ICT-31-2014: Human-centric Digital Age
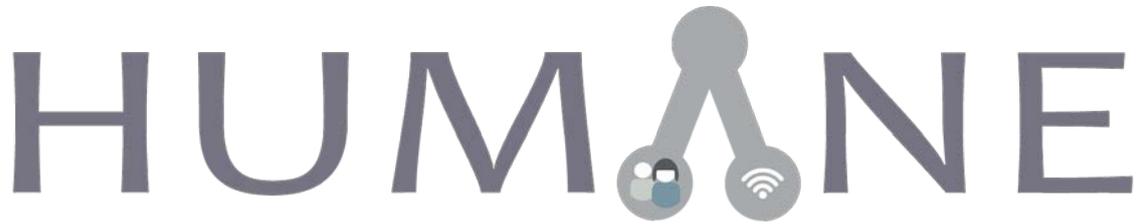**Project number: 645043**

A typology, method and roadmap for HUman-MAchine NEtworks

Deliverable D5.7
Open research data pilot

Editor(s)	Asbjørn Følstad
Lead Partner	SINTEF
Version	V03
Date	30/05/2017
EC Distribution	Public

Project Number	H2020 – 645043
Project Title	HUMANE

Title of Deliverable	Open research data pilot
Date of delivery to the EC	30/05/2017

Editor(s)	Asbjørn Følstad (SINTEF)
Contributors	Vegard Engen (IT Innovation) Brian Pickering (IT Innovation) Eva Jaho (ATC) Taha Yasseri (UOXF)
Reviewer	All

Abstract	This deliverable presents how HUMANE research data has been published to meet the Open Research Data requirements. This includes a description of the publication status for all datasets that has been collected in the project, and a listing of publications conducted in the project and how the publications have been made openly available.
Key-words	Open research data, open access publication

Versioning and contribution history

Version	Date issued	Description	Contributors
V01	11/05/2017	First draft	SINTEF
V02	24/05/2017	Second draft	SINTEF, ATC, IT Innovation, UOXF
V03	30/05/2017	Submitted version	SINTEF, ATC, IT Innovation, UOXF

Definitions and abbreviations

Abbreviation	Definition
APC	Article Processing Charges
CSV	Comma Separated Values
DS	Dataset
DMP	Data Management Plan
Gold open access	A journal-article is immediately provided in open access mode as published. Some journals require a fee (see APC) as a one-off payment by authors.
Green open access	Self-archiving: the published article or the final peer-reviewed manuscripts is archived by the author in an online repository. Most journals operate with an embargo-period, typically 12 months, before the article can be self-archived.
Open access	The practice of providing online access to scientific information that is free of charge to the end-user and that is re-usable.
Open data	Research data that is made openly accessible for any third party to be mined, exploited, reproduced and disseminated free of charge for the user.

Executive summary

HUMANE, as a research and innovation action under ICT 31, participates in the Open Research Data pilot. Hence, the project is required to make access to research data and scientific publications generated in the project as open as possible. In this report, we provide an overview of the guiding principles for open research data and open access publication as applied in HUMANE. We also provide a detailed overview of the gathering and processing of data in HUMANE, as well as an overview of open access publication of scientific results from the project.

Open research data: The guiding principles for open research data in HUMANE are grounded in the principles as laid out by the European Commission. Specifically, we aim to be as open as possible and as closed as necessary.

In total 14 datasets have been gathered and processed during the HUMANE project. Of these the majority (nine) are openly published. However, for five of the datasets considerations regarding confidentiality and challenges pertaining to informed consent made it necessary to opt out of such open publication of data. For two of these, we also share our experiences concerning unforeseen issues on informed consent of open publication of qualitative data.

For all 14 datasets, we in this report provide details on origin, data nature and scale, to whom the data could be useful and scientific publication made or planned on the basis of the data, as well as archiving and preservation. For the nine datasets which are openly published, we provide details on metadata and data sharing. For the remaining five datasets, we provide the rationale for opting out of open data publishing.

Open access to publications: The guiding principles for open access to research publications in HUMANE also follow the principles as laid out by the European Commission. All publications resulting from the project are or will be made available through green or gold open access. Furthermore, submitted or pre-print author versions of the papers are made available through the preprint service arxiv.org.

An overview of all scientific publications generated in HUMANE is provided, along with their open access publication status.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1   Introduction

All projects under ICT 31 are required to participate in the Open Research Data pilot. This implies requirements for open access to research data and open access to scientific publications. Open access is defined by the EC as "the practice of providing online access to scientific information that is free of charge to the end-user and that is re-usable" (European Commission 2017, p. 3).

This report describes the status of publication of datasets collected and processed by HUMANE ,as well as the status of open access to papers based on the research in HUMANE. The status of publication of datasets includes a report on how the data has been shared, and considerations regarding participant anonymity, informed consent, and confidentiality. The status of open access to papers, includes an overview of these

For publishing the HUMANE datasets we have, as laid out in the data management plan (DMP; HUMANE deliverable 5.3) used Zenodo as the project data repository. Project data made available through Zenodo will also be visible at the HUMANE project-page at OpenAIRE. For one of the datasets, however, we have used Figshare for data sharing instead of Zenodo, in consequence of this dataset being published as part of the requirements for publishing a study drawing on these data in PLOS ONE. Also, due to considerations regarding informed consent and confidentiality data from some of the studies have been opted out for open publication.

With regard to open access to scientific publications, papers publishing results of the HUMANE work have been made openly available. Such open availability has been achieved either through publication in open access journals (gold open access), or by making publications behind pay-walls available as pre-prints or final peer-reviewed manuscripts in open online repositories (green open access).

This deliverable is structured as follows. In section 2, we will describe the guiding principles for open research data and open access publication in HUMANE. In section 3 we present the datasets. In section 4 we describe how HUMANE complies with the Horizon 2020 mandate on open access to publications.

2   Guiding principles

2.1   Guiding principles for open research data

The legal requirements for open research data in ICT topic 31-projects are contained in the article 29.3 in the Grant Agreement, stating that:

Regarding the digital research data generated in the action ('data'), the beneficiaries must:

a) *Deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate – free of charge for any users – the following:*

I.   *the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible.*

II.   *other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan':*

b) *provide information – via the repository – about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and – where possible – provide the tools and instruments themselves).*

As can be interpreted from article 29.3 in the Grant Agreement, the objectives of open access to data primarily concern two aspects: to have raw-data available for post-validation of research results; and to permit re-use in future research projects. Relatedly, as emphasized by the EC (2017), open research data can help to accelerate innovation; foster collaboration and avoid duplication of efforts; build on previous research results; and increase the transparency of the scientific process.

However, some considerations may render it necessary and legitimate to fully or partially opt out from open publication of research data at any stage of the research process - if:

- *participation is incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited*
- *participation is incompatible with the need for confidentiality in connection with security issues*
- *participation is incompatible with rules on protecting personal data*
- *participation would mean that the project's main aim might not be achieved*
- *the project will not generate / collect any research data or*
- *there are other legitimate reasons (you can enter these in a free-text box at the proposal stage)*

(EC, 2017, p. 8)

The approach to publication of research data is described by the commission as follows: "as open as possible, as closed as necessary". (EC, 2017, p. 8)

For the datasets in HUMANE, the following considerations are of particular relevance to assess whether it is necessary to opt out from open publication of research data for some of the datasets:

Anonymity: As laid out in the data management plan, open access to research data should not represent a risk for compromising the privacy of the informants who have contributed the data. This is in line with the third bullet-point in the EC guidelines for when opting out is legitimate, cited above.

Confidentiality issues: The collaboration with industry partners in research and innovation actions may require treating research data and findings as confidential. Hence, in cases where it is not feasible to avoid that the collaborating case company is identifiable in the research data,

confidentiality issues may make it relevant to not openly publish data and findings. This is in line with the EC guidelines for when opting out is legitimate, cited above.

