RDA Global Adoption week

15 - 19 June 2020

CC BY SA

Objective

- Originally planned for Plenary 15, the RDA Adoption week aims to **demonstrate the wide variety of RDA adoptable and adopted solutions to data sharing challenges** that people in the field encounter in their daily jobs.

Purpose of the week:

- Learn about RDA Outputs
- Converse with speakers from all around the world who have created and implemented them
- Determine how best to integrate those data sharing solutions into your own projects

CC BY SA

Wednesday 17th June

14:00 UTC

Identity, Store & Preserve

Collections and streams of digital objects are growing at an incredibly rapid pace. We need to understand on what these objects are, how we need to be documenting and storing them, and how they should all link and talk to one another before we get overtaken by a forest of data.

- **Scalable Dynamic Data Citation Methodology:** the CCCA Subset & Dynamic Data Citation Service - Chris Schubert (CCCA)
- **Social Sciences and Humanities Open Cloud (SSHOC) adoption story of CoreTrustSeal across European SSH repositories** - Hervé L'Hours (UK Data Archive - CESSDA)
- **39 Hints to Facilitate the Use of Semantics for Data on Agriculture the Nutrition:** the Food and Agriculture Organization (FAO) adoption story - Caterina Caracciolo (FAO)

Recommendations & Outputs Catalogue short-link: **https://bit.ly/378QOBq**

RESEARCH DATA ALLIANCE

Identity, Store & Preserve

Collections and streams of digital objects are growing at an incredibly rapid pace. We need to understand on what these objects are, how we need to be documenting and storing them, and how they should all link and talk to one another before we get overtaken by a forest of data.

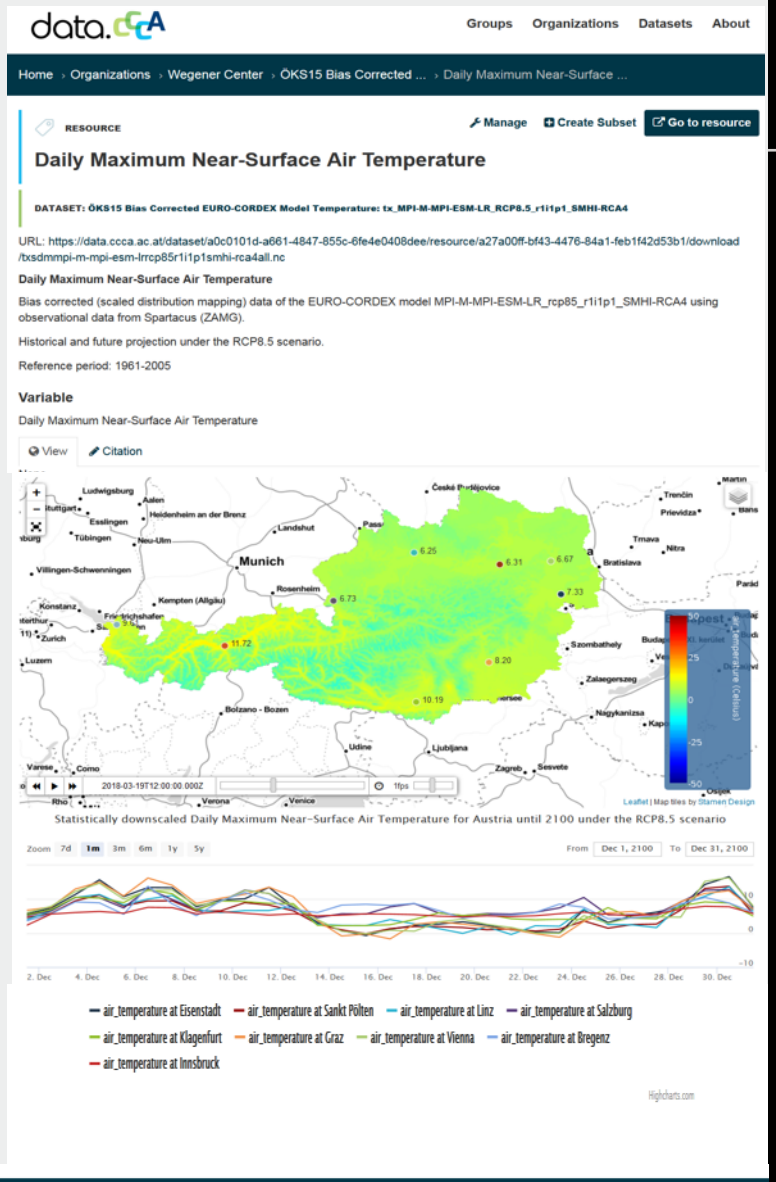
data.ccca

SCALABLE DYNAMIC DATA CITATION METHODOLOGY

... a short view on RDA adoption

Chris Schubert
Head of CCCA – Data Centre
data.ccca.ac.at
1190 Vienna, Austria

chris.schubert[at]ccca.ac.at



CONNECT

CCCA Data Centre, an Austrian research data infrastructure, promotes interoperability and collaboration between different science and research communities

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Forschungsinfrastruktur

http://doi.org/10.17616/R3KS9D

CCCA Data Centre

SERVICES

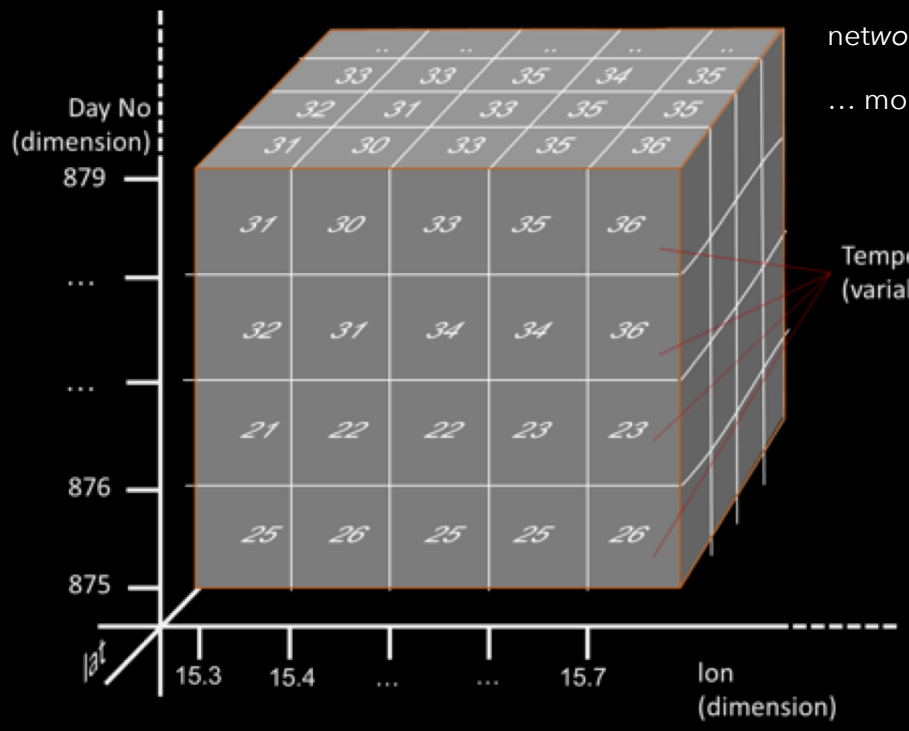
Publish and cite resources & data

Centralized access to relevant meta-information

Storage, Server, VM & HPC facilities

On the fly preview of NetCDF files

Create subsets of large NetCDF files



network *Common Data Form*

… more than a data format

RESPONSIBLE FOR A BETTER

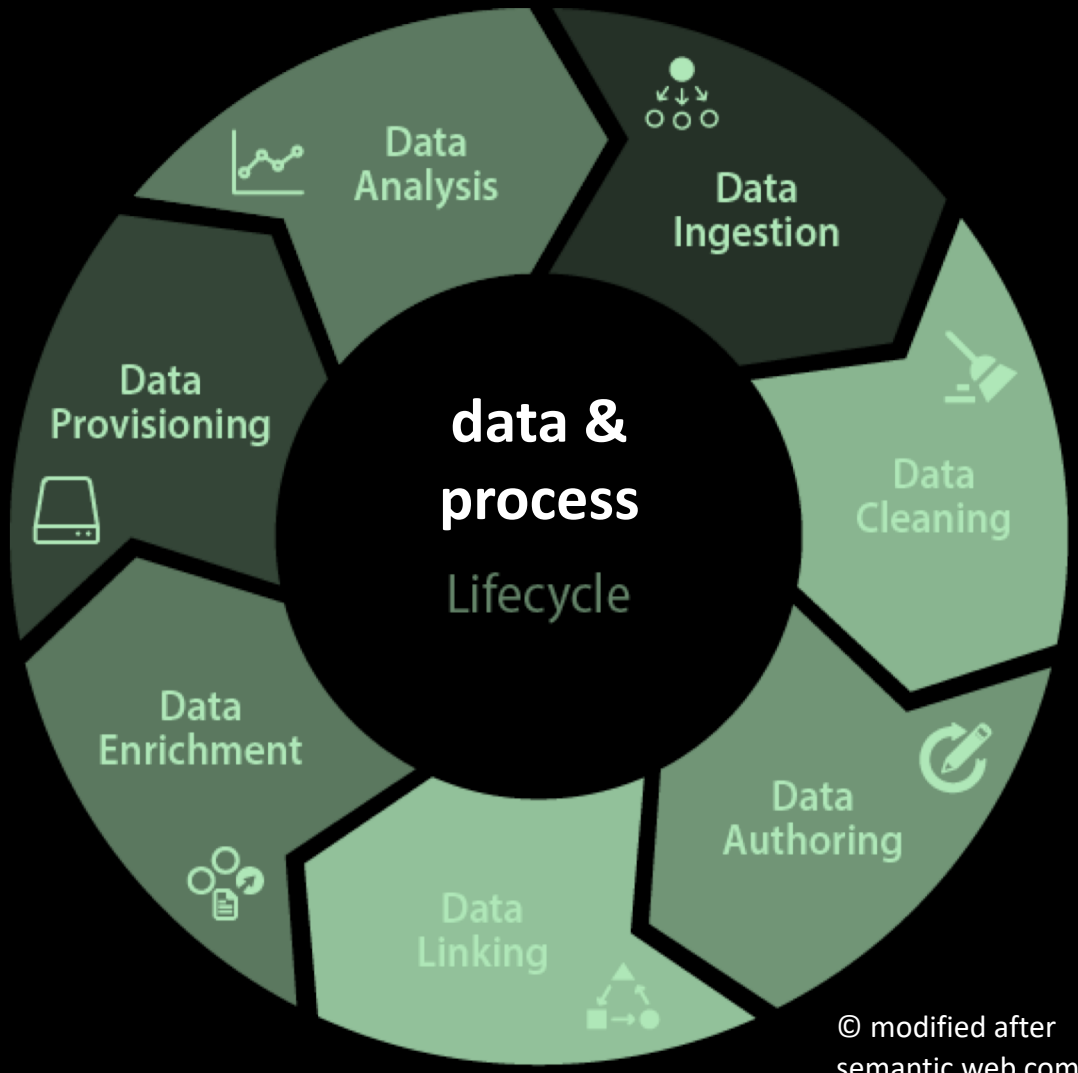
Data Access & Reuse

Data Preservation

Data Processing and analysis

Domain tailored Data Management

Data Life Cycle, Data Provenance



© modified after semantic web company poolparty

DCAT-AP
FOR
DATA PORTALS
IN EUROPE

DCAT application profile for data portals in Europe

Dataset Metadata — Export Metadata -

Contact | Basics | Keywords | Spatial | Time | Specifics | Quality | Conformity

Owner and Contact Information regarding this dataset

Organization	Wegener Center
Metadata Point of Contact (Maintainer):	Heimo Truhetz heimo.truhetz@uni-graz.at
Dataset Creator (Author):	Armin Leuprecht armin.leuprecht@uni-graz.at
Citation Info	Leuprecht et al

