisation would be the best choice to achieve the appropriate level of interoperability, also maximising synergies between civil and military research and development activities.

4.3 Linking past and future research concepts

Based on the work previously carried out in the interactive research mapping (see Section 3), and the information available from the nine SRIA flagship initiatives, the problem of linking past and future research concepts was posed in such a way that the objective was to find out for each project in our current database (as summarised in Section 2.1), to which of the flagship initiatives it most related. Different methodologies were researched in order to achieve this, but due to the high volume of projects and in order to have a working methodology as impartial and unbiased as possible, it was decided to use current state-of-the-art natural language processing (NLP), unsupervised algorithms. These algorithms try to create this link automatically and independently of human input. The objective of the selected algorithm was to rank the similarity of the projects with the flagship initiatives. To achieve this, the algorithm should be able to compare the text extracted from each of the projects with the descriptive text of the SRIA flagship initiatives, obtaining a measure of their similarity. The

descriptive texts of the SRIA [4] flagship initiatives ("Problem statement"; "Description of high-level R&I needs/challenges" and "Expected high-level outcomes and performance objectives") were extracted manually and then pre-processed in a similar way to the texts extracted from the projects. Two main types of similarity algorithms were selected and explored: term frequency-inverse document frequency (TF-idf) and semantic similarity. The first one, TF-idf, is a simple but fast algorithm that looks at words that appear in both texts and scores their similarity based on how often they appear. It is a useful algorithm if it is expected that the same words will appear in both texts, but some words are more important than others. The main shortcoming of this algorithm is that it is not very useful when the texts are short, and it requires a high overlap between the words of both texts. After some tests, the performance of the algorithm was not sufficient and it was decided not to use it further. The second one, semantic similarity, scores the similarity of the texts based on how similar their words are, even if they are not exact matches. These algorithms are more complex but have a better performance in cases where there are short texts or with little word overlap. It was decided to use this algorithm for the creation of the ATM concepts roadmap.

As the semantic similarity algorithm is based on the similarity of words, the first step was to create a language model. Word embeddings are a type of language model that allows words with similar meaning to have a similar representation in a multi-dimensional space. Individual words are represented as real-valued vectors in a predefined vector space which is learned based on the usage of words in training texts. There are already pre-trained word embeddings models created by Google and Facebook, trained with millions of news and/or Wikipedia articles. Their use was assessed but it was decided that they were not suitable as they did not capture well the complexity and specificities of the language related to air transport and the aviation industry in general.

It was therefore decided to create a specific word embedding model using the text of the projects, that of the SRIA flagship initiatives, and other extracted projects[1] (in total, more than 1 million sentences). The word embedding model used was *Word2Vec*. The *Word2Vec* statistical method is an efficient stand-alone learning model that can be used to create word embedding from a text corpus. Once the specific word embedding was trained, it was used to create a similarity matrix that contained the similarity between each pair of words (both of the project and the SRIA flagship initiatives), weighted using the term frequency. Using this similarity matrix, we calculated the soft cosine similarity between the nine SRIA flagship initiative texts and the project texts. The cosine similarity is a useful metric used to measure how similar texts are. It measures the cosine of the angle between two vectors projected in a multi-dimensional space. The cosine similarity is useful as it is non-dependent on the size of the texts. The smaller the angle between the two vectors the higher the cosine similarity. The soft cosine similarity used is a better implementation of the metric as it also takes into account the word similarity. As an end result, a semantic similarity index was obtained for each of the projects in our database with respect to the descriptions of the nine SRIA flagship activities. The projects were then tagged with the SRIA flagship activity with the highest similarity index.

[1] As in Section 3, we also include SIDs papers in this context. (See also Section 2.1.3.)

4.4 Complementary research in other programmes

In the next step, we expanded the analysis to future research directions from other disciplines in as much as they are, or could be, linked to ATM. Table 10 lists the concepts from other disciplines / programmes that are linked with future ATM research.

Table 10: Concepts linked to ATM covered in other programmes

No.	Concept	Description	SRIA flagship initiative link
1	Sustainable aviation fuels	Fully de-carbonised fuels like power-to-liquid synthetic fuels, methane and/or hydrogen.	Aviation green deal
2	Evolutionary propulsion	Reduction of fuel burn through new engine technologies, such as higher by-pass-ratios than previous engines.	Aviation green deal
3	Revolutionary propulsion	Revolutionary engines like open-rotor design, boundary-layer ingestion and electric aircraft are expected to bring significant fuel and emissions savings.	Aviation green deal
4	New airframe configurations	New airframes that are more fuel efficient, and quieter are in development. Design concepts like strut-braced wing, blended wing body, double-bubble and box-joined-wing aircraft. This is coupled with the development of revolutionary materials that would improve weight efficiency and increase aircraft performance.	Aviation green deal
5	Civil drones regulation and certification	Designing a framework to safely operate drones and allowing this industry to remain agile, defining the technical and operational requirements drones need to have.	U-space/Virtualisation and cyber-secure data sharing
6	Urban air mobility	Innovations for air vehicles and frameworks enabling better environmental and operative efficiency, new services, larger networks, optimised frequency and new business opportunities in the urban mobility setting.	U-space
7	Cybersecurity	Advances in cybersecurity, promotion, regulatory activities, and international cooperation aimed at incorporating cybersecurity in aviation systems.	Virtualisation and cyber-secure data sharing
8	Mobility as a service	Providing seamless access to all mobility services contained in multi-modal journeys, coordinating with similar initiatives from other sectors to foster collaboration and coordination.	Multimodality and passenger experience
9	Artificial intelligence advances, regulation and certification	Recommendations on future policy developments, on ethical, legal, societal issues and socio-economic challenges related to AI.	Artificial intelligence for aviation/Virtualisation and cyber-secure data sharing
10	Monitoring climate	Climate monitoring, including observations (ground and satellite-based), and numerical models aimed at predicting climate impacts at seasonal, decadal, and longer term scales.	Aviation green deal

The listed concepts aim at developing areas that are directly, or indirectly, related to ATM. The 'SRIA flagship initiative link' column shows the name of the SRIA flagship initiative on which the identified concept would have the main impact. For example, the main impact of 'sustainable aviation fuels' (to be further developed within the future Clean Aviation partnership (see: https://clean-aviation.eu/)), would be on the 'aviation green deal', as the goal is to decrease greenhouse gas emissions. The first four concepts in Table 10 are the result of the review of the documents describing novel aircraft technologies in the pipeline, namely the Clean Aviation SRIA [5], and IATA's Technology Roadmap [7].

