Deliverable 1.2:
Data Management Plan

Martín López Nores (UVIGO), Ákos Lencsés (KIFÜ) and Ildikó Kádárné Kelemen (KIFÜ)

This document describes the data management policy and life cycle for all datasets that will be collected, processed or generated by the project.

rurAllure
reach out!

This document is part of a project that has received funding from the **European Union's Horizon 2020** research and innovation programme under grant agreement No 101004887


rurAllure
reach out!

Project acronym	rurAllure
Full title	Promotion of rural museums and heritage sites in the vicinity of European pilgrimage routes
Grant agreement number	101004887
Funding scheme	Coordination and Support Action (CSA)
Work programme topic	SOCIOECONOMIC AND CULTURAL TRANSFORMATIONS IN THE CONTEXT OF THE FOURTH INDUSTRIAL REVOLUTION (H2020-SC6-TRANSFORMATIONS-2018-2019-2020)
Project start date	2021-01-01
Project duration	36 months
Work Package	WP1 – Project Management
Deliverable lead organisation	Universidade de Vigo (UVIGO)
Authors	Martín López Nores (UVIGO), Ákos Lencsés and Ildikó Kádárné Kelemen (KIFÜ)
Reviewers	Jaime Padilla (GVAM) and Uxío Novo (UDC)
Version	1.1 (Fixing errata about data repositories from Hungary in the Annex)
Status	For EU review
Dissemination level	Public
Due date	M6 (2021-06-30)
Delivery date	2021-06-30
Updated to v1.1	2021-09-16

Table of contents

1. Introduction

According to Article 29.3 of the Grant Agreement ("Open access to research data"), the rurAllure Consortium must comply with the following obligations regarding the digital research data generated in the project (henceforth, "data") unless doing so would jeopardize the achievement of its goals:

- On the one hand, the project partners must deposit in a research data repository the data (including associated metadata) needed to validate the results presented in scientific publications. They must take measure to allow third parties to access, mine, exploit, reproduce and disseminate the data, free of charge for any user. Other data that might have research value should be included, too.
- On the other hand, the partners must provide information (via the chosen repository) about tools and instruments at the disposal of the beneficiaries and necessary for validating the project's results. Where possible, they must provide the tools and instruments themselves, too.

The partners are not expected to share sensitive data or breach any IPR agreements with industrial partners. It is not necessary to deposit all the data generated during the project either; only data that underpin published research findings and/or have longer-term value.

This document, the Data Management Plan (DMP) for rurAllure, describes the data management life cycle for the data to be collected, processed and/or generated by the project, as part of making research data findable, accessible, interoperable and re-usable (FAIR). According to the Guidelines on Data Management specified in Horizon 2020:

- The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the Consortium with regard to all the datasets that will be generated by the project.
- The DMP is not a fixed document, but evolves during the lifespan of the project. An initial version must be provided at the beginning of the project that may be updated during mid-term and towards the end of the project.
- The DMP should describe all the datasets generated or collected during the project, reflecting the current status of reflection within the Consortium about the data that will be produced.

2. Characteristics of rurAllure affecting the type of research data and their management

The main objective of rurAllure is to help overcome the underdevelopment of rural areas in the vicinity of pilgrimage routes through the allure of their cultural heritage. In essence, this is to be done by building a network of cultural heritage sites and promoting their resources and initiatives through the project's IT platform. Aside from this, it is necessary to have a supporting legislative environment in the project regions, and therefore national and regional policies related to pilgrimage must be analyzed before providing recommendations nearing the end of the project.

The type of datasets generated during the project's lifetime will help analyze pilgrim tourism in the context of the four pilots, plus those of other actions that may be started in other countries once Milestone 1 has been achieved in M12. In particular, data will be collected about the following:

- Places the pilgrims may visit, represented by Point of Interest (POI) info sheets (henceforth, Dataset 1 or DS1).
- Aggregated statistics from the profiles of pilgrims using the IT platform (DS2).
- Aggregated statistics from planned activities, including the pilgrims' willingness to accept recommendations and take detours (DS3).
- Anonymous answers to profiling questionnaires, describing actual travel characteristics and preferences (DS4).
- Lists of laws and policies related to pilgrimage, cultural routes, tourism and rural development classified by NUTS levels (DS5).