Dataset Metadata

Contact | Basics | Keywords | Spatial | Time | Specifics | Quality | Conformity

Keywords

Controlled Keywords	bias correction, scaled distribution mapping
Used Thesauri	

Dataset Metadata — Export Metadata -

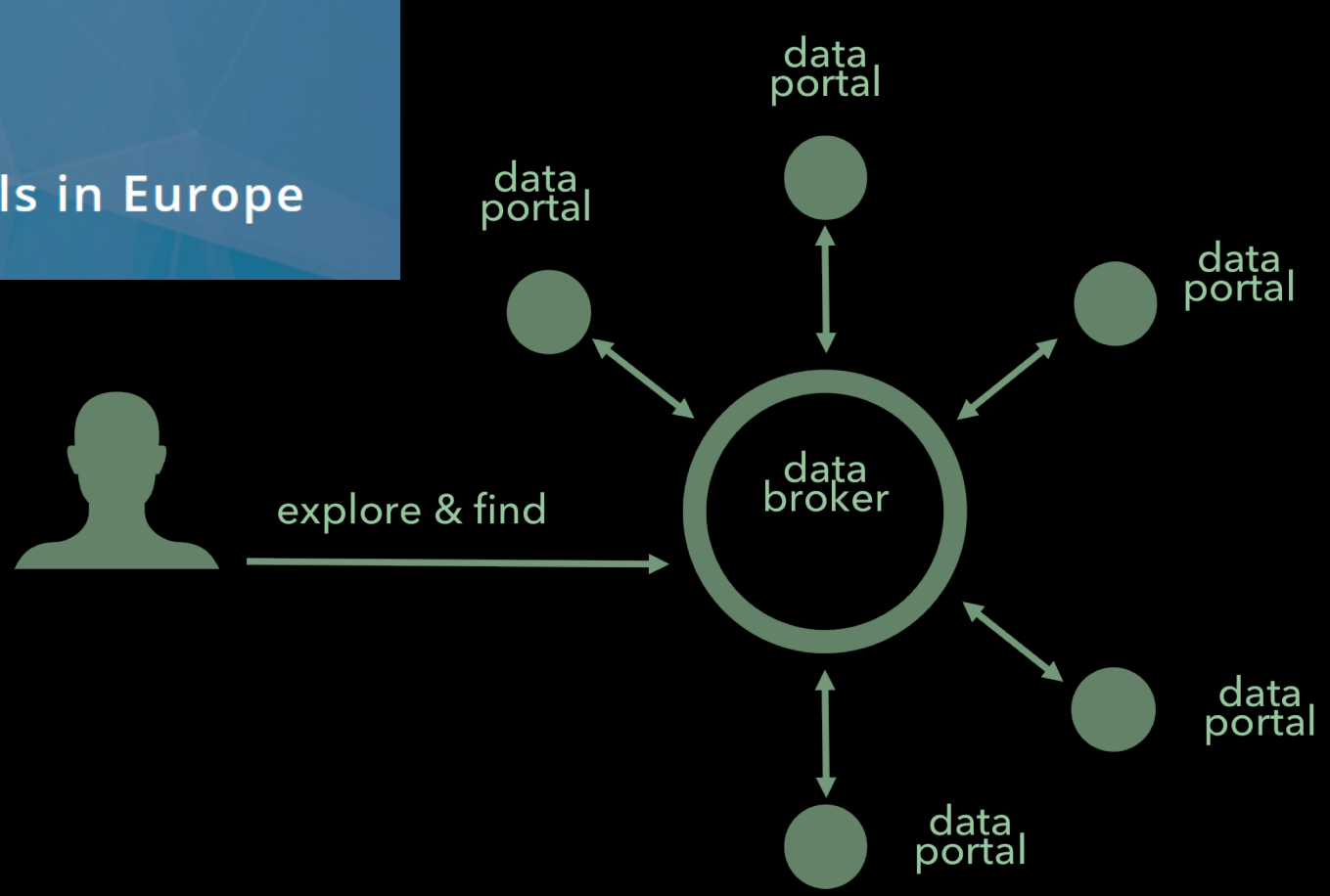
Contact | Basics | Keywords | Spatial | Time | Specifics | Quality | Conformity

Geographic Aspects of the Resources

Polygon

Dataset extent

Coverage Austria

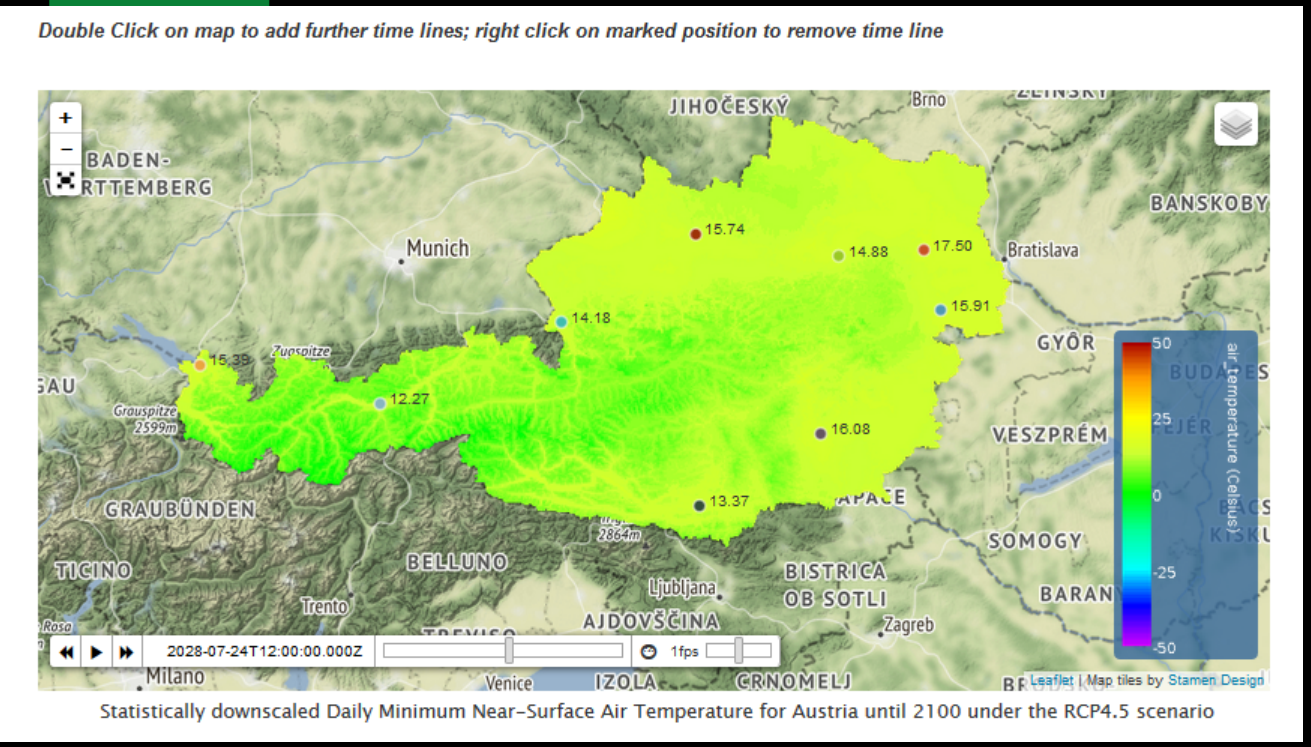


METADATA

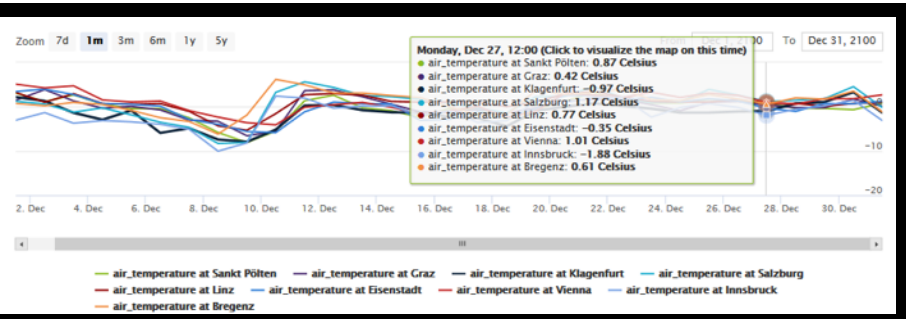
GEO DCAT – AP

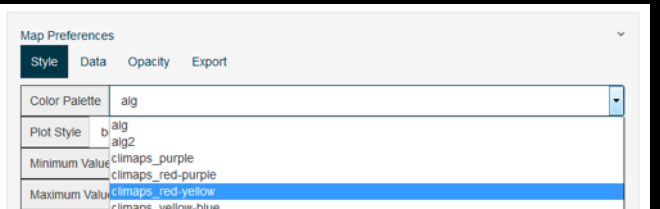
RESEARCH DATA ALLIANCE

data.ccca



Statistically downscaled Daily Minimum Near-Surface Air Temperature for Austria until 2100 under the RCP4.5 scenario