The next concept, civil drones regulation and certification, is a result of the review of both European and US documents, namely U-space [10], the European Drones Outlook Study [11], and NextGen's UTM concept of operations [12] (in addition to the European ATM master Plan [3]). The urban air mobility concept is raised in the European ATM Master Plan [3], the SESAR SRIA [4], also features in other envisioned Horizon Europe programmes, such as the Clean Aviation SRIA [5], and is likely to continue in the main cluster of Horizon Europe [13] dedicated to climate, energy and mobility: as in the current H2020 programme, 'Mobility for Growth' Calls are addressing urban air mobility [22].

The cybersecurity concept is relatively new in the main ATM domain and it will likely become more important if the 'airspace architecture study' [20] objectives are be achieved. The importance of the issues is mentioned across different bodies related to aviation safety and security, such as European cybersecurity policies [18] and EASA's dedicated cybersecurity body [19]. The developments in this area outside ATM should be integrated with the required regulatory and policy requirements. Artificial intelligence advances, with its regulation and certification, is a concept that is also relatively new in ATM, and the possible impacts and required research is mentioned in the SESAR SRIA [4] and Fly AI report [9], for example.

The 'Mobility as a Service' concept has often been mentioned in various transport modes (e.g. the Rail SRIA [6], the SESAR SRIA [4], and by the MaaS Alliance [23]) and as it is inherently a multimodal concept, coordination and collaboration across different modes will be needed.

Climate change is a global issue of increasingly pressing importance and with growing regulatory commitments. There are developments in various specific, and generic, domains, such as the development of monitoring capabilities by the European Space Agency ([15], [16]), and the global climate reports from the Intergovernmental Panel on Climate Change [17]. These should be followed, in order to include the findings in future ATM solutions and research initiatives.

As the proposal for Horizon Europe in the next MFF required the rationalisation and streamlining of all initiatives, these multidisciplinary links have been identified at the programme level, to be able to delineate better the area of each initiative. As a result, most of these links are mentioned in the SRIA, as well as in SRIAs of other initiatives, under "dependencies with other initiatives". Some of the concepts above would have an impact on ATM, such as, for example new aircraft configurations, or new propulsion means. Others are needed for further development and deployment, but are out of scope of a strictly research and innovation (R&I) remit (e.g. up to TRL6), such as civil drones regulation and certification. However, for the quicker deployment a closer cooperation between the R&I and regulatory or standardisation bodies is needed. The concepts presented in Table 10 are outside the main ATM R&I, under the remit of other initiatives, and care should be taken to maintain close collaboration on these topics for harmonised R&I and to speed up the subsequent deployment of developed solutions. As they are mentioned in various SRIAs, it is likely that they will be addressed in the next MFF period, with the goal of becoming operational by 2040.

4.5 Candidate challenges – future research focus areas

The previous section discussed research variously complementary to ATM, in other programmes. A key remit of the concepts roadmap activity is to identify future research focus areas of importance for 2040 and beyond. We have initially seeded the roadmap with example ideas, as shown on the right of Figure 16 and in Table 11. As expected, these challenges overlap to some extent, and variously relate to the SRIA [3] activities (although the links shown in the figure are currently illustrative only). These seed concepts are not addressed with a strong focus in the SRIA *per se*, although they are variously related to existing elements therein, and may thus be future research focus areas.

Table 11: Candidate challenges – future research focus, as seeded in the wiki roadmap

No.	Summary concept	Fuller text
1	**Smart cities, with dynamic travel solutions**	Integrated suburban transport (incl. airports) is improved through digital services and open data governance, embracing greener, door-to-door mobility and ageing populations. Travel service providers, integrators, technology companies and associations (e.g. IATA), support a significant growth in wider, intermodal travel solutions through digital, dynamic ticketing and services: aka (virtual) interlining.
2	**Global warming shift**	A marked shift takes place in previously incremental environmental policies and regulations (e.g.: limiting the operation of short-haul air services; increased scoping of carbon taxes, further extended to include new NO_x regulations) - with a concomitant, increased demand for associated transport and monitoring technologies.
3	**Complexity, emergence, resilience**	Increasing, multi-system integration and dependencies drives a renewed demand for complexity science to model and understand (positive and negative) emergence in transport, with a focus on service resilience and its corresponding metrics, and cybersecurity playing a key role in the delivery of resilient solutions.

These future research focus areas, challenges for 2040 and beyond, will be updated and further populated through continued research in 2021, drawing on outputs from Engage, more widely in SESAR, and beyond ATM, through wiki user inputs and wider programme consultation, including the assessment of interdisciplinary concepts.

The key next steps in this process are presented in Section 7.

4.6 ATM concepts roadmap – live in the wiki

Live link to wiki: https://wikiengagektn.com/EngageWiki:ATM_concepts_roadmap

Figure 16 shows the ATM concepts roadmap as it appears in the wiki. As noted, the "large scale demonstrations" and "transversal projects" clusters are special cases.

Table 7 and Table 8 show the corresponding, explanatory texts in the wiki, for example giving details of the interactive features illustrated in Figure 17.

