Holden Deliverable

D7.4 – Final Data Management Plan

Grant Agreement number	101099491
Action Acronym	HOLDEN
Action Title	Ethical Design of Holography with Dense wireless Networks
Type of action	HORIZON-EIC-2022-PaTHFINDEROPEN-01
Version date of the Annex I against which the assessment will be made	13/12/2022
Start date of the Project	1/6/2023
Due date of the deliverable	31/05/2026
Actual date of submission	29/05/2026
Lead beneficiary for the deliverable	AALTO
Dissemination level of the deliverable	Public

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Innovation Council and SMEs Executive Agency (EISMEA). Neither the European Union nor the granting authority can be held responsible for them.

European Innovation Council

Authors in alphabetical order		
Name	**Beneficiary**	**e-mail**
Dariush Salami	AALTO	dariush.salami@aalto.fi
Ying Liu	AALTO	ying.2.liu@aalto.fi
Yuqing Song	AALTO	yuqing.song@aalto.fi
Stephan Sigg	AALTO	stephan.sigg@aalto.fi

Change history				
Version	**Date**	**Status**	**Partner**	**Description**
0.1	04/12/2023	Initial	AALTO	Baseline D7.3 data management plan submitted.
0.9	15/03/2026	Updated draft	AALTO	End-of-project consolidation based on WP1–WP9 deliverables.
1.0	15/03/2026	Final	AALTO	Final project-state data management plan for submission.

Abstract

This final Data Management Plan updates the initial D7.3 and records the end-of-project status of data handling in HOLDEN. It reflects the datasets, software, simulation outputs, benchmark material, ethical documentation, dissemination assets, and restricted personal-data processes described across the technical, ethics, exploitation, and validation deliverables.

Contents

1 Data Management

1.1 Scope and final-project update

This deliverable is the final version of the HOLDEN Data Management Plan. It supersedes the initial D7.3 plan by replacing forward-looking statements with the final project-state description of data handling across the consortium.

The update is based on the outputs reported in the technical work packages (WP3–WP6), dissemination and exploitation work (WP7), ethics and compliance work (WP8), and late validation and prototype-related work (WP9). In particular, the update incorporates the data categories and governance implications described in the static and dynamic holography deliverables, the gesture-recognition and machine-learning deliverables, the functional and privacy-profile work, the IPR and dissemination plans, the ethics requirement deliverables, the DPIA, the ethics status monitor, and the study- and prototype-related documentation.

1.2 Data summary

HOLDEN generated and managed several distinct classes of research outputs rather than one single dataset. The final data inventory is summarized in Table 1.

Table 1: Final overview of data categories in HOLDEN

Category	Typical contents	Typical formats	Access status at project end
External baseline data	Reused public RF datasets for method development and benchmarking, including binary I/Q and point-cloud data	dat, txt, csv, pkl, dataset documentation	Public, subject to source repository terms
Measured RF data	CSI, RSS, I/Q samples, hologram frames, point clouds, localization and activity traces from lab and pilot measurements	csv, dat, txt, mat, npz, pkl	Mixed: selected subsets open, sensitive/raw subsets restricted
Simulated and synthetic data	FEKO-based electromagnetic simulations, synthetic hologram data, generated point clouds, augmentation data, human phantom simulations	mat, npz, csv, obj, stl, configuration files	Mostly shareable after review, unless tied to restricted validation setups
Labels and annotations	Gesture labels, motion classes, localization targets, experimental conditions, benchmark labels, privacy-profile parameters	csv, json, txt, xlsx	Mixed: open with released datasets, otherwise restricted with parent dataset
Software and models	Pre-processing pipelines, simulation code, analysis scripts, learning models, checkpoints, benchmark code, demonstrator software	py, jl, m, tex, ipynb, yaml, json, model checkpoints	Mixed: open where possible, otherwise restricted by IPR, security, or exploitation constraints
Documentation and governance outputs	DMPs, ethics reports, DPIA, ESM, dissemination and IPR documents, installation records, protocols, meeting and review material	pdf, docx, xlsx, tex	Public or sensitive depending on deliverable classification
Restricted personal data	Consent forms, contact details, limited website logs, and in exceptional cases interview or study-management records	paper forms, txt, docx, server logs	Restricted; not openly shared

The most important scientific data types created during the project were static and dynamic RF holography measurements and reconstructions; CSI, RSS, and binary I/Q time-series from radio channel measurements; point-cloud representations derived from mmWave and related RF sensing pipelines; simulation outputs for indoor scenarios, antenna-array configurations, passive-tag localization, and privacy-

aware data-reduced acquisition; labeled benchmark data for gesture recognition, localization, people counting, and motion/activity analysis; and trained machine-learning models, synthetic-data generators, and evaluation outputs.