visual context

easier exposed and

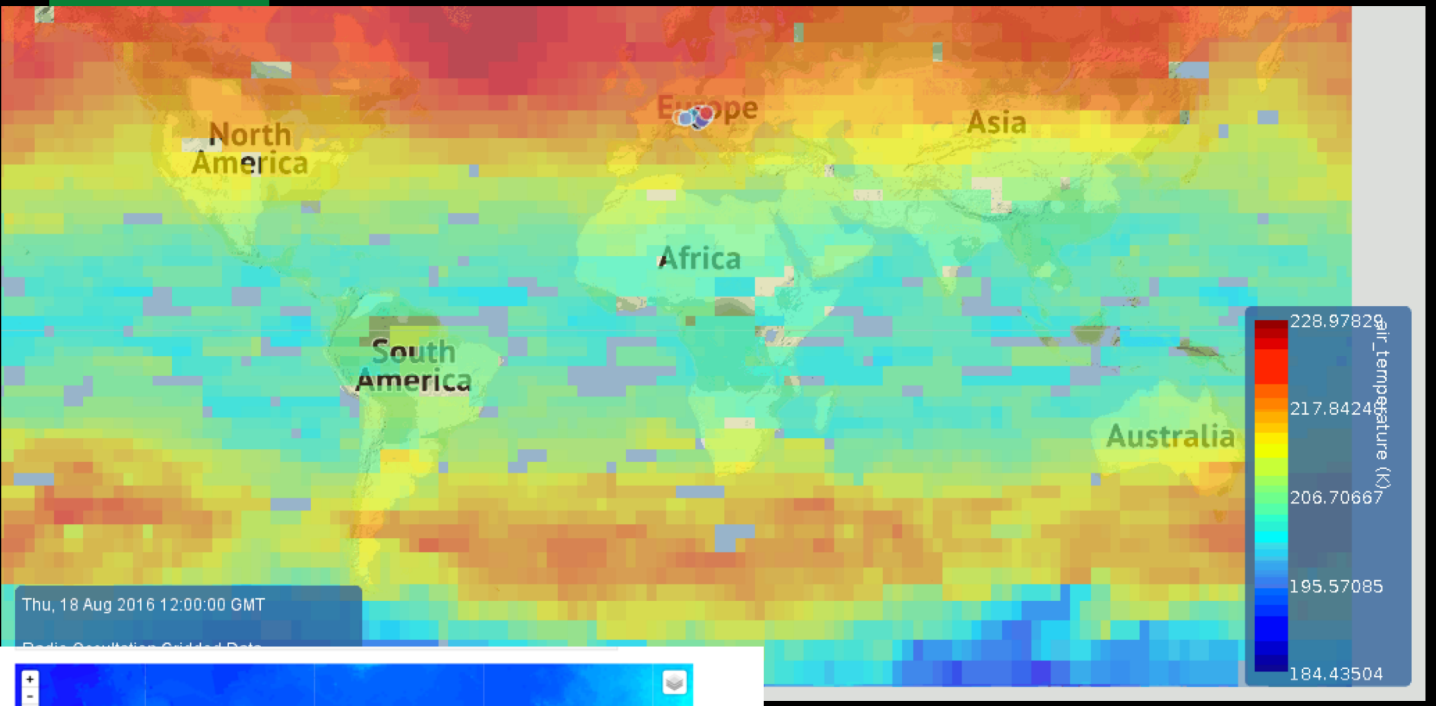
recognized information

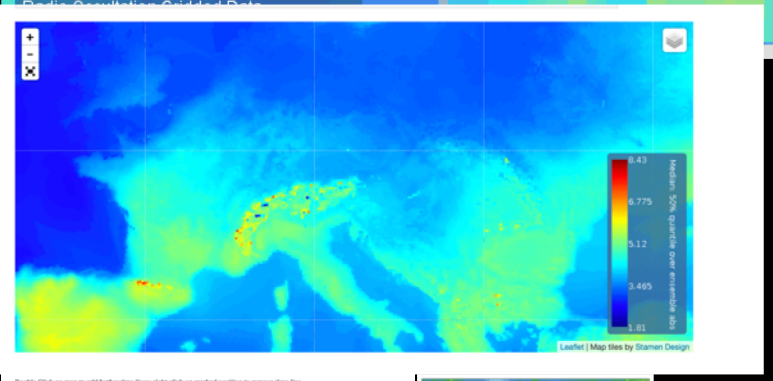
VISUALISATION

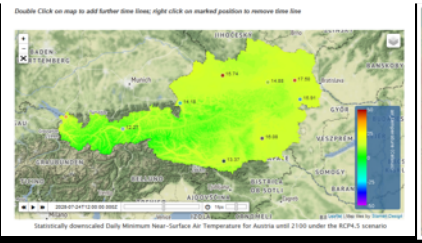
SHOW your data

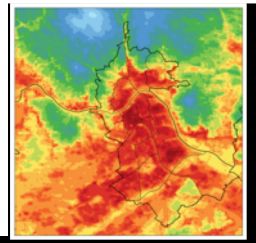
RESEARCH DATA ALLIANCE

data.cccA









providing

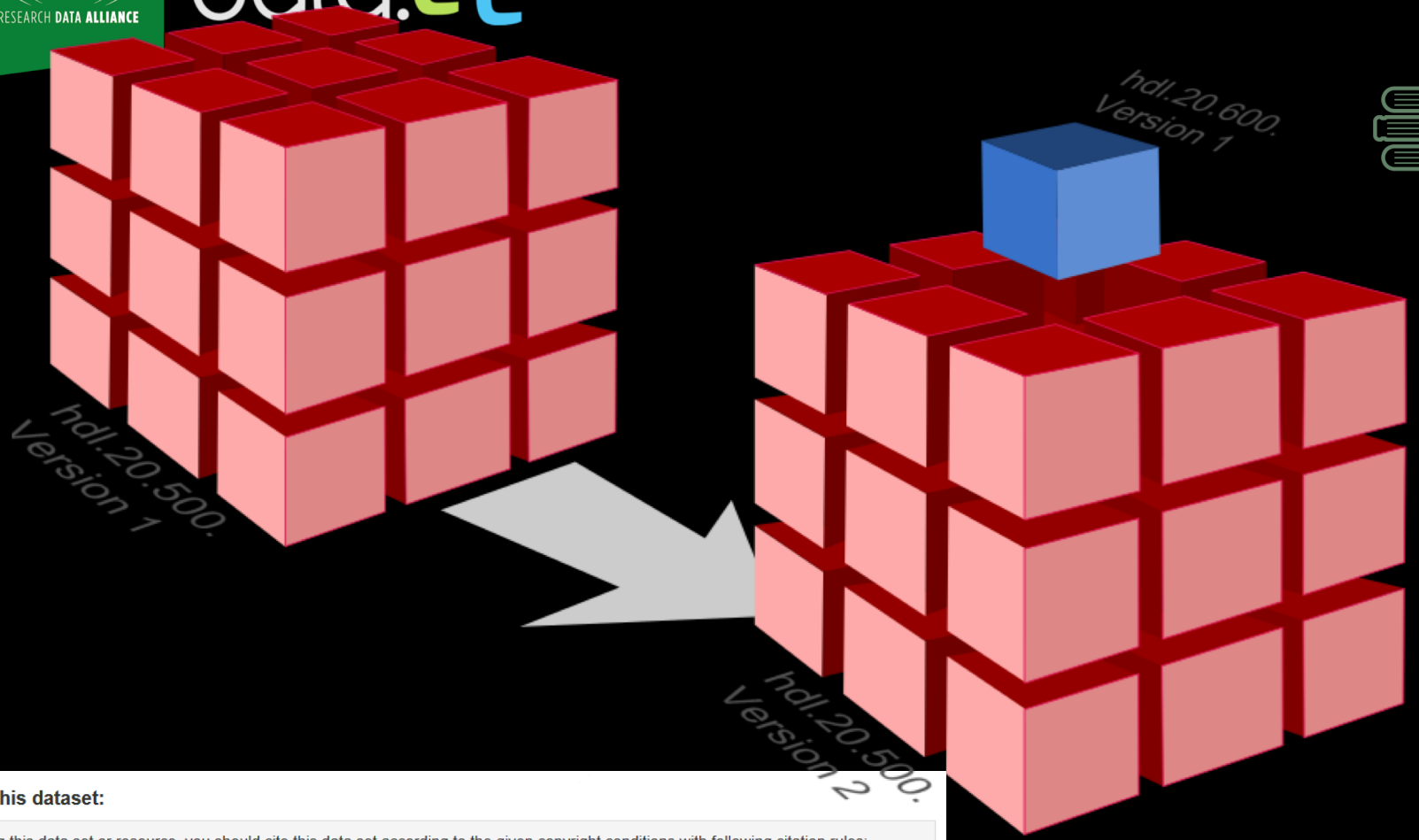
different spatial extent

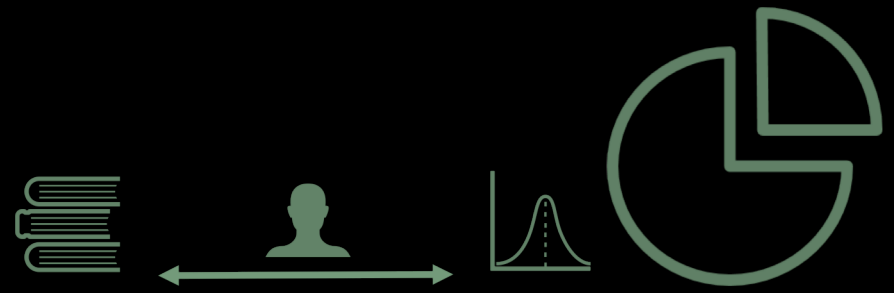
STANDARDISED DATA FORMAT

SHOW your data

RDA RESEARCH DATA ALLIANCE

data.ccca





Citation

Visualisation

Relation

Information

Cite this dataset:

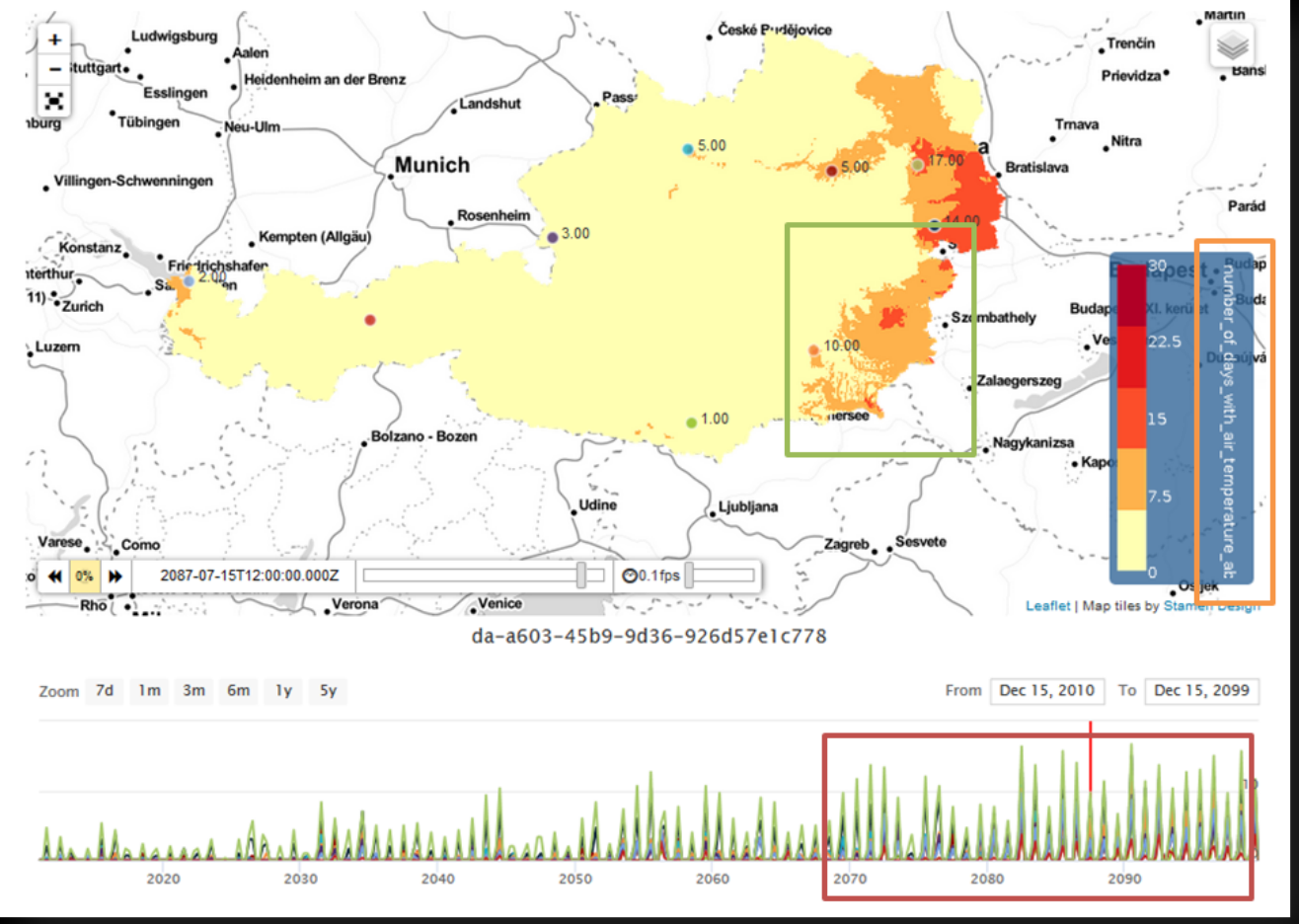
Using this data set or resource, you should cite this data set according to the given copyright conditions with following citation rules:

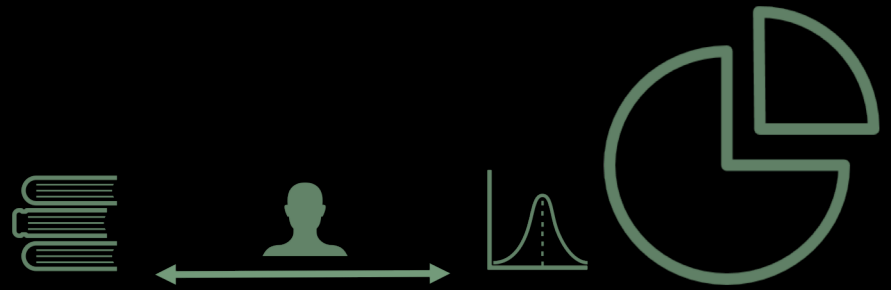
Leuprecht et al (2016). ÖKS15 Bias Corrected EURO-CORDEX Model Precipitaion: pr_CNRM-CERFACS-CNRM-CM5_RCP4.5_r1i1p1_CLMcom-CCLM4-8-17, Version 1. Vienna, Austria. CCCA Data Centre. PID: https://hdl.handle.net/20.500.11756/9df12611. [April 20, 2018]

Copy Text

SUBSETTING + dynamic data citation

Cite your Data





(research) data is dynamic

identify precisely the data at a specific point in time

identify precisely the subset of (dynamic) in a process

Choose a:

PARAMETER

AREA OF INTEREST

TIME RANGE

@KEEP VERSIONING

@KEEP TIMESTAMPS

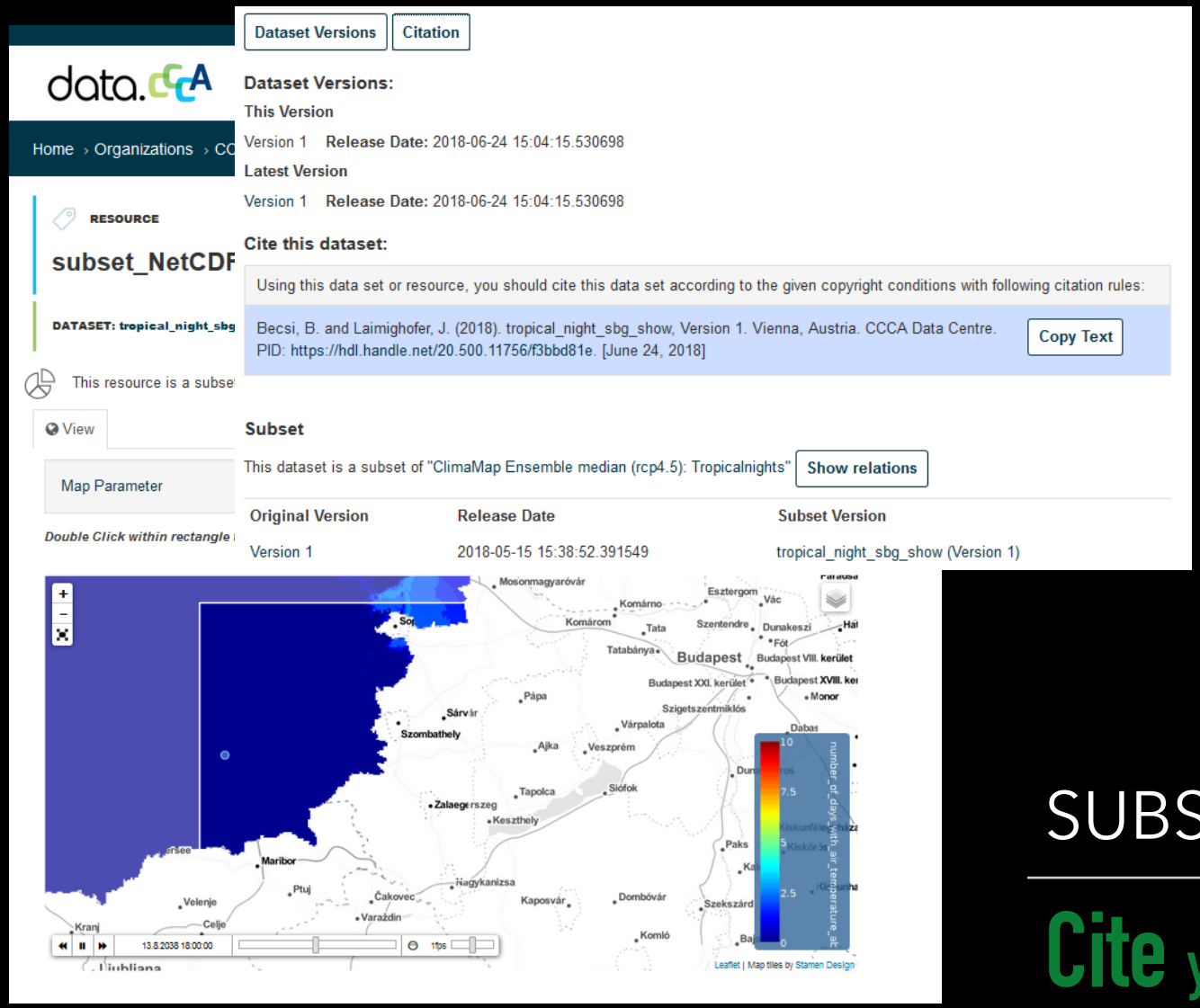
@KEEP & ADAPT METADATA

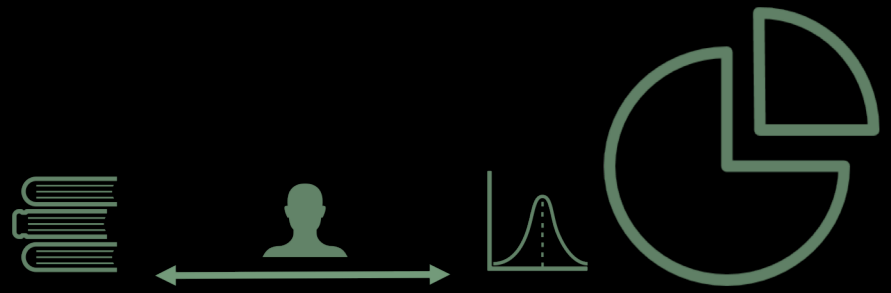
SUBSETTING + dynamic data citation

Cite your Data

RDA RESEARCH DATA ALLIANCE

data.ccca





(research) data is dynamic

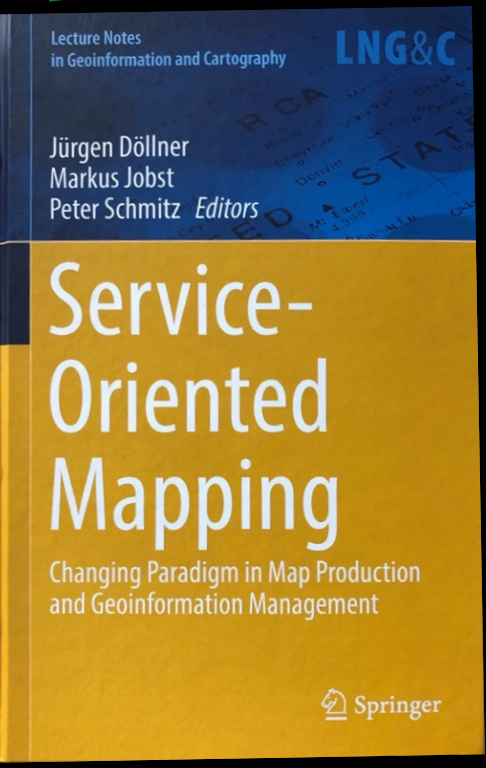
Re-published

avoid redundant storage consumption

keep all relations between updates, original sources & subsets

SUBSETTING + dynamic data citation

Cite your Data



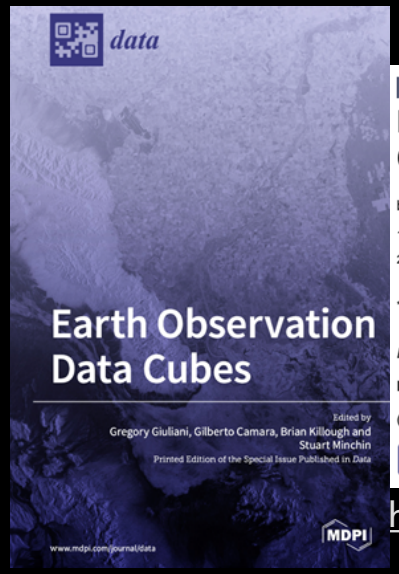
© 2019

Service Oriented Mapping
Changing Paradigm in Map Production
and Geoinformation Management

Handling Continuous Streams for Meteorological Mapping

Chris Schubert[1], Harald Bamberger[2]

[1] CCCA Data Centre, Vienna, Austria, hosted by ZAMG,
[2] ZAMG, Dep. Software Application development and Data Management



Special Issue "Earth Observation Data

Open Access Article

Dynamic Data Citation Service—Subset Tool for Operational Data Management

by Chris Schubert [1,*], Georg Seyerl [1] and Katharina Sack [2]

[1] Data Centre—Climate Change Centre Austria, 1190 Vienna, Austria
[2] Institute for Economic Policy and Industrial Economics, WU—Vienna University of Economics and Business, 1020 Vienna, Austria
* Author to whom correspondence should be addressed.