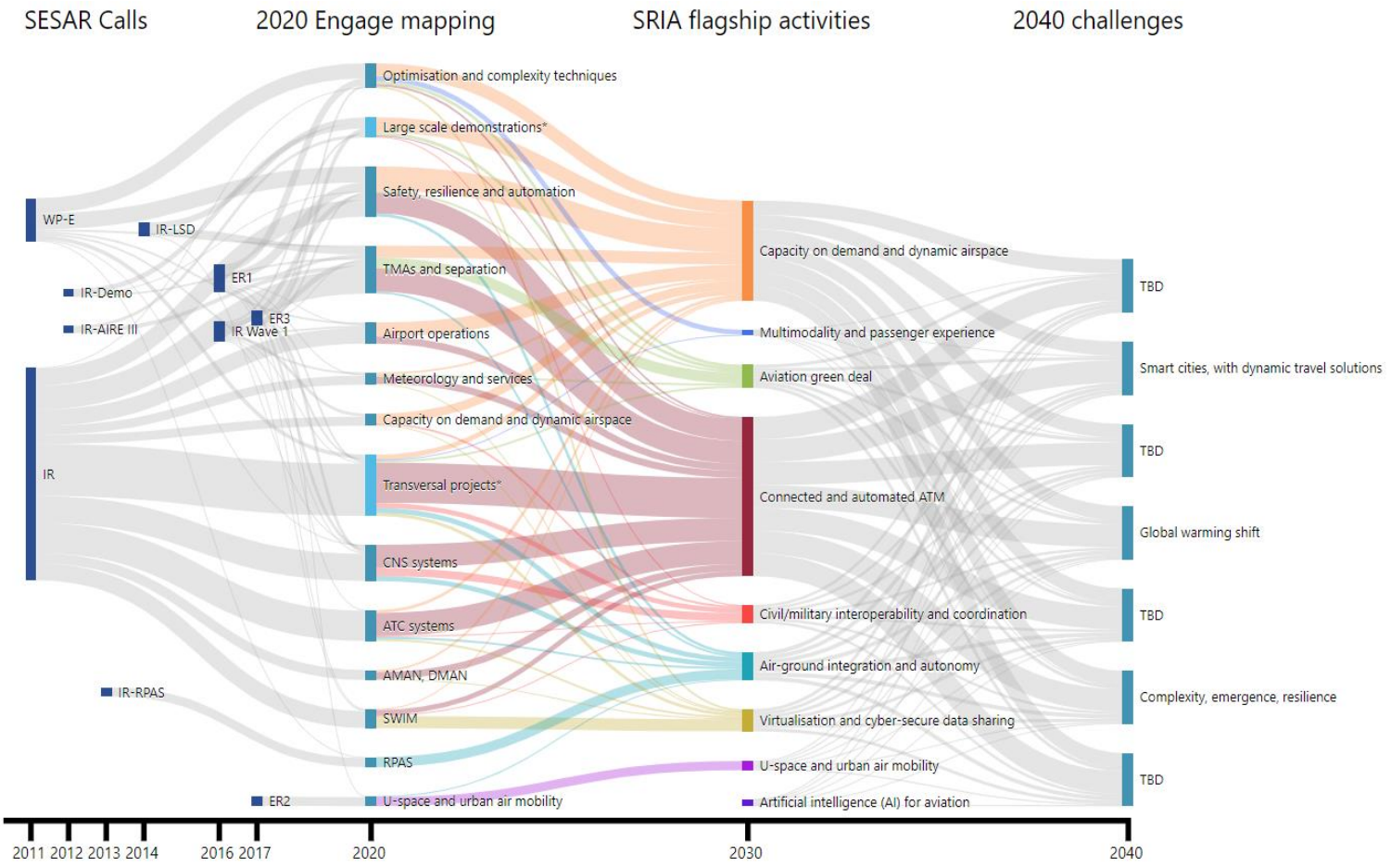

Figure 16: ATM concepts roadmap

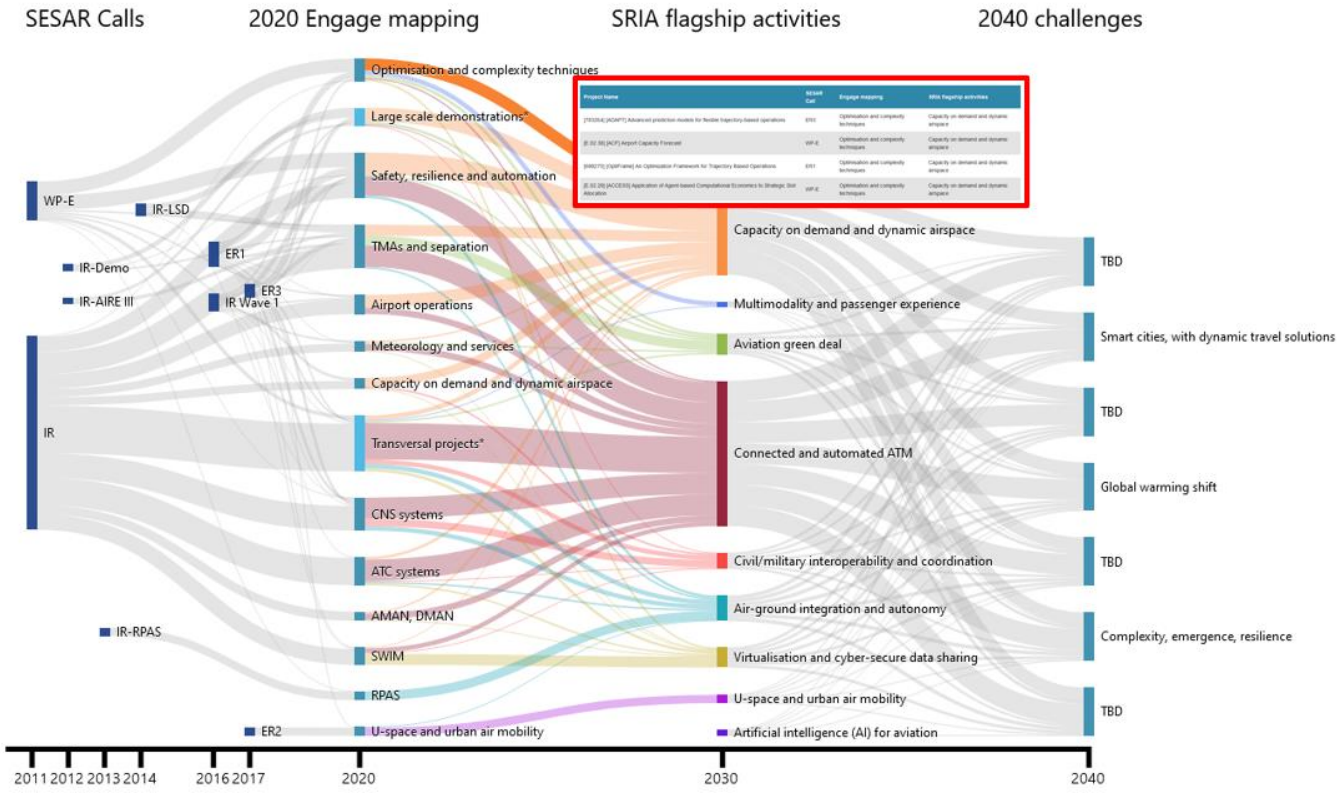

Figure 17: ATM concepts roadmap – with a link-table generated (illustrative)

Table 12: Short summary text in the wiki under the ATM concepts roadmap

Short summary text in the wiki under the ATM concepts roadmap (links below are non-functional)
* The "large scale demonstrations" and "transversal projects" clusters are special cases. These may be represented differently in future versions of the map. The ATM concepts roadmap shows how previous (SESAR) research connects with the Strategic Research and Innovation Agenda roadmaps, via the Engage research map. We used natural language processing and semantic similarity mapping to build this representation. When hovering over the different nodes and links, a tooltip will appear with some basic information. When further clicking on a link, a table appears below the visualisation with more detailed information of the projects in the selected link. There are four stages in the flow corresponding to the partly figurative timeline. The last stage, "2040 challenges", is the future concepts area of the roadmap. These are currently temporary challenges, as examples of topics addressed in less detail in the comprehensive 2020 SRIA. These challenges will be updated and further populated through continued research in 2021, drawing on outputs from Engage, more widely in SESAR and beyond ATM, through wiki user inputs, and including interdisciplinary concepts. We continue to add features suggested by our use case survey, and welcome your feedback. Select 'Page ... Discussion ...' to get involved. Click here for a fuller explanatory text.

Table 13: Longer summary text in the wiki complementing the ATM concepts roadmap

Longer summary text in the wiki complementing the ATM concepts roadmap (links below are non-functional)
The ATM concepts roadmap shows how previous (SESAR) research connects with the Strategic Research and Innovation Agenda roadmaps (described in the 2020 SRIA). This mapping has been built on the work of the Engage research map. Pre-processed project information, as an output from the Engage research map, as well as textual descriptions of the nine SRIA roadmaps ('flagship activities'), were used to create a word-embedding model through natural language processing (NLP) techniques, in order to capture semantic and syntactic similarities. The word-embedding model was used to perform a document semantic similarity analysis, instead of a more traditional lexical similarity analysis. As a result, a semantic similarity index was obtained for each of the projects in our database with respect to the descriptions of the nine SRIA flagship activities. The projects were then tagged with the SRIA flagship activity with the highest similarity index. In the ATM concepts roadmap the x-axis flow represents the (partly figurative) timeline progression of projects and concepts. The size of the nodes and links in the visualisation are directly proportional to the number of projects encompassed. The visualisation is interactive. When hovering over the different nodes and links, a tooltip will appear with some basic information. When further clicking on a link, a table appears below the visualisation with more detailed information of the projects in the selected link. There are four stages in the flow corresponding to the timeline. “SESAR Calls” represents