The project also produced non-data research outputs that must still be managed under FAIR-oriented principles, including algorithms, scripts, model architectures, benchmark protocols, demonstrator descriptions, privacy profiles, ethics analyses, and dissemination assets.

1.3 Origin and purpose of the data

The project built on existing public RF sensing resources and also generated new project data.

Reused data The initial D7.3 and the DPIA identify two main reused sources for early development and benchmarking: the public binary I/Q and point-cloud dataset previously referenced through Zenodo record 4459969, and the Aalto-hosted `sidelinkchanneldataset` repository.

These reused data supported method design, benchmarking, and early model development before or alongside the creation of new HOLDEN-specific data.

Project-generated data Final deliverables show that new data were created in several work packages. WP3 produced indoor scenario simulations, anechoic-chamber and controlled measurement data, passive-holography acquisition data, and synthetic or simulated outputs for imaging and localization. WP4 produced structured hologram stop-motion frames, body-motion databases, received-power or CSI datasets, and dense-array simulations for dynamic holography. WP5 produced massive RF learning inputs, point-cloud datasets, multi-angle gesture datasets, synthetic orientation data, learned representations, and model-evaluation outputs. WP6 produced benchmark and application-validation material, scenario definitions, privacy profiles, and demonstrator-oriented configuration and evaluation data. WP7 and WP8 produced dissemination, IPR, ethics, website, and compliance-related records, including the DPIA and Ethics Status Monitor. WP9 produced prototype-installation, study-preparation, and evaluation documents, some of which include restricted study-management information.

The purpose of these data was to support the project objectives on privacy-aware RF holography, distributed sensing, machine-learning-based recognition, ethical and privacy-compliant application design, and exploitation/dissemination of results.

1.4 Data volume

The initial project estimate of up to the TB scale remains appropriate for the full project lifecycle when raw RF captures, repeated measurements, simulations, derived tensors, and trained-model artifacts are considered together. In practice, documentation, governance, and ethics material are comparatively small; processed benchmark datasets and trained models are moderate in size; and raw measurements, dense-array simulations, and high-dimensional RF or point-cloud data are the main contributors to storage volume.

Not all raw captures are preserved permanently. Temporary or low-value acquisition traces used only for setup verification, connection maintenance, debugging, or failed runs are candidates for deletion after quality control.

2 FAIR Data

2.1 Making data findable, including provisions for metadata

At project end, findability is implemented through repository-level metadata, structured naming, deliverable cross-references, and dataset-level documentation.

For internally curated scientific datasets, the consortium uses descriptive names that encode at least the acquisition context, time or version, data modality, and experiment or label class. The naming convention introduced in D7.3 remains valid in principle:

```
<MeasurementSite>_<timestamp>_<DataType>_<ClassLabel>_<DeviceID>_<AntennaNo>
```

Where this exact pattern is not suitable, an equivalent structured convention is used that still records provenance and acquisition context.

Metadata recorded with datasets typically includes the work package and task of origin, the data modality and sensing setup, the acquisition date or version, the relevant device, antenna, and software configuration, the main preprocessing and annotation steps, the responsible partner and contact point, and the access status, licence status, and reuse restrictions.

Project outputs that are publicly released are intended to be deposited in repositories that provide persistent identifiers or stable repository URLs. Zenodo remains the preferred public deposition platform for datasets and software snapshots requiring durable citation. Institutional or consortium repositories such as `version.aalto.fi` remain suitable for development, collaboration, and controlled distribution.

Documentation is provided through README files, data dictionaries, experiment notes, or deliverable references. For structured datasets, simple open metadata files such as `txt`, `csv`, `json`, or repository landing-page descriptions are preferred to minimize tool dependence.

2.2 Making data accessible

Accessibility in HOLDEN is governed by a staged-release logic rather than unconditional openness for every artifact.

Publicly accessible outputs The following output classes can be made openly accessible where no conflicting constraint applies: selected datasets that are sufficiently anonymized or non-personal; software, scripts, notebooks, and benchmark material; scientific publications and public deliverables; and non-sensitive demonstrator descriptions and dissemination assets.

Public release channels include Zenodo, institutional repositories such as `version.aalto.fi`, open publication channels, preprint archives, and university archives such as ACRIS/Aaltodoc, as well as links from publications, project pages, or EOSC-facing dissemination where relevant.