Data **2019**, *4*(3), 115; https://doi.org/10.3390/data4030115

Received: 31 May 2019 / Revised: 29 July 2019 / Accepted: 30 July 2019 / Published: 1 August 2019

(This article belongs to the Special Issue Earth Observation Data Cubes)

View Full-Text | Download PDF | Browse Figures | Review Reports

https://doi.org/10.3390/data4030115

SUBSETTING + dynamic data citation

PUBLICATION

RESEARCH DATA ALLIANCE

data.ccca

TU WIEN

Dynamic Data Citation for frequently modifying High Resolution Climate Data

RESEARCH DATA ALLIANCE

Climate Change Centre Austria (CCCA) Data Centre adopts Research Data Alliance (RDA) Recommendation on Data Citation of Evolving Data

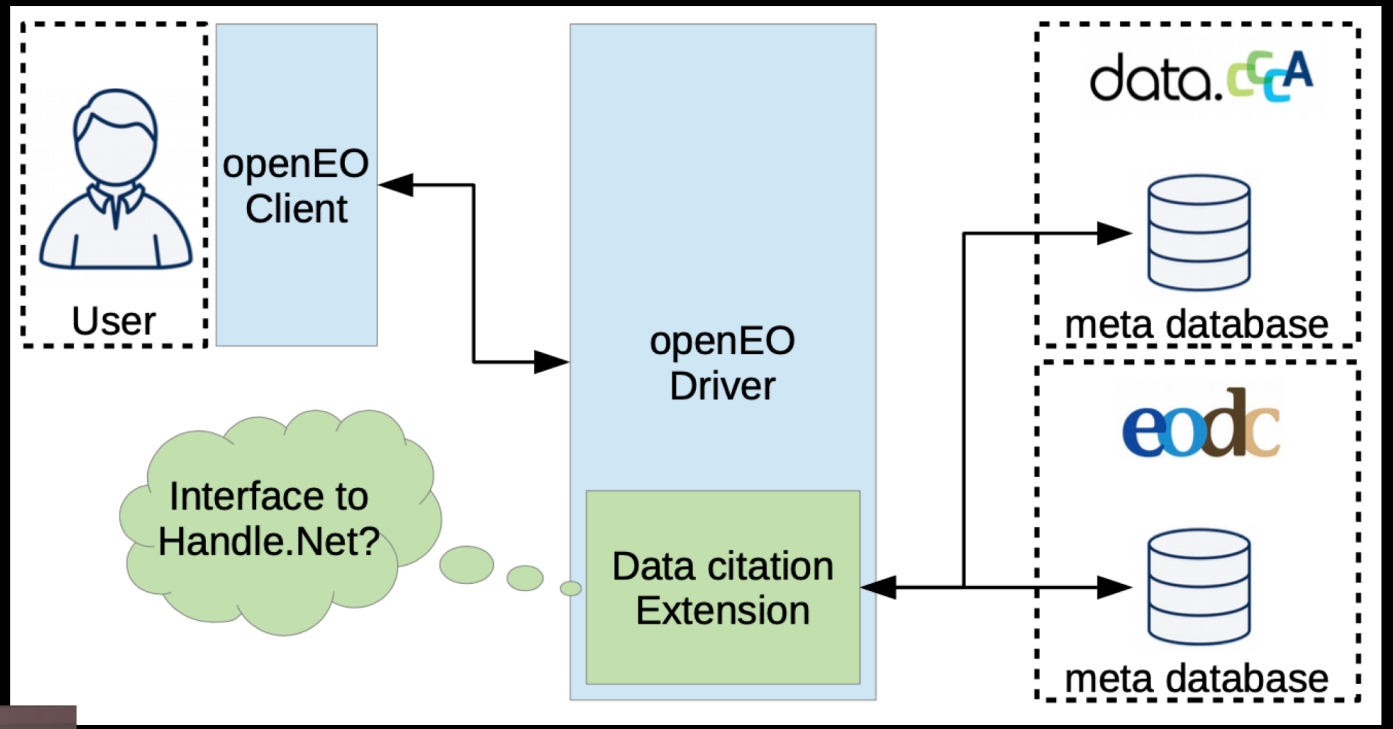
An RDA adoption story written by **Chris Schubert**, Geologist and Geoinformatics,
Head of CCCA – Data Centre, Coordinator for Austria of the Group on Earth Observation (GEO)

Reading time: 6 minutes

The Climate Change Centre Austria (CCCA) Data Centre expected a comprehensive project outcome of completely new simulated High Resolution Climate Scenarios for Austria in the time range from 1965 till 2100 on a daily basis. For consumption, 13 model runs, 5 meteorological parameters like temperature, 3 emission scenarios, over 1600 NetCDF files with an average size of 13 GB were calculated. How could we implement proper data management processes on such data packages? We were looking for best practices on persistent identifiers and sub-setting tools for such big data containers. By chance, I met members of the RDA Data Citation Working Group. The idea of using the RDA recommendation on dynamic data citation as a pilot "NetCDF Pilot Implementation of Climate Scenarios" was born.

"High Resolution Climate Data modify frequently, due to their complex dependencies and statistical methods for downscaling. In order to re-use these data and services in a reproducible manner, to share and cite, data analysts and researchers need a possibility to identify the exact version used."

Chris Schubert, Head of CCCA-Data Centre



SUBSETTING + dynamic data citation

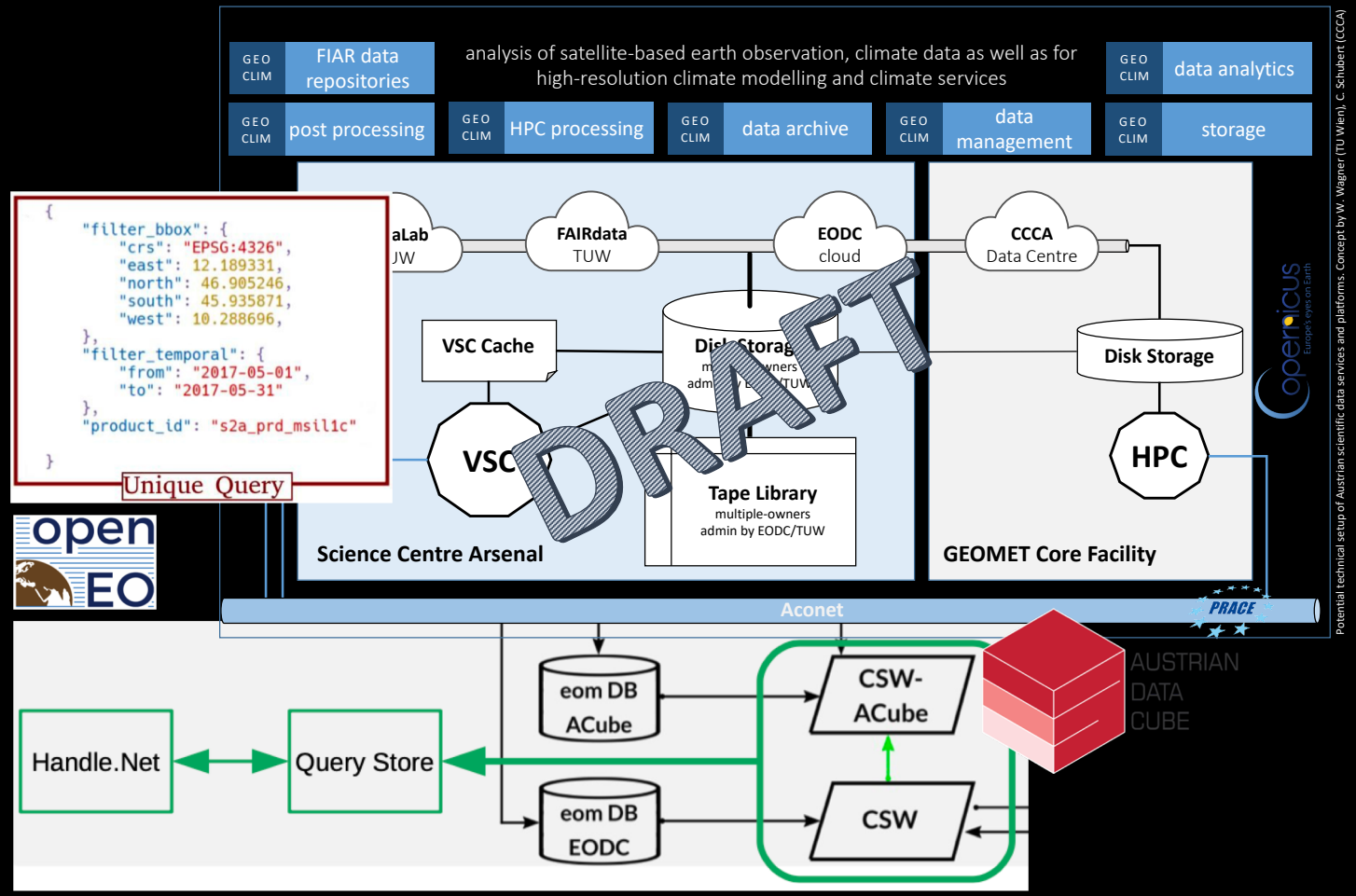
RDA Adoption Grant

RESEARCH DATA ALLIANCE

data.ccca

Design Concepts on data service landscape to establish the research data infrastructure