The datasets will contain periodic (e.g., monthly) snapshots of the information in order to enable research not only about the data accumulated up to the present moment, but also about their evolution over time. Data will be most commonly in csv format (i.e. plain text, with comma-separated values):

- DS1, DS2, and DS3 will contain monthly-generated csv files, starting once the technological platform is ready, after M12.
- DS4 will generate results according to survey progress, starting in M7. Monthly csv files will be shared periodically, subject to having sufficient data available to avoid any possible statistical disclosure of personal data.
- DS5 will be updated whenever new data is added.

At this stage, the amount of data is not predictable, but most probably the file sizes will be up to tens of megabytes.

All datasets will be published under CC-BY license with rich metadata, with no embargo period.

3. Selection of data repositories and naming conventions

The European Commission policies encourage the use of certified repositories or trusted community-recognized repositories as the first choice for depositing data. Researchers are recommend to make use of repositories endorsed by the research community in particular fields, such as, for example, life sciences (ELIXIR, https://elixir-europe.org/), social sciences (CESSDA, https://www.cessda.eu/) or the Humanities (DARIAH, https://www.dariah.eu/tools-services/tools-and-services/). Where a community-recognized repository does not exist, then it is possible to use general data repositories such as Zenodo (https://zenodo.org/), institutional repositories, or national repositories.

Given the diverse nature of the aforementioned datasets as well as inputs from Work Packages 2 and 3, it was decided to use Zenodo, with a mirror in the institutional repository of Universidad Autónoma de Madrid (UAM). Other mirrors may be added later on, chosen from the list included in the Annex. This decision applies to all the datasets. Using Zenodo will help manage all datasets provided by different work packages and/or consortium members in one platform and let researchers have a holistic view of project results. Besides, the possibility to mirror and store certain data sets on local institutional repositories will help each consortium member to analyze data related to its regional interest.

Datasets preserved on Zenodo will be given a Digital Object Identifier (DOI) as versioned records. A new DOI will be generated for all datasets at the monthly updates, however all datasets will have a Concept (all versions) DOI as well. File names will be generated as follows:

- rurallure.01.2022.01 → DS1 for Jan 2022
- rurallure.02.2022.07 → DS2 for July 2022
- rurallure.03.2022.12 → DS3 for Dec 2022
- rurallure.xx.0001 → For DS4, DS5 and any further dataset that might be published will get suffixes generated as consecutive numbers for each new round of data.

4. Practical dispositions

As explained in the WP9 deliverables, the project will take every possible measure to avoid statistical disclosure of personal data. For this reason, mostly aggregated data will be published, and raw data will be stored in the rurAllure server available only to rurAllure admins during the project's lifetime. Long-term preservation of published aggregated data will be provided by Zenodo, and the aforementioned institutional repositories.

The uploads of data to Zenodo and the automated mirrors will be managed by one module of the rurAllure platform called the "FAIR data research service" (see Fig. 1, retrieved from Deliverable 3.1).