In addition to the above considerations, we also became aware of an unforeseen potential ethical concern during the project period. This was a concern regarding **informed consent** of open publication or research data. For all the HUMANE cases, informed consent was gathered prior to any data collection activities involving human participants. Such up front gathering of informed consent is standard practice. However, qualitative data such as the data from semi-structured interviews may contain thoughts and reflections of a character that is seen as sensitive or personal by the interviewee, which is not necessarily foreseen by the interviewee at the time the informed consent was granted. Hence, open publication of qualitative data such as anonymized interview transcripts, without informed consent gathered also after data collection, may possibly be ethically questionable. In the case of semi-structured interviews, open publication of anonymized interview transcripts may require a more extensive process for gathering informed consent, e.g. by gathering informed consent also after the study, possibly also by making the anonymized interview transcript available to the interviewee for review. We find this possible need for a more extensive process for gathering informed consent a relevant lesson learnt from being part of the Open Research Data pilot.

2.2   Guiding principles for open access publication

In addition to open access to research data, HUMANE should comply with the requirements for open access to scientific publications. Following this requirement, the scientific publications from HUMANE should be available as open access. Open access to publications can be ensured either by publishing in Gold open access journals or Green open access journals.

Gold open access means the article is available as open access by the scientific publisher. Some journals, such as PLOS ONE, require an author processing fee for publishing open access.

Green open access means that the article is submitted to, or published in, journals which accept self-archiving on an online repository of pre-prints or final peer-reviewed versions of manuscripts. Some journals, such as within the social sciences and humanities domains require authors to delay self-archiving of accepted versions of manuscripts to 12 months after the article first being published. However, most journals accept immediate self-archiving of the manuscript version that was submitted to the journal.

Figure 1 illustrates the main points for how open access to research data and publications has been ensured in the project. This figure deviates from the figure presented in the data management plan, in that (a) only datasets are published in Zenodo/OpenAIRE, (b) multiple considerations may curb open publication of research data (not just lack of anonymization), and (c) open access to manuscripts are achieved in self-archiving repositories such as arxiv.org (not in Zenodo) or the repositories of the gold open access journals.
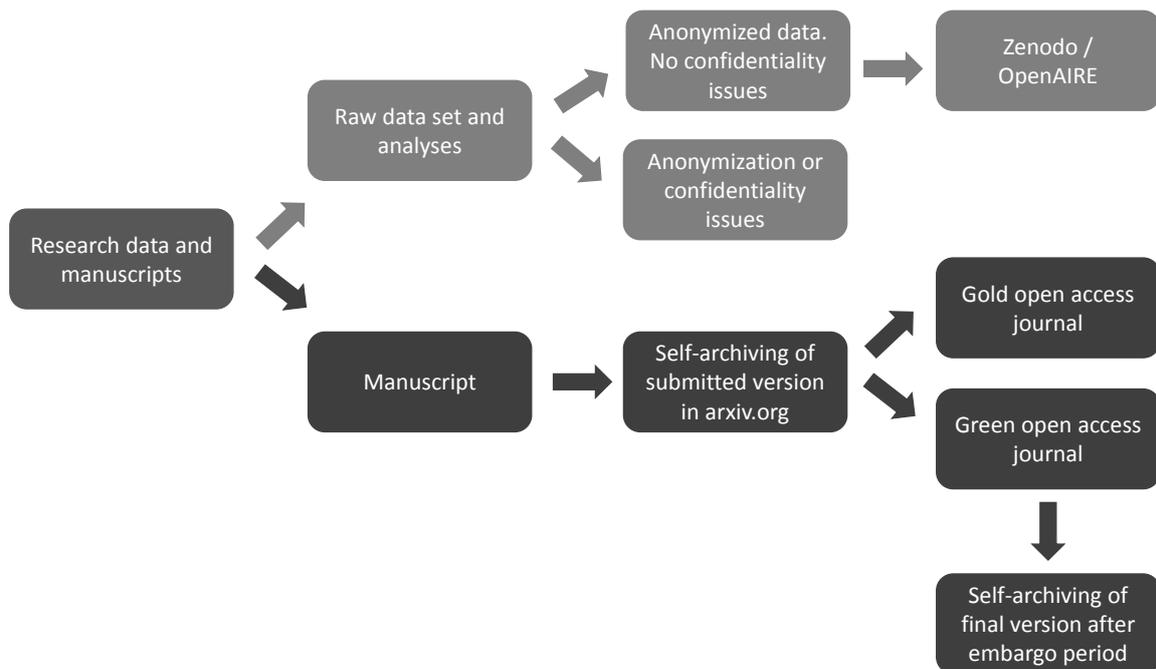
Figure 1: HUMANE open access to data and publications.

3   Datasets gathered and processed in HUMANE

In this chapter we describe the different datasets that have been gathered and processed by the HUMANE-partners. Most of these datasets have been published openly. However, some are for reasons of informed consent or confidentiality not published.

The published datasets are described in accordance with the template provided by the EC for open research data projects in Horizon 2020 (European Commission 2013b). The template is provided in Section 7: Appendix. For each dataset, we: (i) provide an identifier for the dataset to be produced; (ii) provide the dataset description; (iii) refer to standards and metadata; (iv) describe how data will be shared; and (v) describe the procedures for archiving and long-term preservation of the data.)

The published datasets follow many of the same procedures, e.g. with regard to using Zenodo as an open data repository. This means the same wording is often repeated in the different datasets. As each dataset description should give a comprehensive overview of the gathering, processing and open access archiving of data, we assessed it as necessary to repeat the procedures in the different dataset descriptions.

The datasets opted out from open publication due to confidentiality issues are also described according to the same template, but here not all fields of the template are relevant.

The name for each dataset includes a prefix "DS" for dataset, followed by a case-study (CS) or research activity (RA) identification number, the partner responsible for collecting and processing the data, as well as a short title. The H2020 DMP template requires that information about dataset

metadata is provided. We have primarily based the outlining of how and what metadata will be created on the guidelines provided by the European University Institute (2015).

Table 1 gives an overview of the datasets collected and processes in HUMANE, along with a note on their publication status. The descriptions of each dataset are provided in the following sections.

Table 1: Overview of datasets

No.	Identifier/name	Brief description	Publication status
1	DS.C1.SINTEF. Open Innovation dataset	This dataset consists of (a) the anonymized transcripts of interviews with 7 case company representatives in open innovation, structured according to the HUMANE typology dimensions, and (b) the anonymous data from a questionnaire study among users of the open innovation platform in the case company.	Opt out due to confidentiality issues
2	DS.C2.SINTEF. Redistribution markets dataset	This dataset consists of (a) the anonymized transcripts of interviews with 12 users of redistribution market services, and (b) the anonymous analysis of redistribution market ads.	Opt out due to confidentiality issues
3	DS.C3.IT Innovation. eVACUATE dataset #1	This dataset consists of anonymized transcripts from a focus group with IT/system architects from the eVACUATE project and a summary from the focus group analysis.	Published on Zenodo
4	DS.C3.IT Innovation. eVACUATE dataset #2	The dataset consists of (a) the surveys and anonymized survey responses on using the HUMANE typology and on the HUMANE design patterns, (b) anonymized transcripts from two semi-structured focus groups evaluating the HUMANE method and, and (c) summaries from the analysis of the above data sources of data.	Published on Zenodo
5	DS.C4.ATC. REVEAL dataset #1	The dataset consists of content analysis of (1) qualitative data on HUMANE methodology & dimensions' evaluation in the context of REVEAL, (2) data gathered from the REVEAL human – machine network, regarding the dependencies between the network's elements.	Published on Zenodo
6	DS.C4.ATC. REVEAL dataset #2	The dataset consists of anonymous responses from two online user surveys on HMN characteristics and efficacy of design solutions.	Published on Zenodo