- CCCA Data Center wants to establish services to the Earth Observation domain w/o collecting redundant data
- Extent the user community
- Improvement of software architecture
- Find best synergies to OpenEO, OGC Standards
- Find best synergies for the OpenDataCube initiatives
- Developed software modules are still Open Source, see EODC - RDA Data Citation Recommendations on an openEO

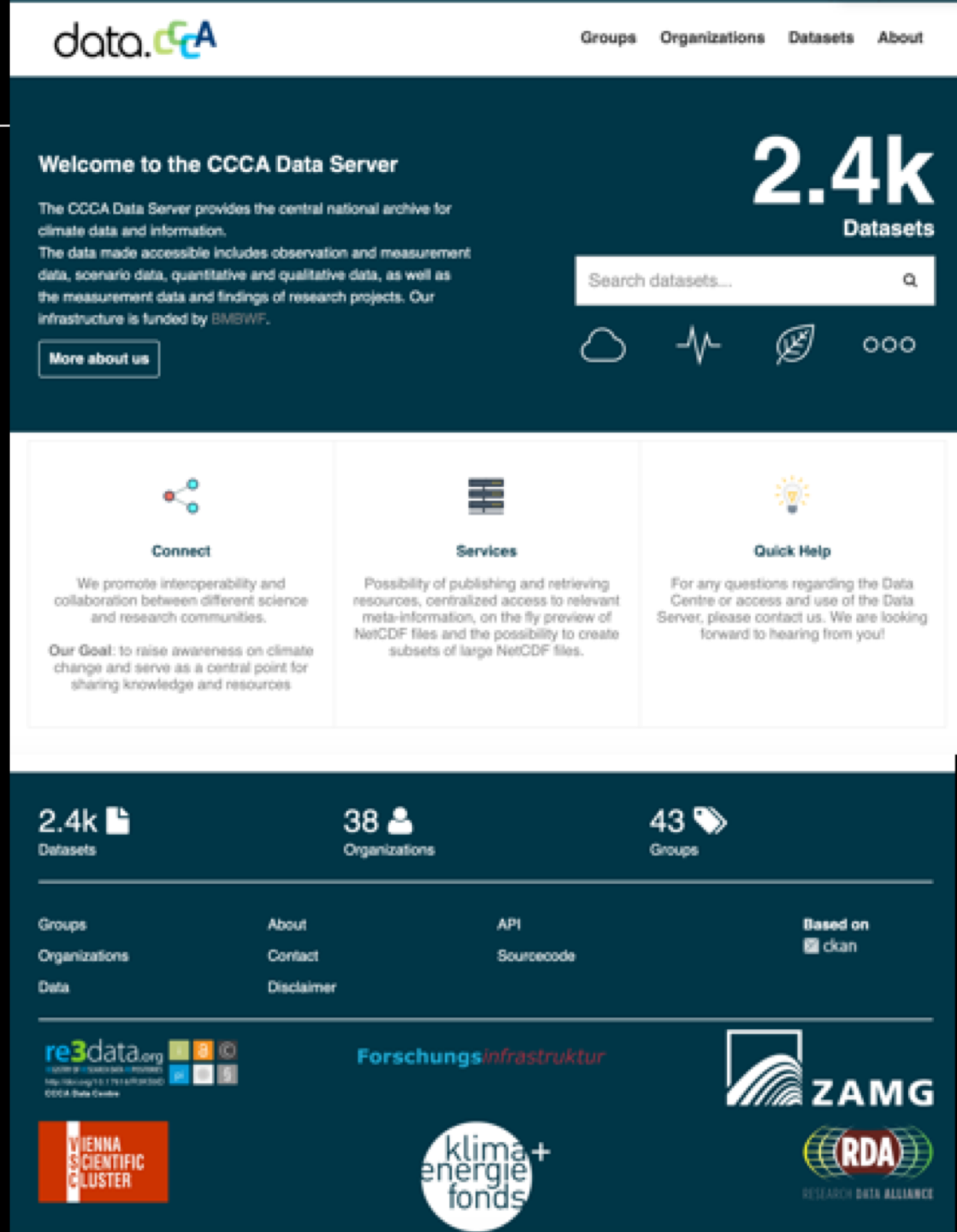


SUBSETTING + dynamic data citation

RDA ADOPTION

Thank you for attention !

Chris Schubert
Head of CCCA – Data Centre
GEO Coordinator for Austria
data.ccca.ac.at
1190 Wien, Hohe Warte 38 Tel: +43136026
2519 chris.schubert[at]ccca.ac.at



CoreTrustSeal in SSHOC
Social Sciences & Humanities Open Cloud

RDA Adoption Stories

Hervé L'Hours, UK Data Service - CESSDA
Franciska de Jong, CLARIN ERIC

RDA Adoption Week: Identify, Store and Preserve
2020-06-20

CC BY

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

Project: SSHOC

social sciences & humanities open cloud

Horizon 2020
European Union Funding
for Research & Innovation

Type of action & funding:

Research and Innovation action
(INFRAEOSC-04-2018)

Partners: 45

(20 beneficiaries + 25 LTPs)

SSH ESFRI Landmarks and Projects
& international SSH data infrastructures

Duration: 40 months

(January 2019 – 30 April 2022)

Project budget:

€ 14,455,594.08

Project website:

www.SSHopencloud.eu

Objectives:

- creating the social sciences and humanities **(SSH)** part of European Open Science Cloud **(EOSC)**
- maximising **re-use** through **Open Science** and **FAIR** principles (standards, common catalogue, access control, semantic techniques, training)
- interconnecting existing and new infrastructures (clustered cloud infrastructure)
- establishing appropriate **governance model** for SSH-EOSC

The Organisation

6 European Research Infrastructures (+partners) from the SSH domain

- Developing
- Established

- CESSDA ERIC
- CLARIN ERIC
- DARIAH ERIC
- ESS ERIC
- SHARE ERIC
- E-RIHS
- LIBER Association of European Research Libraries

40 months to move from disciplinary silos and separate facilities into an integrated, cloud-based network of interconnected data infrastructures.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

The Challenge

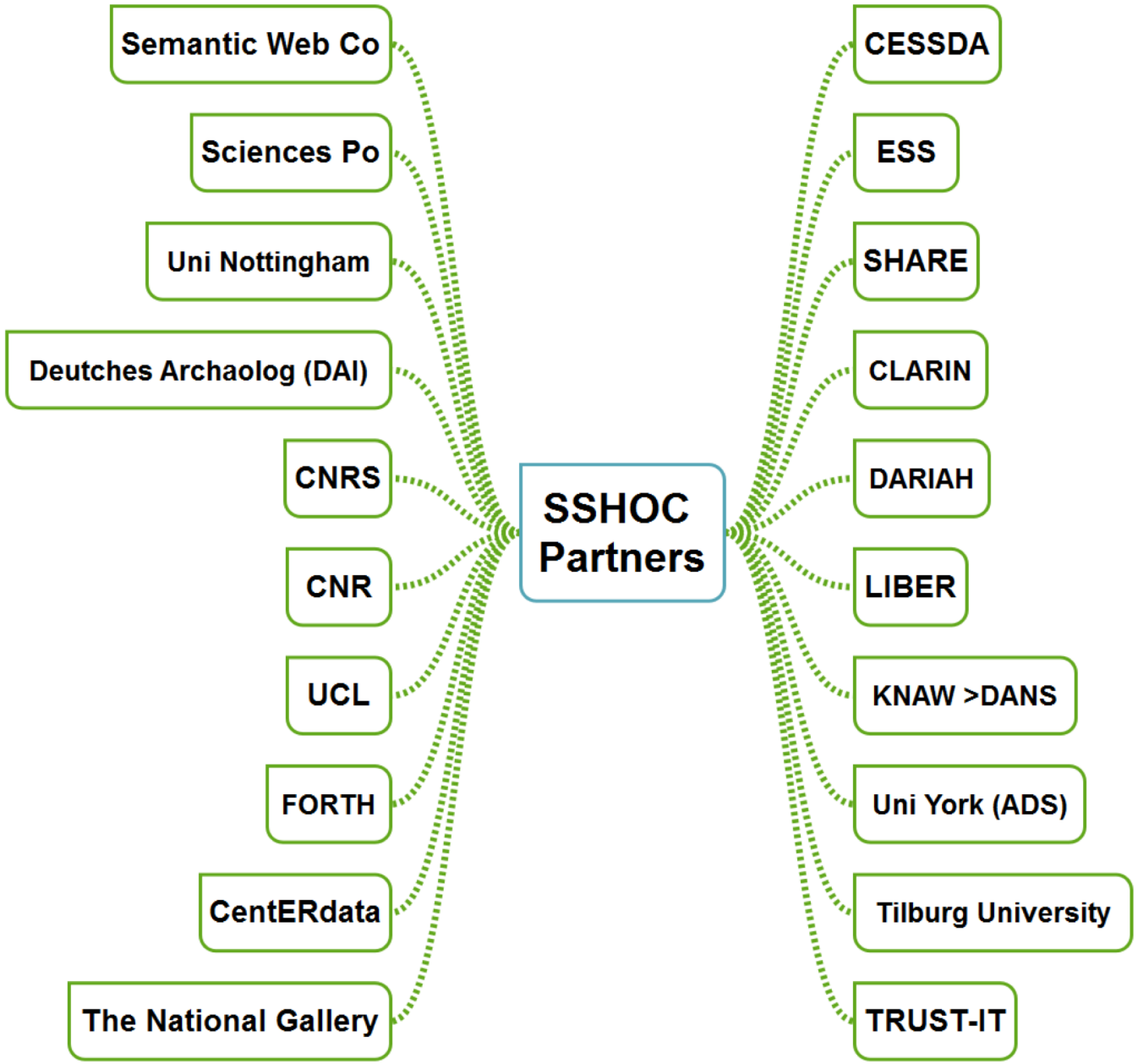
- Research Infrastructures: repository trust & quality
  - CESSDA Social Sciences
  - CLARIN Language data for SSH at large
  - DARIAH Arts & Humanities
  - E-RIHS Cultural & Natural Heritage