the funded SESAR programmes and their project start dates. “2020 Engage mapping” corresponds to the NLP, unsupervised clustering performed for the Engage research map and helps to visualise how the different SESAR programmes feed into the identified research clusters. “SRIA flagship activities” is related to the semantic similarity analysis mentioned above and shows the flow of the projects from the identified research clusters and their similarity with the nine innovation roadmaps. The last stage, “2040 challenges”, is the future concepts area of the roadmap. These are currently temporary challenges, as examples of topics addressed in less detail in the comprehensive 2020 SRIA. They therefore currently reflect impact areas that are outside typical ATM research. These challenges will be updated and further populated through continued research in 2021, drawing on outputs from Engage, more widely in SESAR and beyond ATM, through wiki user inputs, and including interdisciplinary concepts. As expected, these '2040' challenges overlap to some extent, and variously relate to the SRIA activities (although the links shown in the figure are currently illustrative only). We continue to add features suggested by our use case survey, and welcome your feedback. Click here to get involved.

5 Technical update on the wiki migration and launch

5.1 Previous status of the wiki

The starting point for the Engage wiki was to recover the content of the ComplexWorld project wiki (CWW). The wide content already created in this wiki during the ComplexWorld project was considered to be useful for the new Engage wiki, as well as the set up and configuration of some of the features of the MediaWiki platform[2]. Consequently, Innaxis started a handover from the ComplexWorld to the Engage wiki, in September 2019.

Considering that the project ComplexWorld finished in 2016, the CWW had not been maintained for some months. ComplexWorld committed to maintain the wiki for 2 years after the project end date in order to enable a handover to the next SESAR network but the call for networks finally took much longer than expected. This period with no maintenance created some issues with the installation of the MediaWiki platform, including an internal virus that made it impossible for some time to move it to a new server exclusively dedicated to the Engage project. This led to Task 1, as explained below.

5.1.1 Task 1: ComplexWorld wiki content recovery

The CWW included valuable content on WP-E projects (a single page per project with all its results, publications and outcomes), agendas and videos of the events organised, results and material from the PhD theses co-financed by SESAR, etc. It was considered good practice to make use of these results obtained and build the new wiki using this content rather than starting from scratch.

While the CWW look and front-end still worked, the server and back-end had a number of IT issues that needed to be solved towards the end of 2019, before migrating into a new one. A complete copy of the CWW had to be downloaded locally and the files were reorganised. The virus had extended to multiple internal files, so it was extracted and cleaned. Once the virus was removed and the installation was properly organised, the wiki was uploaded back into the ComplexWorld wiki URL. This was successfully assessed since the installation was new and clean and no content was lost in the process.