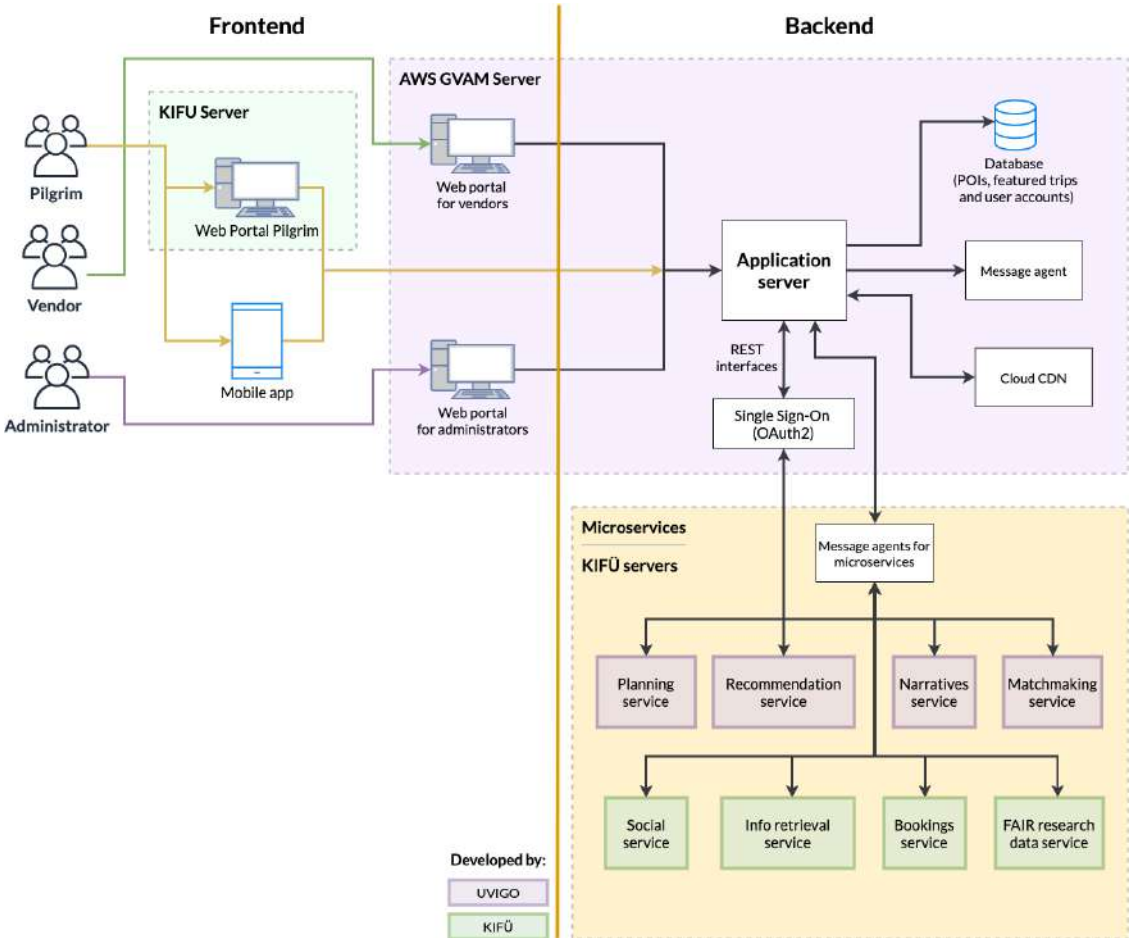

Figure 1. Architectural overview of the rurAllure IT platform.

The FAIR research data service will be responsible for retrieving information from the other platform modules in order to periodically generate data that can be made available to the

scientific community, in the shape of open datasets that are findable, accessible, interoperable, and reusable. It will also serve research data created by other tasks in rurAllure.

The FAIR research data service will be implemented as a set of scripts with a convenient interface that will periodically fetch information, process it and store the periodic snapshots in Zenodo and the automated mirrors. Its interface to the outside world will allow querying the stored information, through the following basic operations:

- Get a list of dataset IDs, names, and descriptions.
- For a given dataset, get the list of snapshots available (identified by their DOIs) and their dates.
- For a given dataset, get the links to the external repositories that keep copies of it.
- Download the whole information in a given dataset (paginated).
- Download the information of a given snapshot of a dataset (paginated).

The platform will not offer frontend tools to perform CRUD operations (create, read, update, and delete) on the internal databases of the FAIR research data service. Likewise, there will be no interface endpoints to add, modify, or remove entries in the lists of external repositories that keep copies of the datasets.

The FAIR Data service will be operated by the appointed data clerks: Mohacsi János and Lencsés Ákos (KIFÜ).

5. Meeting FAIR requirements

The following subsections summarize the key measures to meet the requirements of FAIR research data.

- Making data findable:
  - Datasets will be published with a DOI, data will be described with widely used PIDs (e.g. ISO 639-1, NUTS-3, etc.).
  - File names will reflect on dataset type and time. The system of file names will be easy to learn, still flexible enough to publish newly defined datasets if necessary.
  - Metadata regarding datasets will be published next to each datasets in a separate tab. Metadata will provide data definitions and help to learn how a certain data type has been collected.
  - Keywords will be added to Zenodo records to support discoverability.
- Making data openly accessible:
  - To maximize guarantees with regard to sensitive and personal data, only aggregated datasets will be made openly available.
  - Datasets will be published on Zenodo mostly as csv files (plain text) and will be mirrored on institutional repositories.
- Making data interoperable:
  - Datasets will be published as csv files, no special software is needed for access. All the fields in the csv files is properly described in the metadata description
  - Certain datasets will be fit to link and compare with Eurostat tourism statistics.
  - PIDs (eg. NUT-3 codes) will also help to link rurAllure data with other datasets.
- Increasing data re-use:
  - Datasets will be published under CC-BY license with no embargo period.
  - Monthly datasets will be published at the earliest convenient date, usually every 15th of the following month.