- A common approach to assessing:
  - People
  - Processes
  - Technology

Evaluation against agreed requirements

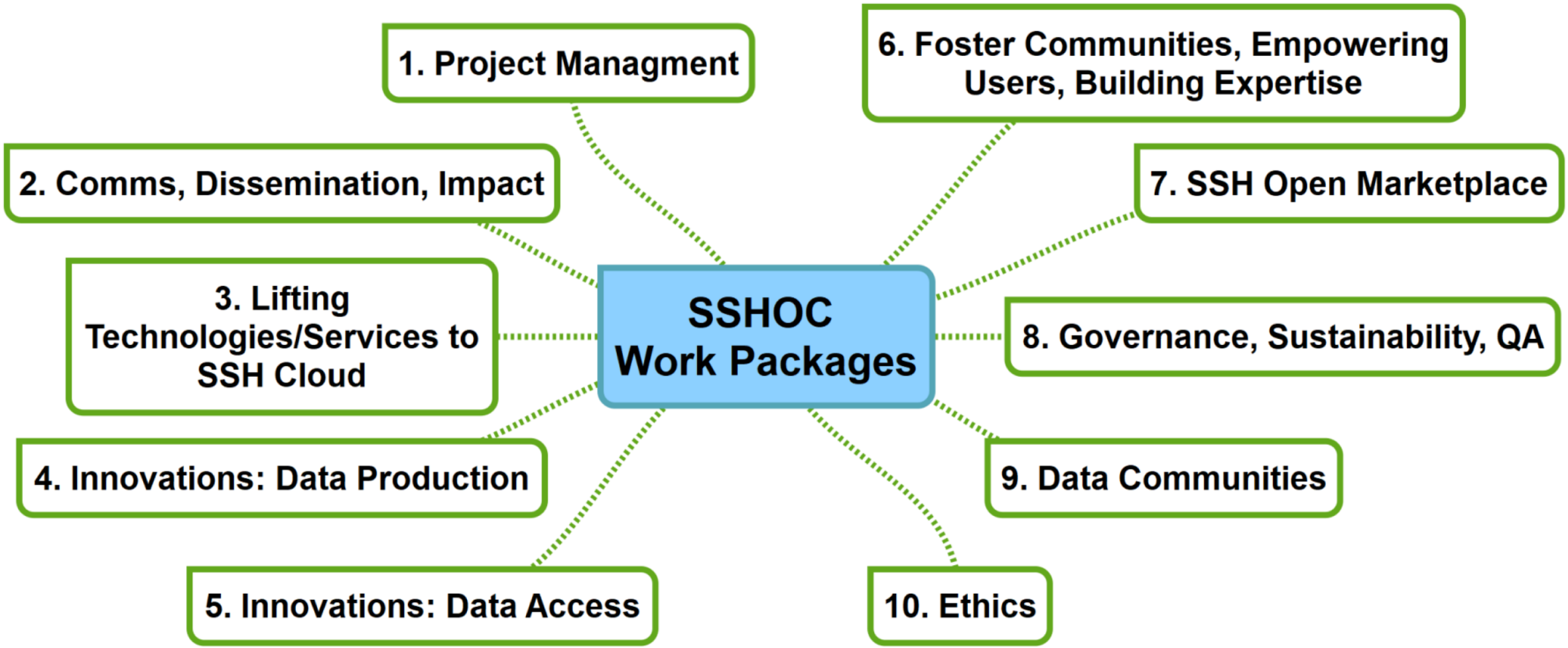
This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud



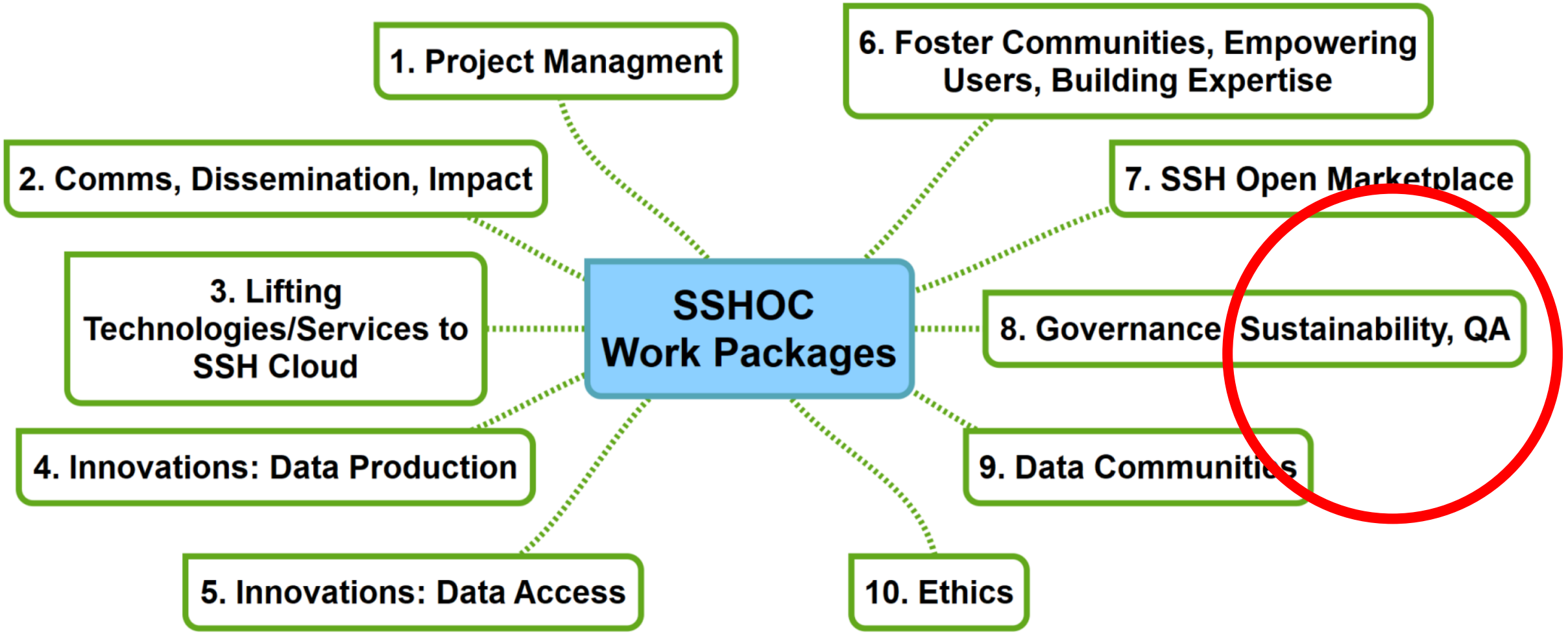
This project is funded from the EU Hor

SSHOC
social sciences & humanities open cloud



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

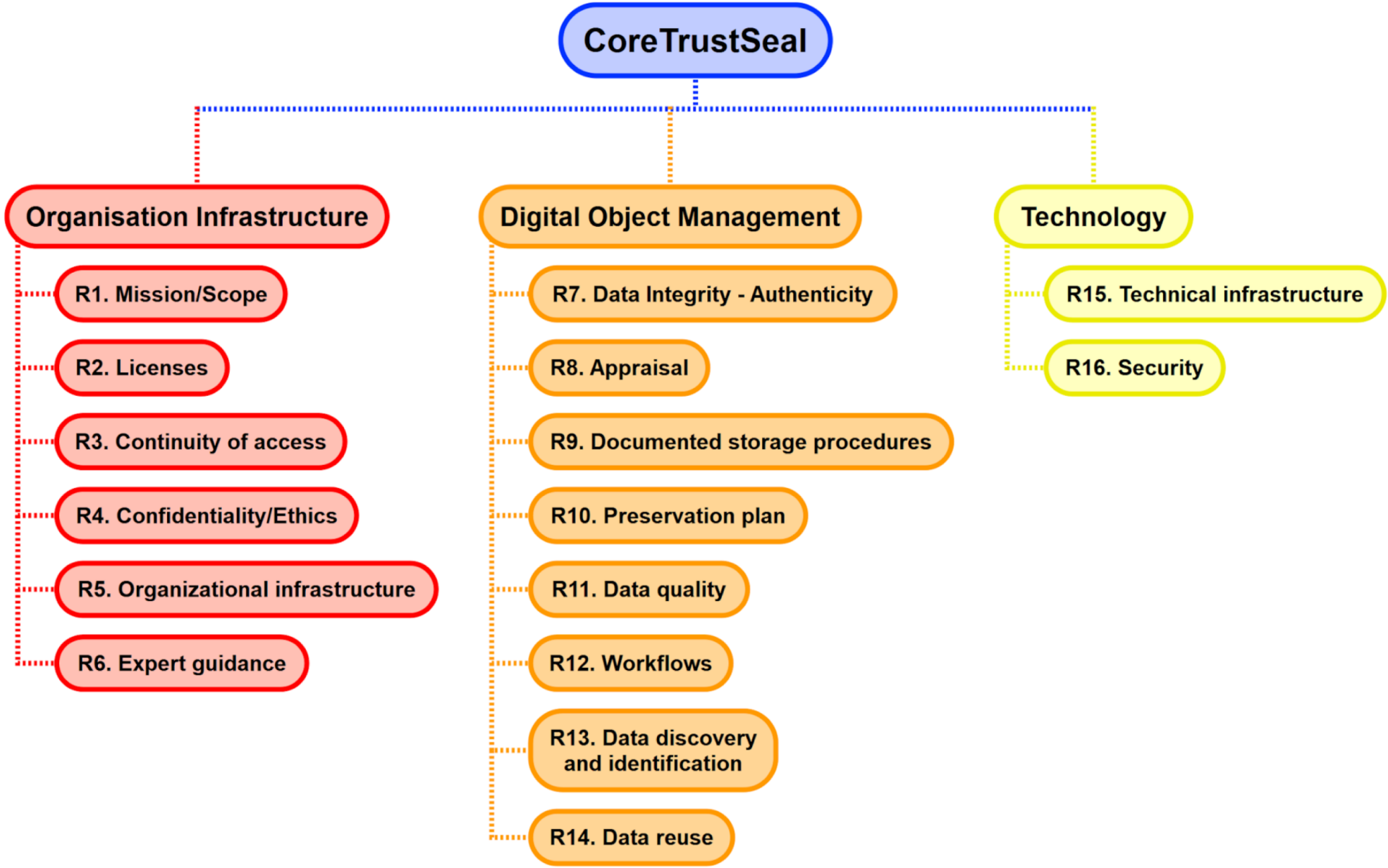
Which RDA recommendation & why?

- CoreTrustSeal Trusted Digital Repository (TDR) Requirements
  - Repository Audit and Certification Catalogues
  - RDA Repository Audit and Certification DSA–WDS Partnership WG
  - RDA/WDS Certification of Digital Repositories IG

Why?