7	DS.C5.UOXF. Wikipedia dataset	This dataset provides the analysed large-scale raw data extracted from selected channels for Wikipedia data, addressing transactions and collaboration between active and contributing Wikipedia-users.	Published on Figshare
8	DS.C6.UOXF. Zooniverse dataset	This dataset provides anonymous logs of contributors' classifications at the citizen science portal Zooniverse.	Published on Zenodo
9	DS.C7.SINTEF. Emergency management system dataset	This dataset consists of the anonymized transcripts of interviews with 6 user and 6 service provider representatives for an emergency management system, structured according to the HUMANE typology dimensions.	Opt out due to confidentiality issues
10	DS.C8.SINTEF. Decision support case dataset	This dataset consists of the anonymized transcripts of interviews with 4 participants in an innovation project addressing automation in decision support systems, structured according to the HUMANE typology dimensions.	Opt out due to confidentiality issues
11	DS.RA1.ATC. stakeholder survey dataset	This dataset consists of the anonymous raw-data from a survey conducted with relevant practitioners to gather data on needs, expectations and experiences with human-machine networks within three specific domains (eHealth, sharing economy, citizen participation).	Published on Zenodo
12	DS.RA2.ATC eHealth roadmapping interviews dataset	This dataset consists an anonymized summary of the feedback received for the eHealth roadmap from two focus groups and one interview.	Published on Zenodo
13	DS.RA3.SINTEF sharing economy roadmapping interviews dataset	This dataset consists of the anonymized transcripts of interviews with 9 service owner representatives, 4 policy maker representatives, and 6 researchers / experts on sharing economy services, exploring the future of human-machine networks in the sharing economy.	Opt out due to confidentiality issues
14	DS.RA4.ITINNO simulation and modelling data and analyses	This dataset was derived from the publicly available data from Wikipedia, for the purposes of simulation modelling in the HUMANE project. The data was derived from the Simple English language edition of Wikipedia, made available via the WikiWarMonitor project:	Published on Zenodo

3.1   DS.C1.SINTEF. Open innovation dataset

The DS.C1 dataset consists of: (1) Qualitative and anonymized interview-transcripts with employees in an enterprise, which uses an online open innovation solution for gathering suggestions and ideas for service innovation; (2) Anonymized and primarily qualitative data from a survey with customers who have contributed ideas and suggestions to the open innovation solution, and (3) Anonymized interview-transcripts from focus-groups with enterprise-employees.

Anonymous data are items of information that cannot in any way identify individuals in the data material directly through names or personal ID numbers, indirectly through background variables, or through a list of names / connection key or encryption formula or code.

As part of the anonymization process for the interview transcripts, any references to named persons, the company role of the interviewee, and the specific background of the employee were removed.

Also the company name and names of company services and projects were removed from the interview transcripts. Nevertheless, we assess it as potentially possible for readers of the interview transcripts to identify the case company; though the individual interviewee cannot be identified. This because the company has a profile not shared by many other companies, and also because the company is the only company of its kind participating in the collaborating project in which the case has taken place.

3.1.1   Dataset description

Origin of data: The data were collected by SINTEF researchers as part of the HUMANE project, and in collaboration with the Norwegian Center for Service Innovation (CSI) (HUMANE case 1).

Nature and scale of data: (1) Transcripts of interview data in the language it was conducted (Norwegian); (2) Completed questionnaire responses in Norwegian. (3) Transcripts of focus-group interviews in Norwegian.

To whom the dataset could be useful: Outside of the consortium, the data in its anonymized form might be useful for other researchers interested in the potentials and limitations of open innovation and crowdsourcing of ideas. However, the transcripts are in Norwegian which clearly delimits the usefulness of the data outside of Scandinavian countries.

Scientific publication: The results of the dataset has been included in HUMANE deliverable D3.2

Existence of similar datasets? To our knowledge qualitative datasets on the experiential aspects of online open innovation are not openly available.

3.1.2   Standards and metadata

Not relevant, as the case needs to opt out from open sharing of research data (see next subsection).

3.1.3   Data sharing – opt out from open publication due to confidentiality issue

It was outlined in the data management plan that the interview and survey data should be published openly. Such open publication was motivated in that the datasets could be adequately anonymized, which they have. However, what was not considered at the time was that open publication of data also would require these not to imply confidentiality issues.

Following the Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 (EC, 2017), a project may partially opt out of the requirement for open publication of research data when confidentiality may be compromised.

In this case, we acknowledge that the research data have a character that require confidentiality, as they concern business critical aspects of the case company provided. In particular, the datasets provide rich descriptions of experiences - positive and negative - from the employees (interviews) and customers (survey) of the case company.

It should also be noted that the case work was planned and conducted under the condition that the report from the case (HUMANE deliverable D3.2) would be a confidential deliverable. Without this confidentiality we assume that the data provided would be less rich in character and, therefore, less valuable as basis for the work presented in deliverable D3.2.

It may also be noted that the gathered interview data would have required particular ethical consideration in the case they were to be openly published. As the anonymized interview transcripts may have a personal character, open publication may possibly have required informed consent to be gathered in a different manner than what was initially envisioned in this case. As discussed in Section 2.1, such open publication might require informed consent to be gathered also after data collection, possibly on the basis of the participant reviewing the anonymized interview transcript.

3.1.4   Archiving and preservation (including storage and backup)

The data are archived in anonymized form on an inhouse secure server at SINTEF.

Audio-recordings and non-anonymized transcripts have all been deleted. So has the participant list.

3.2   DS.C2.SINTEF. Redistribution markets dataset

The DS.C2 dataset consists of (1) qualitative and anonymized interview-transcripts with adult end-users of an online redistribution service and (2) Summaries of a qualitative content analysis.

Anonymous data are items of information that cannot in any way identify individuals in the data material directly through names or personal ID numbers, indirectly through background variables, or through a list of names / connection key or encryption formula or code.

As part of the anonymization process for the interview transcripts, person names and any references to the specific background of the participants was removed.

Also the company name and names of company services and projects were removed from the interview transcripts. Nevertheless, we assess it as potentially possible for readers of the interview transcripts to identify the case company as the object of conversation in the interviews. This because the case company has a profile not shared by many other companies.

3.2.1   Dataset description

Origin of data: The data were collected by SINTEF researchers as part of the HUMANE project, and in collaboration with the Norwegian innovation project Conserve and Consume (HUMANE case 2).

Nature and scale of data: (1) Transcripts of interview data in Norwegian and (2) summaries of qualitative content analysis in Norwegian.

To whom the dataset could be useful: Outside of the consortium, the interviews data in its anonymized form might be useful for other researchers interested in the sharing economy in general and in user-experiences of redistribution markets in particular. The interview transcripts is in Norwegian, which limits the usefulness of the data outside of Scandinavian countries.

Scientific publication: The results of the dataset has been included in HUMANE deliverable D3.2.

Existence of similar datasets? To our knowledge qualitative datasets on the experiential aspects of online redistribution markets are not openly available.

3.2.2   Standards and metadata

Not relevant, as the case needs to opt out from open sharing of research data (see next subsection).

3.2.3   Data sharing – opt out from open publication due to confidentiality issue

It was outlined in the data management plan that this dataset should be published openly. Such open publication was motivated in that the dataset could be adequately anonymized, which they have. However, what was not considered at the time was that open publication of data also would require these not to imply confidentiality issues.

Following the Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 (EC, 2017), a project may partially opt out of the requirement for open publication of research data when confidentiality may be compromised.

In this case, we acknowledge that the research data have a character that require confidentiality, as they concern business critical aspects of the case company provided. In particular, the datasets provide rich descriptions of experiences - positive and negative - from the customers of the case company.

It should also be noted that the case work was planned and conducted under the condition that the report from the case (HUMANE deliverable D3.2) would be a confidential deliverable. Without such confidentiality we assume that the inclusion of case companies such as that in Case 2 would be challenging.

It may also be noted that the gathered interview data would have required particular ethical consideration in the case they were to be openly published. As the anonymized interview transcripts may have a personal character, open publication may possibly have required informed consent to be gathered in a different manner than what was initially envisioned in this case. As discussed in Section 2.1, such open publication might require informed consent to be gathered also after data collection, possibly on the basis of the participant reviewing the anonymized interview transcript.

3.2.4 Archiving and preservation (including storage and backup)

The data are archived in anonymized form on an inhouse secure server at SINTEF.

Audio-recordings and non-anonymized transcripts have all been deleted. So has the participant list.

3.3   DS.C3.IT Innovation. eVACUATE dataset #1 (Year 1)

This dataset was collected in the first (internal) eVACUATE case study, which was conducted on 14 December 2015. The purpose was to evaluate the usefulness of the HUMANE approach as perceived by relevant developers (software engineers), and additionally ask if the HUMANE typology facilitates cross-disciplinary understanding.

The dataset consists of: (1) a qualitative and anonymized transcript from a semi-structured focus group with IT/system architects from IT Innovation working on the eVACUATE project, but unaware of the aims or outcomes of HUMANE; (2) summary from the analysis of the aforementioned focus group.