- Long term preservation will be provided via Zenodo and mirrors in institutional repositories.
- Data clerks will be appointed for all datasets.

6. Structure of the datasets

The following subsections contain detailed information about the datasets DS1 to DS5. This information is subject to changes during the project's lifetime, which will be informed in successive versions of this Data Management Plan.

DS1: Collections of POI info sheets

DS1 will contain POI info sheets excerpted from the database of the AWS GVAM Server (Figure 1), containing at least the fields listed in Table 1.

Table 1. Data fields contained in the POI info sheets of DS1.

FIELD	DESCRIPTION
POI ID	Generated by the IT platform
Language	ISO 639-1 language code
Official name of the site	Open text
Description	Open text, maximum length 700 characters
Location	GPS coordinates
Address	Number, road, town, province, country
Distance from official path	Distance in kilometers
Does the POI represent an activity that ends at a different location?	YES/NO
Position for end of activity	GPS coordinates
Duration of activity	Duration in minutes
Categories	Tags from closed list: Museum, restaurant, hotel, church, …
Activity features	Tags from closed list: Art, monuments, gastronomy, outdoors activities, horse riding, culture, music, architecture, religion, relax, shopping, …
Stamping point for pilgrim's credential	YES/NO
Planner priority	1-5

Rating	1-5
Suitable for	Tags from closed list: Individuals, couples, families, groups, …
Recommended transport	Tags from closed list: Car, walk, bicycle, public transport, …
Opening season	Calendar periods
Opening hours	Hour ranges linked to the calendar periods of "Opening season"
Page views (previous month)	Page view count
Views of attached multimedia contents	View counts
Uploaded by whom	Administrator / Vendor

The intended target group and scientific impact can be summarized into the following points:

- DS1 might be of use for scientific research regarding pilgrim route infrastructure.
- DS1 might help regional development planning, and help to identify main characteristics of certain pilgrim route segments. It will also help planning development programs for areas nearby pilgrim routes that are facing economic challenges.
- DS1 will help marketing purposes regarding pilgrim-related POIs, both for research and financial interests.
- Page view counts will help to focus on pilgrims' interests.

DS2: Aggregations of user statistics

DS2 will provide aggregated data from the user profiles kept in the database of the AWS GVAM Server (Figure 1), containing information on pilgrims' origins, demographics, number of trips, traveling patterns (e.g. number of consecutive walking days, length of stays, …), numbers of inbound trips, number of tourist arrivals (domestic and international), etc. The format will reflect on **tour_occ_ninat**, **tour_occ_ninraw**, **tour_dem_toage** and other Eurostat data sheets. The data will be broken down by NUTS3 regions.

The intended target group and scientific impact can be summarized into the following points:

- DS2 might be of use for scientific research regarding tourism.
- DS2 might help regional development planning, and help to identify main characteristics of certain tourist attractions. It will also help planning development programs for areas nearby pilgrim routes that are facing economic challenges.
- Researchers might compare DS2 data with Eurostat tourism statistics. This would help to gain insight into pilgrim tourism characteristics and thereby develop pilgrim-related services in the vicinity of pilgrimage routes.
- DS2 will help to study seasonality of pilgrim tourism in Europe.

DS3: Collections of pilgrimage plans

DS3 will provide collections of completed pilgrimage plans, with indications of whether the constituting POIs were provided by the planning service (shown in the "Microservices" box of Figure 1), or added by accepting a suggestion from the recommendation service (likewise), or added after the pilgrims browsed POI lists on their own.

The structure of DS3 will be finalized once the IT platform has been made available to pilgrims and tourists, past M12. This Data Management Plan will be updated accordingly. The list of Table 2 includes some of the data fields foreseen.

Table 2. Data fields foreseen for DS3.