[2] MediaWiki is the most used software for wiki platforms. It powers tens of thousands of websites, such as the well-known Wikipedia. It's powerful, multilingual, free and open, extensible, customisable, reliable, and free of charge. More information on MediaWiki can be found in Deliverable 3.7 [25].

5.1.2 Task 2: Migration to an Engage-dedicated server

The CWW was hosted on a server owned by Innaxis together with other internal resources. Once the content of the wiki was safely hosted and ready to be moved, an independent server space was chosen for the Engage wiki (in late 2019), with the idea of simplifying its legacy after the end of the Engage KTN (see Deliverable 3.7, 'Update on the Engage repository and knowledge hub functionality' [25]). The hosting provider chosen was SiteGround. The reasons for this were the reasonable cost, the good customer service provided, and the overall quality of the product, based on previous experience from Innaxis, who own various websites under this provider. At this moment the wiki content was not very heavy or complicated to process, so this server was sufficient. The purchase was made for one year, until November 2020. After a positive experience with the hosting provider, it was decided to leave the wiki hosted in this platform and the host was extended for one more year, with annual extensions going forward.

5.1.3 Task 3: Recovery from attack

The wiki was attacked by many viruses in February 2020, creating thousands of new pages with inappropriate content. The reason for this was that the user registration was left open, an issue that proved to be dangerous. The wiki was down for a few days since the hosting provider noticed the attack and took it down so that the issue could be fixed. The user registration was blocked, the users and pages were deleted, and the wiki was subsequently restored, in March 2020.

5.1.4 Task 4: Upload of European university programmes

Live link to wiki: https://wikiengagektn.com/EngageWiki:Programmes

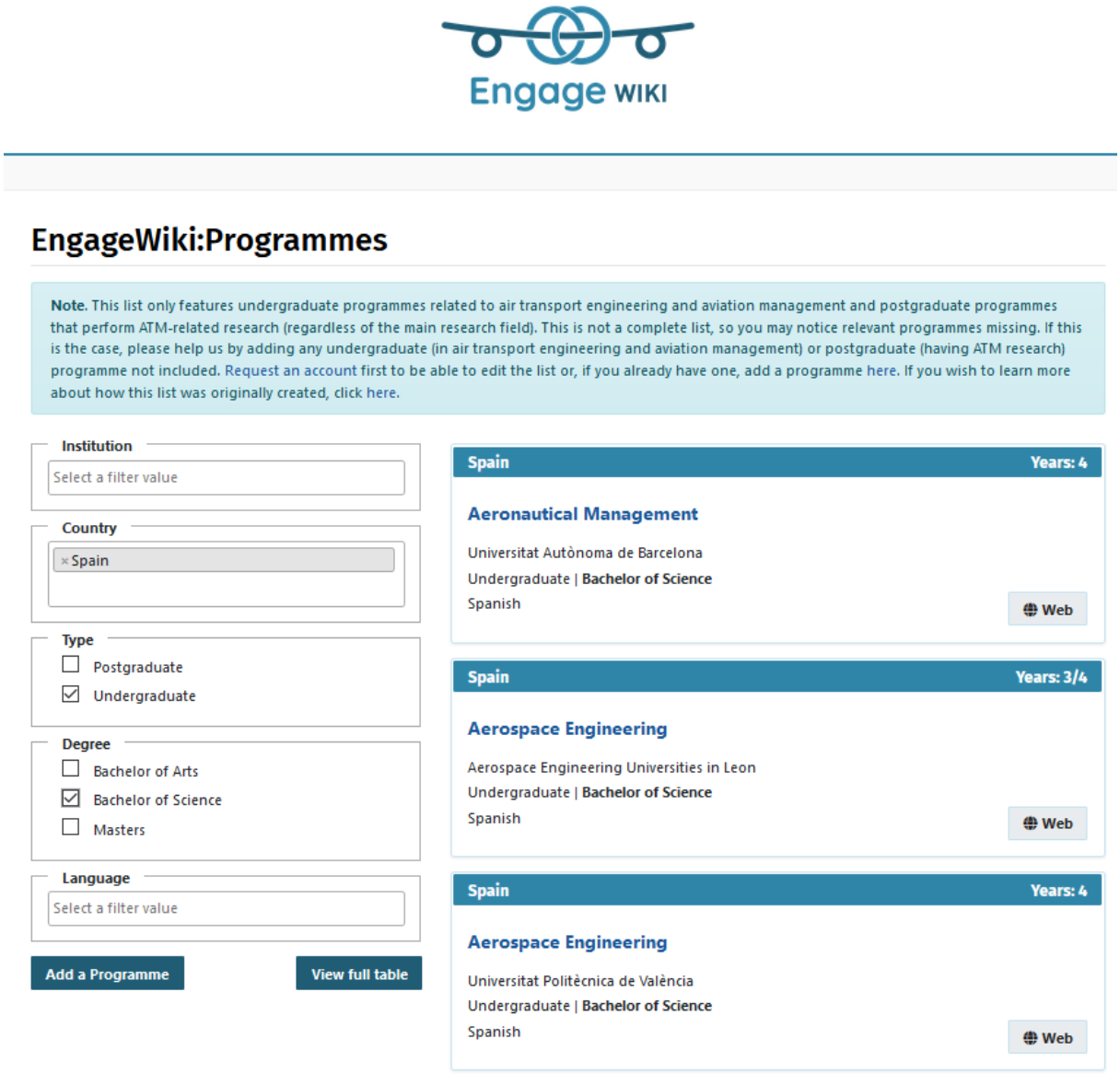

Figure 18: European university programmes page in the wiki

Innaxis led the uploading of a list of undergraduate (UG) and postgraduate (PG) programmes offered in Europe to the wiki. A first list of UG programmes was created by the University of Belgrade while the list of PG programmes was created by University of Trieste (UNITS). These lists are not complete and some clarification of the programmes coverage was added. The list of UG programmes includes only air transport engineering and aviation management related programmes. The PG list includes programmes that perform ATM related research regardless of the main research field. After an internal consultation, it was decided that the ideal would be to have a single list that grouped UG and PG programmes. The information on the programmes was collected from public sources (mainly the information published on the institutions' websites) and through surveys undertaken by Engage. This caused a disparity between the information available between programmes, however. While UG and PG share some fields, many other fields were specific only to one of them and in some cases mostly empty. Therefore, in order to merge the two lists, it was necessary to simplify them so that they both contained the same fields. The result of this merge was a unique list for UG and PG programmes with nine unique fields, detailed in Table 14.