Anonymous data are items of information that cannot in any way identify individuals in the data material directly through names or personal ID numbers, indirectly through background variables, or through a list of names / connection key or encryption formula or code. The dataset does not include the name of the participants, specific eVACUATE use cases or references to other information like projects or locations. The combination of background variables such as gender, age, employee role in the use case and the use case name increases the risk of identifying individuals in the data material. Therefore, at this stage, we deem that withholding the aforementioned information is sufficient to ensure the privacy of participants. Similarly we have excluded IP addresses from the survey responses, although the survey tool automatically collected this information. It was not used as part of the analysis.

3.3.1   Dataset description

Origin of data: The dataset was collected in the HUMANE project based on responses from IT architects/designers from the eVACUATE project. The data was collected and analysed by IT Innovation.

Nature and scale of data: The data is small scale, qualitative, data from 4 participants. There are two types of data: (1) anonymized focus group transcripts in English; (2) anonymized summaries of analysis in English.

To whom the dataset could be useful: Outside of the consortium, the data in its anonymized form might be useful for other researchers interested in the experience of system end-users and system designers of using the HUMANE resources for designing human-machine networks. All data is in English, and is, thus, widely accessible.

Scientific publication: The data has been used as part of research submitted for scientific publication.

Existence of similar datasets? To our knowledge, there are no similar datasets available, except for other datasets generated in the HUMANE project.

3.3.2 Standards and metadata

The dataset is characterized with the following metadata:

- Author/compiler of dataset: Brian Pickering and Vegard Engen, University of Southampton IT Innovation Centre
- Funded by: [HUMANE, H2020 – 645043]
- Format: [PDF/A]
- Content-data: evacuation, eVACUATE use cases, system design, HUMANE typology feedback and evaluation, technology-mediated collaboration, decision making for evacuation, dynamic HMN creation in crises.
- Method of data accumulation: qualitative focus groups.
- Data collection period [from] – [to]: 01.12.2015 – 31/12/2015.
- Conditions of use of data: open access, free of charge
- DOI: 10.5281/zenodo.571843
- Related publications:
  - Følstad, A., Engen, V., Haugstveit, I. M., & Pickering, J. B. (2017). Automation in Human-Machine Networks: How Increasing Machine Agency Affects Human Agency. In Proceedings of International Conference on Man-Machine Interactions – ICMMI 2017 (forthcoming)

3.3.3 Data sharing

Access procedures: The data is made accessible and available for re-use and secondary analysis by uploading the data to Zenodo.

Document format and availability: The dataset will be available as PDF/A at the HUMANE community on Zenodo: https://www.zenodo.org/communities/humane2020/ . From here the fully anonymized data are open accessible for anyone, free of charge.

This dataset is available at the following URL, with an embargo period until the end of May 2017: https://doi.org/10.5281/zenodo.571843

3.3.4 Archiving and preservation (including storage and backup)

Archiving of the anonymized dataset at Zenodo guarantees a long-term and secure preservation of the data at no additional cost for the project. Zenodo informs that *"in the highly unlikely event that Zenodo will have to close operations, we guarantee that we will migrate all content to other suitable repositories, and since all uploads have DOIs, all citations and links to Zenodo resources (such as your data) will not be affected"*.

3.4   DS.C3.IT Innovation. eVACUATE dataset #2 (Year 2)

This case study was conducted in September to October 2016 with the purpose of providing an external validation of the HUMANE typology and method. This eVACUATE case-study comprises four different engagements in order to ensure a comprehensive evaluation: a quantitative online survey on the HUMANE design patterns; a quantitative survey on the HUMANE typology used for characterising Human-Machine Networks (HMNs); and two focus groups evaluating the HUMANE method (covering the profiling process, network diagramming, implication analysis, and design pattern approach). Four self-selecting cohorts, identified on an opportunity sampling basis, took part in one or more of these exercises:

1. Project members from the eVACUATE project. Participants responded to both surveys;
2. Undergraduate students from the Faculty of Physical Science and Engineering at the University of Southampton. Participants responded to the survey on design patterns, and took part in a qualitative focus group to explore using the typology and methodology in connection with their own understanding of a human-machine network of their choice;
3. ICT professionals from the Research Software Engineering group in the UK (http://www.rse.ac.uk/). Participants responded to the same survey as cohort [2] above; and finally,
4. A group of professional software and ICT professionals at the IT Innovation Centre, University of Southampton with experience in a range of pan-European collaborative projects working on HMNs. Participants took part in a similar focus group type qualitative study as cohort [2] above, which was based on a generalised eVACUATE HMN.

The dataset consists of: (1) quantitative and anonymized survey responses on using the HUMANE typology (as per D2.2); (2) quantitative and anonymized survey responses on the HUMANE design patterns; (3) two qualitative and anonymized transcripts from the semi-structured focus groups evaluating the HUMANE method; (4) summaries from the analysis of the three former sources of data; (5) the surveys used in [1] and [2].

Anonymous data are items of information that cannot in any way identify individuals in the data material directly through names or personal ID numbers, indirectly through background variables, or through a list of names / connection key or encryption formula or code. The dataset does not include the name of the participants, specific eVACUATE use cases or references to other information like projects or locations. The combination of background variables such as gender, age, employee role in the use case and the use case name increases the risk of identifying individuals in the data material. Therefore, at this stage, we deem that withholding the aforementioned information is sufficient to ensure the privacy of participants. Similarly we have excluded IP addresses from the survey responses, although the survey tool automatically collected this information. It was not used as part of the analysis.

3.4.1   Dataset description

Origin of data: The dataset was collected in the HUMANE project based on the four engagements described above. The data was collected and analysed by IT Innovation.

Nature and scale of data: The data is small scale, qualitative, data from 4 participants. There are four types of data: (1) anonymized survey responses in English; (2) anonymized focus group transcripts in English; (3) anonymized summaries of analysis in English; (4) the surveys used in [1].

To whom the dataset could be useful: Outside of the consortium, the data in its anonymized form might be useful for other researchers interested in the experience of system end-users and system designers of using the HUMANE resources for designing human-machine networks. All data is in English, and is, thus, widely accessible.

Scientific publication: The data has been used as part of research submitted for scientific publication.

Existence of similar datasets? To our knowledge, there are no similar datasets available, except for other datasets generated in the HUMANE project.

3.4.2   Standards and metadata

The dataset is characterized with the following metadata:

- Author/compiler of dataset: Vegard Engen and Brian Pickering, University of Southampton IT Innovation Centre
- Funded by: [HUMANE, H2020 – 645043]
- Format: [PDF/A]
- Content-data: evacuation, eVACUATE use cases, system design, HUMANE typology and method feedback and evaluation, technology-mediated collaboration, decision making for evacuation, dynamic HMN creation in crises.
- Method of data accumulation: quantitative surveys, qualitative focus groups, qualitative content analysis.
- Data collection period [from] – [to]: 01.09.2016 – 31/10/2016.
- Conditions of use of data: open access, free of charge
- DOI: 10.5281/zenodo.571846
- Related publications:
  - Følstad, A., Engen, V., Haugstveit, I. M., & Pickering, J. B. (2017). Automation in Human-Machine Networks: How Increasing Machine Agency Affects Human Agency. In Proceedings of International Conference on Man-Machine Interactions – ICMMI 2017 (forthcoming)

3.4.3   Data sharing

Access procedures: The data is made accessible and available for re-use and secondary analysis by uploading the data to Zenodo.

Document format and availability: The dataset will be available as PDF/A at the HUMANE community on Zenodo: https://www.zenodo.org/communities/humane2020/ . From here the fully anonymized data are open accessible for anyone, free of charge.

This dataset is available at the following URL, with an embargo period until the end of May 2017: https://doi.org/10.5281/zenodo.571846

3.4.4   Archiving and preservation (including storage and backup)

Archiving of the anonymized dataset at Zenodo guarantees a long-term and secure preservation of the data at no additional cost for the project. Zenodo informs that *"in the highly unlikely event that Zenodo will have to close operations, we guarantee that we will migrate all content to other suitable repositories, and since all uploads have DOIs, all citations and links to Zenodo resources (such as your data) will not be affected"*.

3.5   DS.C4.ATC. REVEAL dataset #1 (Year 1)

The REVEAL project (FP7-ICT-2013-10) was studied as a case study of a Human Machine Network (HMN) that provides advanced content aggregation-filtering-curation and recommendation tools, enabling professionals (mainly journalists) and normal users to reveal hidden 'modalities' such as reputation, influence and credibility. The purpose of the study was to assess the importance of HMN characteristics (such as user motivation, privacy and trust management) and the efficacy of the design solutions that had been applied, as well as to collect recommendations for improvements.