Percentage of pilgrimage plans created from scratch, adopted completely from a featured plan or adapted (modified) by user groups
Percentage of pilgrims who create a plan and follow it, and those who do not
Modifications of pilgrimage plans after initial versions: type of modification (new stops, activities, detours, companions, number of days, …), trigger (recommendation service, manual browsing, social recommendation, …), time (during trip preparation or realization, hours or days in advance, …)
Success rate of POI suggestions received from the recommendation service or the social service
Preferred POIs in the pilgrimages plans within and in the vicinity of a particular route
Characterization of detours to reach POIs out of the official paths (duration, distance, mode of transport, …)
Aggregated data related to pilgrimage plan ownership and sharing
Usage rate of the matchmaking service and the social service
Data related to interest in narratives preselection and consumption
Aggregated data on usage of the booking options (hospitality, programme)

The intended target group and scientific impact can be summarized into the following points:

- DS3 might be of use for scientific research on pilgrim behavior regarding decision making before and during trips.
- DS3 might give information on the willingness of pilgrims to use social media while planning their trips.
- DS3 may also provide information on the preference of pilgrims in various routes for particular types of POIs.

DS4: Collections of answers to profiling questionnaires

The project will develop various questionnaires addressing pilgrims before their travel (in the trip planning phase) and during their trips. The questions look at socio-demographic information, general trip characteristics (e.g. means and length of travel), pilgrim's interests,

reasons and motivations, flexibility regarding dates, amenability to explore the rural surroundings of the official paths, etc. Some questions are related to the effect of COVID-19 on trip planning and realization.

The structure of DS4 will be finalized once the first round of questionnaire answers have been gathered. This Data Management Plan will be updated accordingly. The list of Table 3 includes some of the data fields foreseen.

Table 3. Data fields foreseen for DS4.

Aggregated data on travel motivations
Aggregated data on travel flexibility
Aggregated data of typical trip characteristics by pilgrim profiles
Aggregated data on the effect of COVID on travel
Aggregated data of socio-demographic characteristics of pilgrims and platform users

The intended target group and scientific impact can be summarized into the following points:

- DS4 may be of use for social science research exploring and comparing travel motivations and habits of pilgrims and other travelers.
- DS4 might give information on the effect of COVID-19 on the choice of travel and how it may affect the turnover at pilgrimage routes in the future.
- DS4 may also provide information on the indirect effects of pilgrimages on rural development if detours from the main route can be achieved.

DS5: Lists of relevant policies and regulations

The project is gathering lists of laws and policies, classified by NUTS levels. DS5 will make the data available to the research community, which will be coupled with Deliverable 2.8 ("Policy brief", scheduled on M12) and D2.5 ("White book of recommendations", M30). The structure of the dataset is summarized in Table 4.

Table 4. Data fields foreseen for DS5.

	FIELD	DESCRIPTION
Pilgrimage route		Name of the route
Trail stretch		Initial and final locations
Country		ISO 3166-1 country code
NUTS level 1	Cultural heritage policymaker	Open text
	Cultural heritage policies	List of URLs
	Tourism policymaker	Open text

	Tourism policies	List of URLs
	Rural development policymaker	Open text
	Rural development policies	List of URLs
	Pilgrimage/hiking trails policies	List of URLs
NUTS level 2	Cultural heritage policymaker	Open text
	Cultural heritage policies	List of URLs
	Tourism policymaker	Open text
	Tourism policies	List of URLs
	Rural development policymaker	Open text
	Rural development policies	List of URLs
	Pilgrimage/hiking trails policies	List of URLs
NUTS level 3	Cultural heritage policymaker	Open text
	Cultural heritage policies	List of URLs
	Tourism policymaker	Open text
	Tourism policies	List of URLs
	Rural development policymaker	Open text
	Rural development policies	List of URLs
	Pilgrimage/hiking trails policies	List of URLs

Annex: Data repositories

Name	Website	Institution	Open access	Notes
		EU		
New European Bauhaus - papers and essays	https://ec.europa.eu/eusurvey/runner/new_european_bauhaus_freestyle_contributions	European Commission	Yes	New European Bauhaus platform to share papers, essays and studies
Open Research Europe	https://open-research-europe.ec.europa.eu/	European Commission	Yes	Publishing platform for the research stemming from Horizon 2020
		HUNGARY		
Concorda	https://www.sztaki.hu/en/innovation/projects/concorda	Institute for Computer Science and Control	Partially	General repository for research data.
DKA	http://dka.oszk.hu/indexeng.phtml	National Széchényi Library	Yes	General repository for images.
REAL	http://real.mtak.hu/	Library and Information Centre of the Hungarian Academy of Sciences	Partially	General repository for scientific literature.
		INTERNATIONAL		
Scientific Electronic Library Online (SciELO)	https://www.scielo.org/	FAPESP; CNPq; BIREME	Yes	Bibliographic database, digital library and cooperative electronic publishing model
Zenodo	https://zenodo.org/	CERN; OpenAIRE; European Commission	Yes	International repository for scientific publications