Restricted outputs Not all project data can be openly released. Access remains restricted when the data contains personal data or quasi-identifiers, when it is needed to protect confidentiality, security, or partner obligations, when it is tied to exploitation, patenting, or other IPR-sensitive activities, when the raw data is not meaningful without substantial contextual infrastructure and would create disproportionate documentation and compliance cost, or when the data concerns vulnerable users, pilot installations, or restricted study environments.

Restricted data remains stored on partner-controlled institutional infrastructure with access limited to the responsible researchers and authorized supervisors or managers. Access is granted on a need-to-know basis and logged or governed by normal institutional access-control mechanisms.

Authentication and access control The project does not foresee a separate central data-access committee for ordinary scientific outputs. Instead, access decisions for non-public material are made by the owning partner in line with the Grant Agreement, Consortium Agreement, institutional policy, ethics approvals, and the DPIA. For personal-data-related material, access is limited to the smallest possible number of authorized persons.

Open access to publications The dissemination plan and project practice support open access to publications. Publications are linked, where possible, to the underlying non-sensitive data and software that support reproducibility. Green open access through institutional archiving and preprint sharing remains part of the dissemination strategy.

2.3 Making data interoperable

The project uses a combination of domain-specific and general-purpose data formats. Interoperability is increased by favoring widely used, documented, and tool-neutral formats whenever possible.

Table 2: Representative data and software formats used in HOLDEN

Type	Formats	Interoperability approach
Tabular measurements and labels	csv, txt, dat, xlsx	Prefer csv/txt for release, document units, column names, and label meaning
Scientific arrays and tensors	mat, npz, pkl	Release processed versions in documented formats where feasible; document serialization dependencies
Simulation geometry and configuration	obj, stl, yaml, json, plain-text configs	Preserve software version and configuration context
Software and analysis pipelines	py, jl, m, ipynb, tex	Store with dependency notes, version tags, and example usage
Documentation and governance	pdf, docx, tex, txt	Preserve final public versions and working sources when useful

Interoperability is further supported by explicit description of units, coordinate systems, antenna configurations, and acquisition conditions; separate documentation of pre-processing, denoising, segmentation, and feature-extraction steps; consistent use of labels for gesture classes, activities, localization targets, and privacy-profile settings; and references to related publications and deliverables that explain the data-generation pipeline.

The project considered general metadata approaches such as Dublin Core, RDF, and DDI as useful reference points for documentation. In practice, a lightweight and domain-appropriate metadata strategy is preferred over heavy formalism when this improves usability for downstream researchers.

2.4 Increasing data re-use

At project end, re-use is supported by documentation, provenance tracking, repository publication where possible, and clear separation between open and non-open material.

Documentation quality Reusable datasets and software should be accompanied by a short description of scientific purpose and scope, data-acquisition or simulation conditions, variable descriptions and units, pre-processing and labeling methodology, software requirements and example loading instructions, and a note on limitations, known biases, and conditions under which the data should not be over-interpreted.

Licensing and rights The original D7.3 planned broad open release after anonymization and exploitation review. The final project position remains aligned with that principle, but licensing is applied per output category. Software is released, when appropriate, under an open-source software licence selected by the owning partner. Datasets and documentation are released, when appropriate, under a content or data licence suitable for the material and any imposed restrictions. Where openness conflicts with privacy, confidentiality, security, or exploitation, the output may remain restricted even if a metadata record is made public.

IPR management follows the HOLDEN Grant Agreement, Consortium Agreement, and the guidance summarized in D1.1. Ownership of results remains with the generating partner unless otherwise agreed, and open release must not prejudice legitimate protection or exploitation measures.

Quality assurance Data quality is addressed throughout the data lifecycle through experiment planning to avoid unnecessary redundancy or unusable captures, verification of measurement setups and simulation parameters, cleaning of incomplete or corrupted files, separation of temporary or debugging material from preserved scientific outputs, and review of documentation prior to publication or long-term archiving.

The technical deliverables also show an important additional re-use mechanism: synthetic and simulation-based generation is used to reduce dependence on large personal-data collections while still supporting algorithm validation and benchmarking.

3 Other Research Outputs

The final DMP covers not only datasets but also other research outputs generated during HOLDEN, including source code for simulation, preprocessing, learning, benchmarking, and visualization; trained models, checkpoints, and learned representations; benchmark definitions, evaluation scripts, and test protocols; demonstrator descriptions, privacy profiles, and scenario specifications; scientific manuscripts, preprints, posters, videos, and dissemination assets; and ethical guidance outputs such as the DPIA, ESM versions, privacy analyses, and appropriation-related materials.