This dataset refers to the first year of the REVEAL case study (2015). It consists of two summaries of qualitative content analysis: (1) Qualitative data on HUMANE methodology & dimensions' evaluation in the context of REVEAL, (2) Data gathered from the REVEAL human – machine network, regarding the dependencies between the network's elements.

As in the datasets of the other use cases, the data were anonymous, meaning that it cannot in any way be used in order to identify individuals in the data material directly through names or personal ID numbers, indirectly through background variables, or through a list of names / connection key or encryption formula or code.

In order to ensure confidentiality, the lists with names and reference-number to the participants are kept separate from the empirical data. These lists are not stored together with the main material, but stored in an isolated computer belonging to the institution conducting the different case studies, and accessible only for the person in charge of the case-study.

3.5.1   Dataset description

Origin of data: The dataset was gathered through two focus groups on journalists-experiences with the REVEAL platform, in order to identify the credibility of several sources in the internet. The data in this dataset was collected and analysed by ATC.

Nature and scale of data: There were two focus groups conducted in which 11 journalists participated. The dataset contains two summaries of qualitative content analysis in English.

To whom the dataset could be useful: Outside of the consortium, the data in its anonymized form might be useful for other researchers interested in the investigation of sources' credibility for journalists.

Scientific publication: The results of the data set has been included in HUMANE deliverable D3.2.

Existence of similar datasets? To our knowledge qualitative datasets on the sources' credibility for journalists are not openly available.

3.5.2   Standards and metadata

The dataset is characterized with the following metadata:

- Author/compiler of dataset: George Bravos, ATC

- Funded by: [HUMANE, H2020 – 645043]
- Format: [PDF/A]
- Content-data: trustworthiness, online sources credibility, human-machine networks
- Method of data accumulation: qualitative summaries.
- Data collection period [from] – [to]: 01.11.2015 – 15.12.2015
- Conditions of use of data: open access, free of charge.
- DOI: 10.5281/zenodo.580690 and 10.5281/zenodo.580695
- Related publications n/a

3.5.3 Data sharing

Access procedures: The datasets are uploaded to Zenodo. In addition, the datasets were contained in D3.2 (Report on first set of case-studies).

Document format and availability: The dataset will be available as PDF/A at the HUMANE community on Zenodo: https://www.zenodo.org/communities/humane2020/ . From here the fully anonymized data are open accessible for anyone, free of charge.

This dataset is available at the following URLs, with an embargo period until the end of May 2017: https://doi.org/10.5281/zenodo.580690 and https://doi.org/10.5281/zenodo.580695.

3.5.4 Archiving and preservation (including storage and backup)

Archiving of the anonymized dataset at Zenodo guarantees a long-term and secure preservation of the data at no additional cost for the project. Zenodo informs that "*in the highly unlikely event that Zenodo will have to close operations, we guarantee that we will migrate all content to other suitable repositories, and since all uploads have DOIs, all citations and links to Zenodo resources (such as your data) will not be affected.*"

3.6   DS.C4. REVEAL dataset #2 (Year 2)

The related dataset refers to the second year of the REVEAL case study (2016). It consists of data from two online user surveys. The online user surveys were conducted through LimeSurvey (https://www.limesurvey.org/) between 19 and 31 October 2016.

3.6.1   Dataset description

Origin of data: The dataset was collected in the HUMANE project, based on responses from IT architets/designers of the REVEAL platform. The data was collected and analysed by ATC.

Nature and scale of data: There are two surveys related to this dataset. The survey on the importance of HMN characteristics received 15 responses and the survey on the efficacy of the design solutions received 12 responses. The dataset from the online user surveys consist of tables of responses. The online surveys were completely anonymous and the datasets will not contain any personally identifiable information.

 To whom the dataset could be useful: Outside of the consortium, the data in its anonymized form might be useful for other researchers interested in the design of content aggregation-filtering-curation and recommendation systems. All data is in English, and is, thus, widely accessible.

Scientific publication: The results of the data set has been included in HUMANE deliverable D3.3.

Existence of similar datasets? Similar datasets were produced for other case studies conducted within the HUMANE project.

3.6.2   Standards and metadata

The dataset is characterized with the following metadata:

- Author/compiler of dataset: Eva Jaho and Nikos Sarris, ATC
- Funded by: [HUMANE, H2020 – 645043]
- Format: [PDF/A]
- Content-data: trustworthiness, online sources credibility, REVEAL use case, HUMANE typology and method feedback and evaluation
- Method of data accumulation: quantitative surveys.
- Data collection period [from] – [to]: 19.10.2016 – 31.10.2016
- Conditions of use of data: open access, free of charge.
- DOI: 10.5281/zenodo.571106 and 10.5281/zenodo.571105
- Related publications: n/a

3.6.3   Data sharing

Access procedures: The datasets are uploaded to Zenodo. In addition, the datasets were contained in D3.3 (Report on second set of case-studies).

Document format and availability: The dataset will be available as PDF/A at the HUMANE community on Zenodo: https://www.zenodo.org/communities/humane2020/ . From here the fully anonymized data are open accessible for anyone, free of charge.

This dataset is available at the following URLs, with an embargo period until the end of May 2017:

https://doi.org/10.5281/zenodo.571106 and https://doi.org/10.5281/zenodo.571105.

3.6.4   Archiving and preservation (including storage and backup)

Archiving of the anonymized dataset at Zenodo guarantees a long-term and secure preservation of the data at no additional cost for the project. Zenodo informs that "*in the highly unlikely event that Zenodo will have to close operations, we guarantee that we will migrate all content to other suitable repositories, and since all uploads have DOIs, all citations and links to Zenodo resources (such as your data) will not be affected.*"

3.7   DS.C5.UOXF. Wikipedia dataset

Wikipedia is the focus of DS.C5. An outstanding feature of Wikipedia is that every single action of its editors is tracked and recorded. This includes all edits on articles, posts on talk pages, page deletions or creations, changes in page titles, uploading multimedia files, etc. Apart from the practical advantages of this complete archiving, it is also extremely valuable from scientific point of view.

There are three main channels for collecting Wikipedia data:

- Live data: There are two convenient ways to access live data of Wikipedia. i) "Wikimedia Toolserver databases (http://toolserver.org/), which contains a replica of all Wikimedia wiki databases, and ii) "MediaWiki web service API" (https://www.mediawiki.org/wiki/API).
- Dumped data: Wikipedia also offers archived copies of its content in different formats (http://dumps.wikimedia.org), e.g., XML and HTML and different types, e.g., snapshots of full history of articles or a collection of latest version of all articles.
- Semantic Wikipedia: "Semantic Wikipedia", as a general concept would be a combination of Semantic Web and WP data to provide structured datasets through query services. There are various projects providing access to Semantic WP. Examples are "DBpedia" (http://dbpedia.org) "Semantic MediaWiki" (http://semantic-mediawiki.org), and "Wikipedia XML corpus" (http://www-connex.lip6.fr/~denoyer/wikipediaXML), and most notably, Wikidata (https://www.wikidata.org).

3.7.1   Dataset description

Nature and scale of data: Most of the data described about are either numeric or textual (action logs and article content respectively). The size of the data is at the order of few Tera Bytes. Therefore, it is essential to use live access to publicly available replica (a few of which are named above) rather than locally host the data. The analysed datasets however can be locally host and shared with other interested parties.

To whom the dataset could be useful: Outside of the consortium, other researchers with interest in analysing Wikipedia activity data can use the data packages produced in this case study.

3.7.2   Standards and metadata

The dataset is characterized with the following metadata:

- Author/compiler of dataset: Milena Tsvetkova, Ruth Garcia, Taha Yasseri, UOXF
- Funded by: [HUMANE, H2020 – 645043]
- Format: [CSV]
- Content-data: Wikipedia, collective action, editorial activity and readership
- Method of data accumulation: large-scale statistical analysis.
- Date-range coverage of dataset: 01.01.2001 – 15.12.2011
- Conditions of use of data: open access, free of charge.
- DOI: http://dx.doi.org/10.6084/m9.figshare.4597918

- Related publications
    - Tsvetkova, M., García-Gavilanes, R., Floridi, L., & Yasseri, T. (2017). Even good bots fight: The case of Wikipedia. *PloS one, 12*(2), e0171774.
    - Tsvetkova, M., García-Gavilanes, R., & Yasseri, T. (2016). Dynamics of Disagreement: Large-Scale Temporal Network Analysis Reveals Negative Interactions in Online Collaboration. *Scientific Reports*, 6

3.7.3   Data sharing

Access procedures: The data are accessible and available for re-use and secondary analysis at Figshare.com.