ViaTourism Review - International Interdisciplinary Review of Tourism	https://journals.openedition.org/viatourism/		Yes	International, multilingual and interdisciplinary online journal on tourism
		ITALY		
Alma Digital Library	https://www.unibo.it/it/servizi-e-opportunita/servizi-online/servizi-online-per-studenti-1/guida-servizi-online-studenti/alma-digital-library	UNIBO	Yes	UNIBO institutional archive
Almatourism	https://almatourism.unibo.it/	UNIBO	Yes	Electronic journal
Bollettino della Società Geografica italiana (BSGI)	https://societageografica.net/wp/2021/03/17/bollettino/ https://riviste.fupress.net/index.php/bsgi/index	Società Geografica Italiana; Firenze University Press	Yes	Online and printed
Bononia University Press (BUP)	https://buponline.com/	UNIBO	Partially	UNIBO Publishing house
Collana Koiné del Dipartimento di Scienze per la qualità della vita		UNIBO		Series of the Department of Sciences for the quality of life
Memorie della Società Geografica italiana + Ricerche e Studi della SGI	https://societageografica.net/wp/2021/03/18/memorie-della-sgi/ https://societageografica.net/wp/2021/03/18/ricerche-e-studi-della-sgi/	Società Geografica Italiana	No	Scientific monographs
Padua Research Archive	https://www.research.unipd.it/	UNIPD	Yes	Institutional archive of the scientific production of UNIPD
		NORWAY		
Current Research Information	https://www.cristin.no/english/	Royal Ministry of	Yes	National research

System in Norway (CRIStin)		Education and Research; UNIT		information system of Norway
Norwegian Open Research Archives (NORA)	https://nora.openaccess.no/	CRIStin	Yes	Both a search engine and a harvester of approx. 70 Norwegian institutional repositories
NTNU OPEN	https://ntnuopen.ntnu.no/ntnu-xmlui/	NTNU	Yes	Institutional repository of the Norwegian University of Science and Technology
PORTUGAL				
Portugal open access scientific repository (RCAAP)	https://www.rcaap.pt/		Yes	Scientific contents from Portuguese institutional repositories
SciELO Portugal	http://www.scielo.mec.pt/	FAPESP; CNPq; BIREME; FCT	Yes	Electronic library
ROMANIA				
Romanian Journal of Information Science and Technology (ROMJIST)	https://www.romjist.ro/index.html	Romanian Academy	Yes	Online and printed
Romanian Journal of Regional Science (RJRS)	http://www.rjrs.ase.ro/	Romanian Regional Science Association	Yes	Online journal
SLOVAKIA				
FiF CU	https://uniba.sk/en/	Univerzita Komenského	Yes	Faculty website
FIIT STU (Fakulta	https://www.fiit.stuba.sk/	STUBA	Yes	Faculty website

informatiky a informačných technológií - Slovenská technická univerzita)				
Spektrum	https://spektrum.stuba.sk/	STUBA	Yes	University magazine
STUBA (Slovenská technická univerzita v Bratislave)	https://www.stuba.sk/	STUBA	Yes	University website
	SPAIN			
Biblos-e Archivo	https://repositorio.uam.es/	UAM	Yes	Institutional repository of the Universidad Autónoma de Madrid
e-cienciaDatos	https://edatos.consorciomadrono.es/	Consorcio Madroño	Yes	Consorcio Madroño data repository
Investigo	http://www.investigo.biblioteca.uvigo.es/xmlui?locale-attribute=es	UVIGO	Yes	Institutional repository of the Universidade de Vigo
Recolecta	https://www.recolecta.fecyt.es/portada?language=es	FECYT	Yes	National aggregator of open access repositories
RUC	https://ruc.udc.es/dspace/	UDC	Yes	Institutional repository of Universidade da Coruña

reach out!

www.rurallure.eu