These outputs are managed according to the same general principles as data: public where possible, restricted where privacy, security, partner obligations, or exploitation require it, and documented sufficiently for inspection, re-use, or audit where release is allowed.

4 Allocation of Resources

Responsibility for data management is distributed but coordinated. AALTO coordinates project-level data-management planning and the preservation strategy of this deliverable. Each partner remains responsible for the secure handling, curation, documentation, and lawful sharing of the data and outputs it generates. WP leaders are responsible for ensuring that major datasets, models, and code connected

to their deliverables are documented and stored appropriately. Ethical and compliance work in WP8 provides project-wide constraints and recommendations that affect collection, processing, retention, publication, and deletion decisions.

The project mostly relies on existing institutional infrastructure and established public repositories, so no exceptional dedicated platform cost is foreseen for basic publication and preservation. The main resource burden lies in researcher time needed for documentation and metadata preparation, anonymization and compliance review, repository deposition and software packaging, and retention, review, and deletion of restricted material after project closure.

The coordinator and responsible partners will ensure that preserved public outputs remain discoverable for at least five years after project end, and longer where repository policy or institutional archiving permits.

5 Data Security

5.1 Storage and backup

Throughout the project, working data has been stored on partner-controlled institutional systems, including Aalto-managed infrastructure such as `version.aalto.fi`, and on equivalent institutional environments used by other beneficiaries. These services provide access control and routine backup practices suitable for research data.

5.2 Security principles applied in the project

The final project-state security model is based on least-privilege access to working data, separation of personal identifiers from research data, storage of restricted material only on partner-controlled systems, avoidance of unnecessary duplication of sensitive data, deletion of temporary or unnecessary raw material after its purpose has been fulfilled, and publication only after anonymization and IPR or compliance review.

5.3 Security by data class

The project handled different data classes differently. Open scientific outputs can be published in public repositories, internal scientific working data remains on institutional systems with consortium-partner access as needed, and restricted personal-data-related material is only accessible to specifically authorized staff and is never published openly.

5.4 Temporary and low-value data

Consistent with the initial D7.3, low-value traces used only for calibration, connection maintenance, failed runs, or debugging do not need to be preserved long term. This reduces storage volume and lowers unnecessary risk exposure.

6 Ethics and personal data protection

6.1 General ethics framework

Ethics and data protection are central to HOLDEN. The project's final data-management position is informed by the ethics requirements deliverables, the DPIA, the ethics advisory reports, and the Ethics Status Monitor.

The key principle is that RF sensing, although privacy-advantaged relative to cameras in many settings, can still create meaningful privacy, autonomy, profiling, bias, and surveillance risks. Therefore, the project does not equate anonymization alone with complete ethical adequacy. Data minimization, scope limitation, contextual safeguards, and use restrictions remain necessary.

6.2 Human participants

Where human participants were involved, participation was voluntary and based on informed consent. Vulnerable groups and minors were excluded unless a separate, explicitly approved study design stated otherwise. Study spaces were bounded and signposted where appropriate, participants were informed of their right to withdraw and request deletion where still technically possible, and signed consent documentation was stored separately from scientific research data.

The DPIA further clarifies that consent forms with names and contact details should be deleted shortly after the end of the study, and that anonymized research data may be retained for further scientific use where re-identification is not reasonably possible.

6.3 Types of personal data actually handled

Across the project, the relevant personal-data categories were limited and strongly controlled. They comprised consent forms and participant contact information, minimal participant statistics needed at aggregate level for study documentation, limited website log data such as IP address, browser, and page requests, and in specific validation or ethics-study contexts restricted interview or study-management material where applicable.

The scientific RF datasets themselves were designed to avoid direct identifiers. The project states repeatedly that names, age, and gender are not stored together with individual RF records. Where aggregate demographic statistics are needed, they are kept in non-linkable form.

6.4 Special handling of ground truth and visual material

The early D7.3 text anticipated possible temporary video use in some studies, whereas the later DPIA adopts a stricter minimization approach and states that no image or video ground-truth is required for the main RF studies. The final project position follows the later, more privacy-preserving interpretation: image and video data are not part of the normal retained scientific dataset in HOLDEN; if visual material is exceptionally used for setup verification or restricted study support, it must remain under tightly limited access and be deleted as soon as the immediate purpose is fulfilled; and openly released datasets must not contain directly identifying visual material.