Table 14: Unique fields for UG and PG programmes

Field name	Short description
Programme	Programme name
Institution	Teaching institution
Faculty	Faculty responsible for the programme (not in all cases known)
Country	Country of the institution
Type	Postgraduate or Undergraduate
Degree	(PhD) Master's, Bachelor of Science or Bachelor of Arts
Duration	In years
Language	Language, or languages, in which the programme is taught (English, Spanish, German...)
Link	Web URL of programme or institution

The final list contained a total of 42 UG programmes and 66 PG programmes (although 11 of these were PhD programmes, it was decided not to include these in the wiki due to their partial coverage and difficulties in locating and defining them). The final list was reviewed and updated, where needed, by the University of Belgrade and UNITS. This list was then uploaded to the wiki in two different page formats. One wiki page was created where the data was presented in a more traditional spreadsheet way with a simple filter/search bar. The second wiki page had a more advanced number of filtering options and the data were presented in a *pageblock* type format where all the data for each programme is encapsulated in an individual text block. Once both pages were created, internal feedback from the Engage partners was collected. The feedback was then distilled into specific changes and the pages were modified accordingly. After implementing these changes, feedback was again gathered. This iterative process lasted a couple of weeks until two stable versions of wiki pages were obtained.

The final decision was to make the *pageblock* type format page the main UG/PG programmes accessible in the wiki. On this page a disclaimer note (see top of Figure 18) was added notifying the users of the limitations of the scope of UG and PG programmes presented in the list. A variety of filters are available (institution, country, type, degree and language) to the user on the page. Lastly, two buttons were added to this page. One of them redirects the users to the page where the data is presented in full in a traditional spreadsheet format and a new option was added that enabled users to download the whole list in Excel format. The second button redirects users to a page where authorised users (account holders) can make modifications to programme information or add new UG/PG programmes.

The latest page layout in the wiki is shown in Figure 18. The next steps for this task are presented in Section 7.

5.1.5 Previous project hosting

It is currently intended not to maintain project-by-project pages in the wiki, where such project texts are already maintained elsewhere. Where this information is already maintained on the SJU, and/or CORDIS and/or projects' dedicated websites, the added value of creating a further source of the same information is minimal, especially in cases where there is a significant risk that these could become desynchronised, or regarding confusion as to which one is to be cited as definitive (e.g. in a SESAR brochure). Furthermore, voluntary participation from live and very recent projects regarding content updates runs the risk of patchy inclusion in the wiki.

Associated activities, where Engage is adding value (i.e. creating content and/or functionality that is not available anywhere else), is seen as more valuable, for example:

1. archiving WP-E summaries and materials[3] that are typically not available anywhere else;
2. the creation of the research repository (e.g. of deliverables; functionality under development: see Section 2.1);
3. other activities described through this report, *such as* the European universities programme listing (Section 5.1.4), the interactive research map (Section 3) and the ATM concepts roadmap (Section 4).

Figure 19 shows the corresponding page for (1) in the wiki.

[3] For ease of reference re. the original intention, the Call for the KTN referred to: " ... expected to host the open source framework models, tools and metrics developed by SESAR 1 WP-E projects and in addition contain information (e.g. scientific articles, PhD theses, conference presentations, etc.)"

Live link to wiki: https://wikiengagektn.com/Research_programmes_and_projects_(WP-E_or_others)

Figure 19: (Previous) research programmes and projects page in the wiki

5.2 Wiki launch at the 2020 SIDs

A first version of the wiki was available in May 2020. This version was not made public, but was available online such that Engage consortium members and SJU members were able to visit and work on it while having it live. After some feedback received from the SJU, changes to the visual aspects of the wiki were made. The final version, and currently public wiki, was built from May to December 2020. The wiki content was finalised in November 2020. User registration was filtered with a Captcha confirmation to avoid bugs and registration was tested until the end of November 2020. During early December 2020, the final wiki domain was configured in order to make it public at the SIDs 2020 closing session, on 10 December 2020, as introduced in Section 1. User registration was opened (see Figure 3) and the wiki has been available since this date.

6 Initial lessons learned for SESAR 3

As previewed and summarised in Section 7, multiple activities relating to the wiki will be updated in 2021, and reported in Deliverable 3.9. Hence this section identifies *initial* lessons learned for SESAR 3, which will be subsequently expanded upon, e.g. after re-running the bottom-up cluster analysis of Section 3.2, using the fuller (e.g. SESAR 2020) data described in Section 7.1.

Nevertheless, some initial conclusions may be drawn:

- For the purposes of the interactive research map (intended as a primary search tool of the wiki) and for the ATM concepts mapping, it would be helpful in SESAR 3 to standardise on the formatting of deliverables with regard to:
    - maintaining standard meta data (including common key words) and tagging each deliverable with them (see Section 2.1.1);
    - restricting the appearance of personal names (e.g. authors) and references, to given deliverable fields, such that they can be flagged / removed prior to analysis and uploading to a repository;

      (author removal will, however, prevent researchers from using the tool to search for and contact other researchers, unless each project had some form of generic contact address, e.g. “EngageKTN-coordinator@westminster.ac.uk”);
    - further standardising and enforcing the reporting (content) area of deliverable templates (see Section 3.2).