Document format and availability: The dataset is available as CSV at http://dx.doi.org/10.6084/m9.figshare.4597918. From here the data are open accessible for anyone, free of charge.

3.7.4   Archiving and preservation (including storage and backup)

Archiving of the dataset at Figshare guarantees a long-term and secure preservation of the data at no additional cost for the project. Figshare informs that "Not only do we provide a place for you to share, but we also keep your data (and metadata) stored securely and backed up using AWS. With many funders now making it a requirement to keep your data, figshare provides a safe environment do so."

3.8   DS.C6.UOXF. Zooniverse dataset

DS.C6 will consider Zooniverse, the citizen science portal. The dataset used in this case study consist of logs of contributors' classification. The dataset was produced in collaboration with the Zooniverse team and is not originally publicly available. The Zooniverse User Agreement describes how usage information (e.g. log-ins, page-requests, classifications made) are recorded and made available for collaborators to the Citizens Science Alliance (Oxford as one of the collaborators) for research purposes (see D3.1). The dataset is anonymized, meaning no directly or indirectly identifiable information will be disclosed in the process of sharing the data through Zenodo.

3.8.1   Dataset description

Nature and scale of data: The dataset under study consist of logs of 3.5 years of 35,000,000 contributions to 17 projects of Zooniverse by 345,000 users form 198 countries.

To whom the dataset could be useful: This is a unique dataset in the area of citizen science studies. No project has been growing at this scale and no aggregate data at this size is publicly available.

3.8.2   Data sharing

Access procedures: The anonymized dataset has been made accessible and available for re-use and secondary analysis by uploading the data to Zenodo.

Document format and availability: The dataset is available as an SQL database at http://www.zenodo.org/collection/datasets. From here the fully anonymized data are open accessible for anyone, free of charge.

3.8.3   Standards and metadata

The dataset is characterized with the following metadata:

- Author/compiler of dataset: Taha Yasseri, UOXF.
- Funded by: [HUMANE, H2020 – 645043]
- Format: [SQL]
- Content-data: Citizen science, large-scale collaboration, crowdsourcing
- Method of data accumulation: large scale statistical analysis.
- Date-range coverage of dataset: 09.11.2009 – 01.06.2013.
- Conditions of use of data: open access, free of charge.
- DOI: http://doi.org/10.5281/zenodo.583182
- Related publications

    Khairunnisa Ibrahim, Samuel Khodursky, & Taha Yasseri (2016). Spatiotemporal patterns of classifications to the Zooniverse. Under review.

3.8.4 Archiving and preservation (including storage and backup)

Archiving of the anonymized dataset at Zenodo guarantees a long-term and secure preservation of the data at no additional cost for the project. Zenodo informs that "in the highly unlikely event that Zenodo will have to close operations, we guarantee that we will migrate all content to other suitable repositories, and since all uploads have DOIs, all citations and links to Zenodo resources (such as your data) will not be affected."

3.9   DS.C7.SINTEF. Emergency management system dataset

The DS.C7 dataset consists of: Qualitative and anonymized interview-transcripts with six employees of a HUMANE case company and six users of a system provided by the case company.

Anonymous data are items of information that cannot in any way identify individuals in the data material directly through names or personal ID numbers, indirectly through background variables, or through a list of names / connection key or encryption formula or code.

As part of the anonymization process for the interview transcripts, any references to named persons, the company role of the interviewee, and the specific background of the employee were removed.

Also the enterprise name and names of related services and projects were removed from the interview transcripts. Nevertheless, we assess it as potentially possible for readers of the interview transcripts to identify the case enterprise; though the individual interviewee cannot be identified. This because the company has a profile not shared by many other enterprises.

3.9.1   Dataset description

Origin of data: The data were collected by two SINTEF researchers as part of the HUMANE project.

Nature and scale of data: Transcripts of interview data in the language it was conducted (Norwegian).

To whom the dataset could be useful: Outside of the consortium, the data in its anonymized form might be useful for other researchers interested in decision support systems. However, the transcripts are in Norwegian which clearly delimits the usefulness of the data outside of Scandinavian countries.

Scientific publication: The results of the dataset has been included in HUMANE deliverable D3.3 and in a submitted conference paper (submitted version available on https://arxiv.org/abs/1702.07480)

Existence of similar datasets? To our knowledge qualitative datasets on researchers perspectives on automation in decision support systems are not openly available.

3.9.2   Standards and metadata

Not relevant, as the case needs to opt out from open sharing of research data (see next subsection).

3.9.3   Data sharing – opt out from open publication due to confidentiality issue

Following the Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 (EC, 2017), a project may partially opt out of the requirement for open publication of research data when confidentiality may be compromised.

In this case, we acknowledge that the research data have a character that require confidentiality, as they concern the opinions of employees in a specific company as well as the opinions and experiences of users of one of the enterprise's products.

It should also be noted that the case work was planned and conducted under the condition that the report from the case (HUMANE deliverable D3.3) would be a confidential deliverable. Without this confidentiality we assume that the data provided would be less rich in character and, therefore, less valuable as basis for the work presented in deliverable D3.3.

3.9.4   Archiving and preservation (including storage and backup)

The data are archived in anonymized form on an inhouse secure server at SINTEF.

Audio-recordings and non-anonymized transcripts have all been deleted. So has the participant list.

3.10 DS.C8.SINTEF. Decision support system dataset

The DS.C7 dataset consists of: Qualitative and anonymized interview-transcripts with four employees of an enterprise. The employees research automation in future decision support systems.

Anonymous data are items of information that cannot in any way identify individuals in the data material directly through names or personal ID numbers, indirectly through background variables, or through a list of names / connection key or encryption formula or code.

As part of the anonymization process for the interview transcripts, any references to named persons, the company role of the interviewee, and the specific background of the employee were removed.

Also the enterprise name and names of related services and projects were removed from the interview transcripts. Nevertheless, we assess it as potentially possible for readers of the interview transcripts to identify the case enterprise; though the individual interviewee cannot be identified. This because the company has a profile not shared by many other enterprises.

3.10.1 Dataset description

Origin of data: The data were collected by a SINTEF researcher as part of the HUMANE project.

Nature and scale of data: Transcripts of interview data in the language it was conducted (Norwegian).

To whom the dataset could be useful: Outside of the consortium, the data in its anonymized form might be useful for other researchers interested in decision support systems. However, the transcripts are in Norwegian which clearly delimits the usefulness of the data outside of Scandinavian countries.

Scientific publication: The results of the dataset has been included in HUMANE deliverable D3.3 and in a submitted conference paper (submitted version available on https://arxiv.org/abs/1702.07480)

Existence of similar datasets? To our knowledge qualitative datasets on researchers perspectives on automation in decision support systems are not openly available.

3.10.2 Standards and metadata

Not relevant, as the case needs to opt out from open sharing of research data (see next subsection).

3.10.3 Data sharing – opt out from open publication due to confidentiality issue

Following the Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 (EC, 2017), a project may partially opt out of the requirement for open publication of research data when confidentiality may be compromised.

In this case, we acknowledge that the research data have a character that require confidentiality, as they concern the opinions of employees in a specific company. In particular, the datasets provide reflections of the employees on ongoing research work.

It should also be noted that the case work was planned and conducted under the condition that the report from the case (HUMANE deliverable D3.3) would be a confidential deliverable. Without this confidentiality we assume that the data provided would be less rich in character and, therefore, less valuable as basis for the work presented in deliverable D3.3.

3.10.4 Archiving and preservation (including storage and backup)

The data are archived in anonymized form on an inhouse secure server at SINTEF.

Audio-recordings and non-anonymized transcripts have all been deleted. So has the participant list.

3.11 DS.RA1.ATC. Stakeholder survey dataset

A user survey was performed for gathering stakeholder feedback in order to help build the roadmaps in the domains of sharing economy, eHealth and citizens' participation. A common online survey was conducted, which included general questions, as well as specific questions for each domain.