6.5 Anonymization and pseudonymization

The project implemented anonymization and pseudonymization through separation of direct identifiers from measurement data, assignment of anonymous or pseudonymous participant codes, storage of any

linkage information outside the scientific dataset and only for as long as needed, and design choices that favor privacy-preserving sensing representations, for example point-cloud abstractions and direction-agnostic representations that reduce identifiable detail.

The project also explored technical privacy-enhancement approaches such as data reduction, targeted sensing, ghosting/obfuscation, privacy filters, and synthetic-data generation to reduce the need for broad or intrusive real-world collection.

6.6 Data minimization

Data minimization is a core principle across the final deliverables: collect only the information necessary for the specific research question, prefer synthetic or simulated data where scientifically adequate, prefer processed or abstracted data representations over rich raw recordings where possible, avoid retaining identifiers once no longer needed for lawful study management, and limit secondary use and data sharing beyond the originally approved scope.

7 Retention, archival, and deletion

7.1 Preserved outputs

The following outputs are candidates for medium- to long-term preservation: public datasets and software released during or immediately after the project, final public deliverables and linked supporting material, curated scientific datasets and code required to support publications or follow-up research, and non-personal simulation, benchmark, and model artifacts with continuing research value.

7.2 Restricted retained outputs

Some restricted outputs may still be retained after project end if lawful and necessary, including internal project records required for audit or reporting, restricted technical artifacts needed for exploitation or follow-up work, and anonymized or strongly de-identified research data retained for future scientific work.

7.3 Deletion or disposal

The following materials should not be kept longer than necessary: direct identifiers such as consent forms and contact lists once their retention purpose has expired, temporary raw captures, debugging traces, or low-value setup data, and restricted material for which no lawful retention basis remains.

Deletion is carried out by the owning partner according to institutional policy, ethics approvals, and applicable data-protection rules. Physical consent documents, if used, are destroyed through the institution's standard confidential-disposal process.

8 Other Issues

No project-wide need was identified for a bespoke departmental data-management framework beyond the Grant Agreement, Consortium Agreement, institutional procedures, ethics approvals, and the controls documented in HOLDEN deliverables. However, the later ethics work makes clear that data governance cannot be reduced to a one-time checklist. The final position is therefore that data management in HOLDEN must remain sensitive to context, misuse scenarios, and the ethical implications of aggregation, repurposing, and deployment.

9 Conclusions

HOLDEN did not produce only one open dataset; it produced an ecosystem of scientific data, software, simulations, models, benchmark artifacts, and governance records with different sensitivity levels and publication paths. The final Data Management Plan therefore adopts a differentiated approach: open release for non-sensitive and reusable outputs where possible, restricted handling for personal, confidential, security-relevant, or exploitation-sensitive material, strong emphasis on data minimization, privacy-preserving representations, and separation of identifiers from research data, and preservation of curated outputs with continued scientific or exploitation value together with deletion of unnecessary sensitive material.

This final version records the end-of-project state and the principles that govern post-project preservation, access, and disposal.

A Annex 1 – Privacy Notice for HOLDEN website

A.1 Privacy Notice

The HOLDEN website supports communication and dissemination activities of the project. It therefore processes a limited amount of personal data related to website visitors.

Aalto University, as project coordinator, is responsible for the execution and maintenance of the website and serves as the contact point for data subjects with questions about website-related personal-data processing.

A.2 Why the website processes personal data

Personal data is processed in order to enable access to the website, maintain and improve the website, enable communication related to project activities, monitor visitor traffic as part of dissemination-impact assessment, and maintain data security and detect misuse.

A.3 What personal data is processed

The website logs may contain the IP address, browser information, and requested pages or page requests.

A.4 Legal basis

The legal basis is legitimate interest in operating the project website, communicating project results, assessing dissemination impact, and maintaining website security.

A.5 Sharing and transfers

Personal data is shared only where necessary for website operation, service provision, or legal compliance. The preferred hosting location is within the EU or EEA. If third-country processing occurs through service provision, it must comply with applicable law.

A.6 Retention

Website-related personal data is retained only for as long as necessary for the stated purpose and applicable institutional practice.

A.7 Rights of data subjects

Visitors may have rights of access, rectification, erasure, objection, and restriction, subject to the legal basis and applicable law.

A.8 Contact

For website privacy questions, data subjects may contact the project contact point at Aalto University. For data-protection matters, Aalto's Data Protection Officer may be contacted through the University's official channels.