      NB1. These changes would not resolve issues regarding SIDs papers (unless similar approaches were adopted) or non-SESAR documents (where such control could not be exerted).

      NB2. The desirability of the above depends on the utility of the functionality of the tool in the wiki – if researchers are not using it, the effort is not justified.
- With regard to the control of user registration after the end of the Engage KTN:
    - user registrations should be through a simple form with Captcha confirmation to reduce viral attack;
    - *annual* hosting licences maintain good flexibility, allowing the licence to be renewed with a new name.

The value of the processes described in Section 7, for identifying new research focus areas for SESAR 3, will need to be reviewed by the Engage consortium and SJU colleagues, to identify those potentially bringing most benefit to the new programme, along with any corresponding recommendations (to be reported in Deliverable 3.9).

7 Key next steps

The wiki activities span many tasks and workpackages in the Engage network. The text below summarises *key* next steps, grouped by certain themes.

7.1 Data sources

Detailed consideration to missing data was given throughout Section 2. In summary, the following data updates are required for analyses by the research map and ATM concepts roadmap tasks in time for the next version of the wiki:

1. subject to the availability of the required SESAR 1 metadata, missing content will be added to help describe projects and deliverables;
2. the remaining SESAR 2020 deliverables will be sourced, and metadata prepared;
3. complete metadata are required for fields being considered as new filters for the research map (e.g. project budget, TRL);
4. papers presented at the tenth SIDs (December 2020) will be sourced, and metadata prepared.

Regarding the research repository, this cannot be fully implemented until permission is granted to allow SESAR 1 deliverables to be republished in the wiki. SESAR 2020 deliverables will be initially reachable via links to CORDIS, and the way forward here remains to be discussed with SJU colleagues, in particular with regard to the ideal scenario of having all deliverables in one place in the Engage repository, rather than as links (which often become redundant).

7.2 Extending the ATM concepts roadmap

As introduced in Section 4.5, future research focus areas, challenges for 2040 and beyond, will be updated and further populated through continued research in 2021, drawing on outputs from Engage, more widely in SESAR, and beyond ATM, through wiki user inputs and wider programme consultation, including the assessment of interdisciplinary concepts. The key next steps in this process are:

1. re-running the bottom-up cluster analysis of Section 3.2, using the data described in Section 7.1, to ascertain if certain clusters[4] are no longer being significantly pursued through research, and to consider whether this represents a significant omission from the research landscape;

2. assessing topics in ER4 that resulted in no associated projects, and to assess their potential role in the future funding landscape and whether this is likely to remain a significant omission;

3. considering key areas of future research identified through the Engage catalyst fund projects and how these may play a part in the formulation of the future research focus areas;

[4] As noted, (some of) the current 14 clusters may also be re-defined as a result of this process.

4. exploring the latest data on the status of other research activities such as, but not limited to, the EIC (European Innovation Council) Pathfinder[5] initiative, the Clean Aviation partnership (see Section 4.4), and the wider Horizon Europe programme itself, and to judge whether these identify future research focus areas (challenges) that require specific attention for ATM;

5. engaging the wiki community, through active promotion in the dedicated concepts roadmap discussion forum[6], to comment on the clustering and to *propose* and comment on future research focus areas, including omissions identified in (1) through (4).

Identifying these new research focus areas builds an important bridge from the current programme activities and on into SESAR 3. Whilst an update on the ATM concepts roadmap will be presented in Deliverable 3.9 (see Section 7.4), a mechanism for reviewing the above activities, (1) through (5), between the Engage consortium and SJU colleagues, should be established in 2021 (to review a concise, informal interim paper).

7.3 Other wiki-related activities

In addition to the next steps identified above, other activities will be continued / initiated in 2021 in furtherance of the wiki development:

1. continuing the assessment of feedback from users in the wiki regarding improving the content and functionality; this could include TRL filtering (see Section 2.1.5) in the ATM concepts roadmap[7];

2. continuing to assess the viability of some of the changes proposed through the use case survey summarised in Section 2.1.4;

3. continuing the discussion between the Engage consortium and SJU colleagues on the wiki's legacy beyond 2021 (to be reported in Deliverable 3.9);

4. initiating a survey of universities (planned for Q1 2021) to promote and support the development of the European university programmes listing, *inter alia* (to be reported in Deliverable 3.9).

[5] https://ec.europa.eu/research/eic/index.cfm?pg=pathfinder; the new home for the previously known Future and Emerging Technologies (FET) Programme. This offers grants to "promote collaborative, inter-disciplinary research and innovation on science-inspired and radically new future technologies." It will be active until the end of Horizon 2020, as a transition to Horizon Europe.

[6] https://wikiengagektn.com/EngageWiki_talk:ATM_concepts_roadmap

[7] It is also intended to extend the functionalities of the current ATM concepts roadmap by increasing the connection with the interactive research map. The similarity algorithm obtains a semantic similarity index for each of the projects with respect to each of the nine SRIA flagship activities. We are looking to present this information also in an interactive and visual way so users can not only identify to which of SRIA flagship activities a project is more related but to also see how it is related to the other activities.

7.4 Update to this report in Deliverable 3.9

Deliverable 3.9 will essentially function as an update to this report (D3.8), including on the activities identified in Sections 7.1 through 7.3, above, in addition to content originally planned for D4.20 (Impacts of the ER programme on the longer-term evolution of the European ATM system).

8 References

[1] SESAR Joint Undertaking, 2015. Annual Activity Report 2014, March 2015.

[2] SESAR Joint Undertaking, 2020. Single Programming Document 2020-2022, Edition 02.00, April 2020.