This dataset consists of the anonymized raw-data from the survey. The survey was developed to systematize knowledge on stakeholder needs, expectations and previous experiences with human-machine networks. The survey included a combination of closed and open survey-questions.

The survey dataset was anonymized at source. That is, no personal data was collected as part of any questionnaires, except a broad demographic label. The participants only filled in their area of employment (public/private sector, academia, or other), and whether or not they are engaged professionally in one of the roadmap domains. Hence, the dataset cannot in any way be used in order to identify individuals in the data material directly through names or personal ID numbers, indirectly through background variables, or through a list of names / connection key or encryption formula or code.

For the survey data relating to roadmapping, data were anonymized at source. That is, no personal data was collected as part of any questionnaires, except a broad demographic label (identifying participant stakeholder group, such as *citizen*, *academic researcher* and so forth in the Citizen Participation survey). Since the participant pool cannot be separately identified (i.e., there is no list of their names and assumed participant stakeholder group), publication of the anonymized data on its own or in conjunction with any other dataset lead to the identification of any individual. Consequently, the anonymized survey data can be released along with other project data.

3.11.1 Dataset description

Origin of data: The dataset is collected in the HUMANE project, and is based on the survey launched at the beginning of January 2017 and lasted until the end of March 2017. The data in this dataset was collected and analysed by ATC and IT Innovation.

Nature and scale of data: Overall 75 people have participated in the online survey. The data is available as a CSV file and can hence be accessible and read with e.g. excel and SPSS.

To whom the dataset could be useful: Outside of the consortium, the data in its anonymized form might be useful for other researchers interested in the design and implementation of roadmaps related to the operation of human – machine networks.

Scientific publication: It is our objective to use the dataset as a basis for at least one scientific publication in the near future.

Existence of similar datasets? To our knowledge qualitative datasets related to the design of roadmaps for the implementation of human – machine networks are not openly available.

3.11.2 Standards and metadata

The dataset is characterized with the following metadata:

- Author/compiler of dataset: Marina Klitsi and Eva Jaho, ATC.
- Funded by: [HUMANE, H2020 – 645043]
- Format: [CSV]
- Content-data: human – machine networks operation, roadmaps
- Method of data accumulation: Survey
- Data collection period [from] – [to]: 03.01.2017 – 31.03.2017.
- Conditions of use of data: open access, free of charge.
- DOI: 10.5281/zenodo.571141
- Related publications n/a

3.11.3 Data sharing

Access procedures: The data is made accessible and available for re-use and secondary analysis by uploading the data to Zenodo.

Document format and availability: The dataset will be available as a CSV-file at the HUMANE community on Zenodo: https://www.zenodo.org/communities/humane2020/. From here the fully anonymized data are open accessible for anyone, free of charge.

This dataset is available at the following URL, with an embargo period until the end of May 2017: https://doi.org/10.5281/zenodo.571141.

3.11.4 Archiving and preservation (including storage and backup)

Archiving of the anonymized dataset at Zenodo guarantees a long-term and secure preservation of the data at no additional cost for the project. Zenodo informs that "*in the highly unlikely event that Zenodo will have to close operations, we guarantee that we will migrate all content to other suitable repositories, and since all uploads have DOIs, all citations and links to Zenodo resources (such as your data) will not be affected*".

3.12 DS.RA2.ATC. eHealth roadmapping interviews dataset

Separate focus groups/interviews were conducted to gather feedback as reported in the table below.

Number Focus group/interview	Number of participants	Stakeholders	Implementation Period
1 focus group on eHealth roadmap	5	Policy maker representatives	20 January 2017
1 interview on eHealth roadmap	2	Service owners – eHealth services	11 January 2017
1 focus group on eHealth roadmap	3	Researchers working in relevant eHealth projects	16 December 201

This dataset consists an anonymized summary of the feedback received for the eHealth roadmap.

3.12.1 Dataset description

Origin of data: The dataset is collected in the HUMANE project, and is based on the focus groups/interviews conducted in December 2016 and January 2017. The data in this dataset was collected and analysed by ATC.

Nature and scale of data: Overall 10 people have participated in the focus groups/interviews. The data is available as a PDF file.

To whom the dataset could be useful: Outside of the consortium, the data in its anonymized form might be useful for other researchers interested in the design and implementation of roadmaps related to the operation of human – machine networks in the field of eHealth.

Scientific publication: The data has been used as part of research submitted for scientific publication.

Existence of similar datasets? To our knowledge, there are no similar datasets available, except for other datasets generated in the HUMANE project for the other roadmaps (sharing economy and citizens' participation).

3.12.2 Standards and metadata

The dataset is characterized with the following metadata:

- Author/compiler of dataset: Marina Klitsi and Eva Jaho, ATC
- Funded by: [HUMANE, H2020 – 645043]
- Format: [PDF/A]
- Content-data: human – machine networks operation, roadmap in eHealth.
- Method of data accumulation: qualitative content analysis.
- Data collection period [from] – [to]: 16.12.2016 – 20/01/2017.
- Conditions of use of data: open access, free of charge
- DOI: 10.5281/zenodo.580294

- Related publications: see https://humane2020.eu/2017/05/15/a-roadmap-for-future-human-machine-networks-in-ehealth/

3.12.3 Data sharing

Access procedures: The data is made accessible and available for re-use and secondary analysis by uploading the data to Zenodo.

Document format and availability: The dataset will be available as a PDF-file at the HUMANE community on Zenodo: https://www.zenodo.org/communities/humane2020/. From here the fully anonymized data are open accessible for anyone, free of charge.

This dataset is available at the following URL, with an embargo period until the end of May 2017: https://doi.org/10.5281/zenodo.580294.

3.12.4 Archiving and preservation (including storage and backup)

Archiving of the anonymized dataset at Zenodo guarantees a long-term and secure preservation of the data at no additional cost for the project. Zenodo informs that "*in the highly unlikely event that Zenodo will have to close operations, we guarantee that we will migrate all content to other suitable repositories, and since all uploads have DOIs, all citations and links to Zenodo resources (such as your data) will not be affected*".

3.13 DS.RA3.SINTEF. sharing economy roadmapping interviews dataset

The DS.RA3 dataset consists of: Qualitative and anonymized interview-transcripts with (a) sharing economy service owners, (b) researchers with a special research interest in the sharing economy, and (c) policy maker representatives from government and stakeholder organisations.

Anonymous data are items of information that cannot in any way identify individuals in the data material directly through names or personal ID numbers, indirectly through background variables, or through a list of names / connection key or encryption formula or code.

As part of the anonymization process for the interview transcripts, any references to named persons, the company role of the interviewee, and the specific background of the employee were removed.

Also the company name and names of company services and projects were removed from the interview transcripts. Nevertheless, we assess it as potentially possible for readers of the interview transcripts to identify the organizations of the interviewees; though the individual interviewee cannot be identified. This because the organizations has a profile not shared by many other organizations.

3.13.1 Dataset description

Origin of data: The data were collected by SINTEF researchers as part of the HUMANE project, and in collaboration with the Norwegian innovation project Conserve and Consume (HUMANE case 2).

Nature and scale of data: (1) Transcripts of interview data in the language it was conducted (Norwegian).

To whom the dataset could be useful: Outside of the consortium, the data in its anonymized form might be useful for researchers with a research interest in the sharing economy. However, the transcripts are in Norwegian which clearly delimits the usefulness of the data outside of Scandinavian countries.

Scientific publication: The results of the dataset has been included in HUMANE deliverables D4.2 and D4.4. One journal paper presenting findings from the dataset is planned to be submitted within 2017.

Existence of similar datasets? To our knowledge qualitative datasets on the human-machine networks of the sharing economy are not openly available.

3.13.2 Standards and metadata

Not relevant, as the case needs to opt out from open sharing of research data (see next subsection).

3.13.3 Data sharing – opt out from open publication due to confidentiality issue

Following the Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 (EC, 2017), a project may partially opt out of the requirement for open publication of research data when confidentiality may be compromised.

We acknowledge that the dataset DS.RA3 has a character that require confidentiality, as the interviews include rich descriptions of the interviewees' reflections on internal processes and business critical aspects and of the involved organisations.

Without this confidentiality we assume that the data provided would be less rich in character and, therefore, less valuable as basis for the work presented in the deliverables D4.2 and D4.4.