[3] SESAR Joint Undertaking, European ATM Master Plan, Digitalising Europe's Aviation Infrastructure, Executive view, 2019. url: https://www.sesarju.eu/sites/default/files/documents/reports/European%20ATM%20Master%20Plan%202020%20Exec%20View.pdf

[4] SESAR Joint Undertaking, Strategic Research and Innovation Agenda Digital European Sky, September 2020, url: https://www.sesarju.eu/sites/default/files/documents/reports/SRIA%20Final.pdf

[5] Clean Aviation, Strategic research and innovation agenda - The proposed European Partnership on Clean Aviation, July 2020, url: https://clean-aviation.eu/files/Clean_Aviation_SRIA_16072020.pdf

[6] European Rail Research Advisory Council, Rail Strategic Research and Innovation Agenda, 12 October 2020, url: https://shift2rail.org/wp-content/uploads/2020/10/RAIL-Strategic-Research-and-Innovation-Agenda_Draft-converted.pdf

[7] International Air Transport Association, Aircraft Technology Roadmap to 2050, 2020, url: https://www.iata.org/contentassets/8d19e716636a47c184e7221c77563c93/technology20roadmap20to20205020no20foreword.pdf

[8] SESAR Joint Undertaking, Automation in air traffic management - Long-term vision and initial research roadmap, 2020, url: https://www.sesarju.eu/sites/default/files/documents/reports/automation%20atm%20vision%20roadmap.pdf

[9] European Aviation Artificial Intelligence High Level Group, Fly AI Report - Demystifying and Accelerating AI in Aviation/ATM, 5 March 2020, url: https://www.eurocontrol.int/sites/default/files/2020-03/eurocontrol-fly-ai-report-032020.pdf

[10] SESAR Joint Undertaking, U-space, Supporting safe and secure drone operations in Europe, A preliminary summary of SESAR U-space research and innovation results (2017–2019), 2020, url: https://www.sesarju.eu/sites/default/files/documents/u-space/U-space%20Drone%20Operations%20Europe.pdf

[11] SESAR Joint Undertaking, European Drones Outlook Study - Unlocking the value for Europe, November 2016, url: https://www.sesarju.eu/sites/default/files/documents/reports/European_Drones_Outlook_Study_2016.pdf

[12] NextGen, Unmanned Aircraft System (UAS) Traffic Management (UTM), Concept of Operations v2.0, March 2020, url: https://www.faa.gov/uas/research_development/traffic_management/media/UTM_ConOps_v2.pdf

[13] European Commission, Horizon Europe - the proposal for work programme, strategic planning, implementation, news, accessed 22 December 2020, url: https://ec.europa.eu/info/horizon-europe_en

[14] United Nations, Transforming Our World: The 2030 Agenda For Sustainable Development, accessed 22 December 2020, url: https://sustainabledevelopment.un.org/content/documents/21252030 Agenda for Sustainable Development web.pdf

[15] European Space Agency, Monitoring climate from space, website, accessed 22 December 2020, url: https://www.esa.int/Applications/Observing_the_Earth/Space_for_our_climate/Monitoring_climate_change_from_space2

[16] European Space Agency, Monitoring climate, website, accessed 22 December 2020, url: https://www.esa.int/Applications/Observing_the_Earth/Space_for_our_climate/Monitoring_climate_change_from_space2

[17] IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp., url: https://www.ipcc.ch/site/assets/uploads/2018/02/SYR_AR5_FINAL_full.pdf

[18] European Commission, Cybersecurity policy, website, accessed 22 December 2020, url: https://ec.europa.eu/digital-single-market/en/cyber-security

[19] European Union Aviation Safety Agency, European Centre for Cybersecurity in Aviation (ECCSA), website, accessed 22 December 2020, url: https://www.easa.europa.eu/eccsa

[20] SESAR Joint Undertaking, A proposal for the future architecture of the European airspace, 2019, url: https://www.sesarju.eu/sites/default/files/documents/reports/Future%20Airspace%20Architecture%20Proposal.pdf

[21] European Commission, Flightpath 2050, Europe's Vision for Aviation, Report of the High Level Group on Aviation Research, 2011, url: https://ec.europa.eu/transport/sites/transport/files/modes/air/doc/flightpath2050.pdf

[22] European Commission, Horizon 2020 Work Programme for Research & Innovation 2018-2020, Mobility for Growth calls, Presentation at Info Day, accessed 22 December 2020, url: https://ec.europa.eu/inea/sites/inea/files/cefpub/7_h2020transport_info_day_mobility_for_growth_global_leadership.pdf

[23] MaaS Alliance, website, accessed 22 December 2020, url: https://maas-alliance.eu/

[24] Engage project, 2019. D3.5 Opportunities for innovative ATM research (interim report), Edition 01.02.00, July 2019.

[25] Engage project, 2020. D3.7 Update on the Engage repository and knowledge hub functionality (initial), Edition 01.02.00, March 2020.

[26] European Commission, CORDIS website, accessed July-October 2020, url: https://cordis.europa.eu/

9 Acronyms

AAS	Airspace Architecture Study
AI	Artificial intelligence
ATM	Air traffic management
ATSU	Air traffic service unit
BAFO	Best and Final Offer
CNS	Communications, navigation and surveillance
CORDIS	Community Research and Development Information Service
CWP	Controller working position
CWW	ComplexWorld wiki
D2D	Door-to-door
EATMA	European ATM architecture
EC	European Commission
EIC	European Innovation Council
ER	Exploratory research
ESSIP	European Single Sky ImPlementation (plan)
H2020	Horizon 2020 research programme
HMI	Human-machine interface
IR	Industrial research
KTN	Knowledge transfer network
LSD	Large scale demonstration
MFF	Multi-annual financial framework
ML	Machine learning
NLP	Natural language processing
NLTK	Natural language toolkit (Python)
PG	Postgraduate
R&I	Research and innovation

RPAS	Remotely piloted aircraft system
SESAR	Single European Sky ATM research
SIDs	SESAR Innovation Days
SJU	SESAR Joint Undertaking
SRIA	Strategic research and innovation agenda
TF-IDF	Term frequency - inverse document frequency
TRL	Technology readiness level
UAM	Urban air mobility
UG	Undergraduate
URL	Uniform Resource Locator
UTM	Unmanned aircraft system traffic management
VLD	Very large scale demonstration
WBS	Work breakdown structure
WP-E	SESAR Workpackage E (long-term and innovative research)

-END OF DOCUMENT-