3.13.4 Archiving and preservation (including storage and backup)

The data are archived in anonymized form on an inhouse secure server at SINTEF.

Audio-recordings and non-anonymized transcripts have all been deleted. So has the participant list.

3.14 DS.RA4.IT Innovation. simulation and modelling dataset

This dataset was derived from the publicly available data from Wikipedia, for the purposes of simulation modelling in the HUMANE project.

The data was derived from the Simple English language edition of Wikipedia, made available via the WikiWarMonitor project: http://wwm.phy.bme.hu/light.html

The data we use contains information about the chosen nick-name of people who contribute to Wikipedia, or IP addresses of anonymous contributions. This could be used to trace contributions to an individual. However, this information is already available in the public domain, viewable to anybody visiting a Wikipedia article or using the Wikipedia datasets made available to the public via Wikidata: https://www.wikidata.org/. Users have deliberately put data online publicly. We do not expose information that is not already in the public domain.

Consent already collected by Wikipedia in their terms and conditions for using their services: https://wikimediafoundation.org/wiki/Terms_of_Use

All content is available under a free public domain licence: https://creativecommons.org/publicdomain/zero/1.0/

3.14.1 Dataset description

Origin of data: Wikipedia data is freely available via the Wikidata[1] website, which was processed to generate a smaller, derived, dataset in the WikiWarMonitor project[2] (part of the EC FP7 ICTeCollective project). This dataset has been derived by IT Innovation from the latter source in the HUMANE project.

Nature and scale of data: The data is considered large scale, comprising data about 74,880 articles and contributions to these articles by approximately 150,000 users of Wikipedia (including bots). The dataset was prepared so as to be able to bootstrap a simulation model for Wikipedia.

To whom the dataset could be useful: Outside of the consortium, the data might be useful for other researchers interested in simulation modelling of human-machine networks, especially researchers wishing to reproduce results from the HUMANE project.

Scientific publication: The data is intended to be used as part of research submitted for scientific publication.

Existence of similar datasets? Yes, this dataset is derived from existing Wikipedia data, as noted above.

---

[1] https://www.wikidata.org/
[2] http://wwm.phy.bme.hu/index.html

3.14.2 Standards and metadata

The dataset is characterized with the following metadata:

- Author/compiler of dataset: Vegard Engen and Juri Papay, University of Southampton IT Innovation Centre
- Funded by: [HUMANE, H2020 – 645043]
- Format: [CSV]
- Content-data: Statistics of contributions by human and machine actors to the Simple English language edition of Wikipedia.
- Method of data accumulation: N/A.
- Data collection period [from] – [to]: N/A.
- Conditions of use of data: open access, free of charge
- DOI: 10.5281/zenodo.573223
- Related publications: N/A

3.14.3 Data sharing

Access procedures: The data is made accessible and available for re-use and secondary analysis by uploading the data to Zenodo.

Document format and availability: The dataset will be available as a collection of CSV files at the HUMANE community on Zenodo: https://www.zenodo.org/communities/humane2020/ . From here the data is open accessible for anyone, free of charge.

This dataset is available at the following URL, with an embargo period until the end of December 2017: https://doi.org/10.5281/zenodo.573223

3.14.4 Archiving and preservation (including storage and backup)

Archiving of the anonymized dataset at Zenodo guarantees a long-term and secure preservation of the data at no additional cost for the project. Zenodo informs that *"in the highly unlikely event that Zenodo will have to close operations, we guarantee that we will migrate all content to other suitable repositories, and since all uploads have DOIs, all citations and links to Zenodo resources (such as your data) will not be affected"*.

4   Open access to publications

 In the HUMANE project, open access to publications based on the project results has been achieved through publication either in gold open access journals or through publishers allowing green open access through self-archiving of manuscripts.

Furthermore, upon submission the pre-print version of the manuscripts have been made public through arxiv.org.

Table 2 gives an overview of the hitherto submitted and/or published HUMANE publications and their open access.

Table 2: Overview of HUMANE peer-reviewed publications and open access options

Publication	Status	Open access option
Tsvetkova, M. et al., (2017). Understanding Human-Machine Networks: A Cross-Disciplinary Survey. ACM Computing Surveys, 50(1), Article No. 12, doi:10.1145/3039868	Published	Green open access. Author version of accepted manuscript available at arXiv:1511.05324v2
Engen, V., Pickering, J. B., & Walland, P. (2016). Machine agency in human-machine networks; impacts and trust implications. In Proceedings of the International Conference of Human-Computer Interaction - HCI International 2016 (pp. 96-106). Springer International Publishing.	Published	Green open access. Author version of pre-print manuscript available at arXiv:1602.08237 (Author version of accepted manuscript to be made openly available after 12 month embargo)
Tsvetkova, M., García-Gavilanes, R., & Yasseri, T. (2016). Dynamics of Disagreement: Large-Scale Temporal Network Analysis Reveals Negative Interactions in Online Collaboration. Scientific Reports, 6, Article No. 36333.	Published	Gold open access.
Eide, A. W., et al.. (2016). Human-machine networks: towards a typology and profiling framework. In: Proceedings of the International Conference of Human-Computer Interaction - HCI International 2016 (pp. 11-22). Springer International Publishing.	Published	Green open access. Author version of pre-print manuscript available at arXiv:1602.07199 (Author version of accepted manuscript to be made openly available after 12 month embargo)
Lüders, M. (2016). Innovating with users online? How network-characteristics affect collaboration for innovation, Journal of Media Innovations, 3(1), 4-22.	Published	Gold open access

Rudas, C., Surányi, O., Yasseri, T., & Török. J. (2017). Understanding and coping with extremism in an online collaborative environment. PLoS ONE, 12(3), e0173561.	Published	Gold open access
Tsvetkova, M., Garcia, R., Floridi, L, & Yasseri, T. (2017). Even good bots fight. PLoS ONE 12(2), e0171774.	Published	Gold open access
Pickering, J. B., Engen, V., & Walland, P. (2017). The Interplay between Human and Machine Agency. Accepted for publication at the International Conference of Human-Computer Interaction - HCI International, 2017	Accepted for publication	Green open access. Author version of pre-print manuscript available at arXiv:1702.04537 (Authors version of accepted manuscript to be made openly available after 12 month embargo)
Følstad, A., Engen, V., Haugstveit, I. M., & Pickering, B. (2017). Automation in Human-Machine Networks: How Increasing Machine Agency Affects Human Agency.	Accepted for publication	Green open access. Author version of submitted manuscript available at arxiv:1702.07480 (Authors version of accepted manuscript to be made openly available after 12 month embargo)
Ibrahim, K., Khodursky, S., & Yasseri, T. (submitted). Spatiotemporal patterns of classifications to the Zooniverse.	Under review	Gold open access. Authors version of the accepted version will be available on arXiv.

5 Conclusions

In this report we have described how we have complied with the requirements imposed on HUMANE as a participant in the Open Research Data pilot with regard to open access to research data and open access to publications.

The project partners have used Zenodo as the open research data repository. Project data made openly published Zenodo will be linked to a HUMANE project-site at OpenAIRE. Scientific publications generated in HUMANE is made available open access, through self-archiving of authors versions of the submitted manuscripts at arxiv.org and publishing with gold or green open access publication channels.

Details on the gathering and publication status of the HUMANE datasets was provided in Section 3. Here, we described the details of open publication for nine of the 14 gathered datasets. Furthermore, we provided the details of gathering and storing for five of the datasets which we assessed as necessary to opt out from the requirement of open data publication. The reason for this was issues pertaining to confidentiality. Furthermore, it may be noted that the open publication of data from qualitative interviews in the form of anonymized transcripts might also require a more extensive procedure for gathering informed consent than what was envisioned in two of the HUMANE cases. For these, both issues pertaining to confidentiality and issues pertaining to informed consent procedures made us judge against open publication of the datasets.

We trust that the openly published research data and open access to research publication will be useful to other researchers.

6 References

European Commission (2013a). Guidelines on Open Access to Scientific Publicaitons and Research Data in Horizon 2020.

European Commission (2013b). Guidelines for Data Management in Horizon 2020, Version 1.0.

European University Institute (2015). Research Data Guide, Version 3.0. Retrieved from http://www.eui.eu/Documents/Research/Library/ResearchGuides/Economics/PDFs/EUIResearchDataServicesGuide3.0July2015.pdf