- Core, low barrier to entry
- Community developed & managed
- Recommended (EOSC, Turning FAIR Data into Reality etc.)
- Responsive to community feedback

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

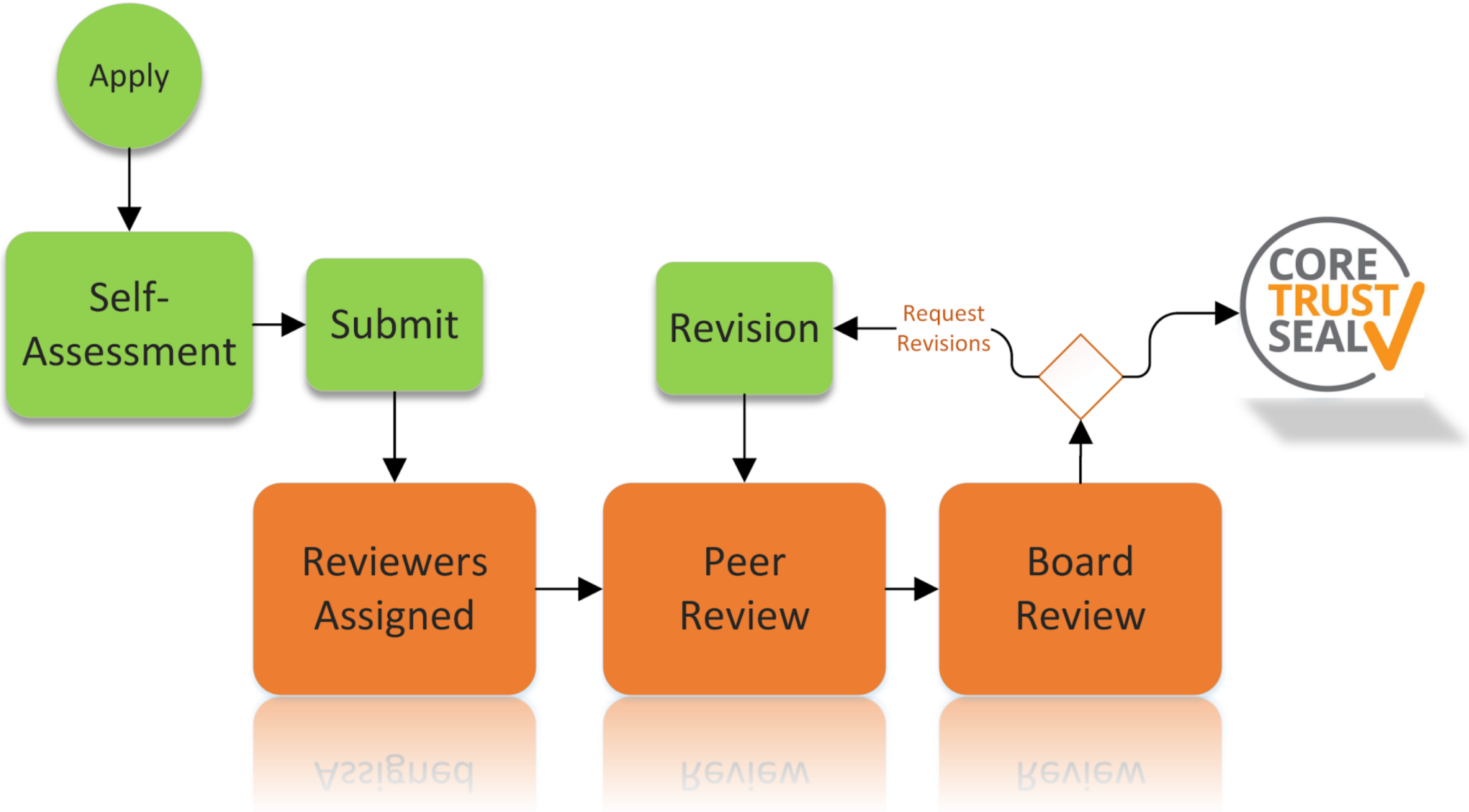
The Adoption Process (SSHOC T8.2)

- Support in reaching certification goals
  - Self-Assessment
  - Achieve CoreTrustSeal
  - Renew CoreTrustSeal (for the >30 certified centres)

- Flexible approaches:
  - Engagement and awareness raising (benefits, approaches, evidence)
  - Information outputs fully open
  - Workshops and webinars on repository certification
  - Events targeted at each community, but open to all
  - SSHOC-internal review of self-assessments (confidential)

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud



HLH-CoreTrustSeal-Process_00_03.vsdx

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

HLH-CoreTrustSeal-Process-SSHOC_00_03.vsdx

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Benefits of adoption and impact

- Benefits beyond formal certification:
  - Self-assessment improves internal communication/understanding
  - Facilitates comparison with peers
  - Supports cooperation with partner
  - Feeds back into data quality and process improvements

- Benefits beyond the obvious candidates
  - Wide variety of potential applications (beyond repositories)
  - Identify cases of ‘not applicable’ (and why)
  - Feedback into the CoreTrustSeal

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Benefits of adoption and impact

- Feedback into the CoreTrustSeal to better support:
  - Repositories
  - Wider infrastructure actors

- Active CoreTrustSeal engagement to work with:
  - general-purpose galleries, libraries, archives and museums
  - complex partnership models
  - technical infrastructure service providers

- Varied SSHOC applicants inform wider data steward support
- Sustainable cross-SSH ERIC trust work beyond SSHOC

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Lessons Learned from Adoption?

- CoreTrustSeal: flexible and broadly applicable
- Those who don't identify as traditional applicants see relevance
- All actors in these partnerships are critical EOSC components
    - Potential TDRs
    - Supporting data service providers
- CoreTrustSeal transparency ethos
    - SSHOC information shared wherever possible
    - Applied to a wide range of data stewards
    - Used for 'gap analysis

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

https://bit.ly/2MRXueg

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Join our community!

www.sshopencloud.eu

@SSHOpenCloud

company/sshoc

info@sshopencloud.com

CC BY

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Food and Agriculture Organization
of the United Nations

39 Hints to Facilitate the Use of Semantics for Data on Agriculture the Nutrition

FAO Adoption story

Caterina Caracciolo
FAO , Statistics Division

RDA Global Adoption Week
Online, June 17, 2020

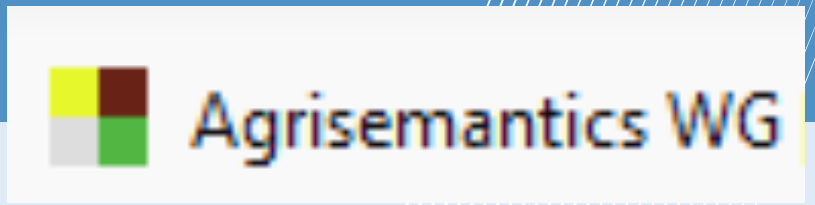
Food and Agriculture Organization of the UN

Part of the “UN family”, goal is food security

- HQ in Rome, offices in 130 countries
- Activities on policy, data

- FAO in RDA:
  - Agricultural Data IG (IGAD)
  - Agrisemantics WG
  - Fisheries Data Interoperability WG
  - Wheat Data WG

Agrisemantics WG



The challenge: Semantic data interoperability

Output focuses on:

- Tools: generic web framework for collaborative creation of semantic resources
- Reusability of semantic resources
- Adoption of standards
- ..across the entire lifecycle
- ..considering all user profiles involved

Why Agrisemantics is relevant to ag statistics

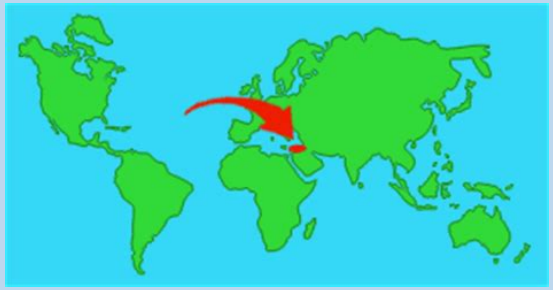
Statistical data heavily relies on classifications for their collection, use, analysis, integration.

But they are still largely treated “ad-hoc”. Utlimately, interoperability is hampered.

Agrisemantics recommendations offer a comprehensive view on how to make it better

Because… Words are ambiguous, codes too

Turkey



1079

Statistical classifications in practice

Domain Code	Domain	Area Code	Area	Element Code	Element	Item Code	Item	Year Code
QC	Crops	2	Afghanistan	5510	Production	01709.90	Pulses nes	2018
QC	Crops	3	Albania	5510	Production	01701	Beans, dry	2018
QC								

FCL Item code	FCL Title	CPC ver.2 code	*	CPC ver.2 title
0015	Wheat	01111		Wheat, seed
0015	Wheat	01112		Wheat, other
0016	Flour of Wheat	23110		Wheat and meslin flour
0016	Flour of Wheat	23130	*	Groats, meal and pellets of wheat and other cereals
0017	Bran of Wheat	39120	*	"Bran and other residues from the working of cereals or legumes; vegetable materials and vegetable waste, vegetable residues and by-products, whether or not in the form of pellets, of a kind used in animal feeding n.e.c."

Adoption @FAO , Statistics Division

Caliper, a web platform dedicated to statistical classifications

http://stats-class.fao.uniroma2.it/caliper/

Ultimate goal = have better statistics – more timely, of higher quality, better comparable

Note: The platform is not yet an official service of FAO. Data can be accessed and tested but should NOT be taken as reference.

Caliper

Statistical Classifications in an Open Linked World

HOME | CLASSIFICATIONS | BROWSE | DOWNLOAD | QUERY | EDIT | SOURCES | DOCUMENTATION | FAQ | NEWS | CONTACTS

Statistical Classifications in an Open Linked World

The timeliness and interoperability of agricultural statistics largely depend on the availability of classifications in usable and standard formats. Instead, they are o
in formats that are hardly or not-machine readable (e.g., PDF), or information is scattered in different files (e.g, multilingual titles and correspondences), or metada
hoc and not-machine readable formats.

The goal of this project is to address those limitations by applying open technologies for the web. We look at all phases in the life cycle of statistical classificatior

- **Publication**:
  - classifications can be **browsed and visualized** in different ways. We use three tools: SKOSMOS, PMKI, and the content management system Drupal.
- **Use**:
  - download classifications in machine readable formats
  - query and extract specific pieces of classifications on demand
  - programmatic access classifications, via APIs, or through content negotiation of web addresses (URIs).
- **Editing**:
  - VocBench as a editing platform.

Credentials: “caliper”, “caliper”

CREDITS: Caliper is part of a project run by the Food and Agriculture Organization of the UN (FAO) and supported by the Bill and Melinda Gates Foundation (BMGF
Vergata (Rome, Italy) provides technical and scientific support.

Caliper is

An aggregator of resources

- Based on semantic technologies and open standards
- Serving data to humans and machines

Recommendations implemented

1. Standard modelling and formats
   - Mostly, RDF-tech stack
   - SKOS, XKOS, URIs..
2. Reuse of common metadata models
   - DC, DCAT, VoID, for machine readability
   - Versioning and provenance information
3. A web-based platform
   - Addressing different users needs and roles
   - Open source tools and licenses

Some highlights

Coverage

Caliper

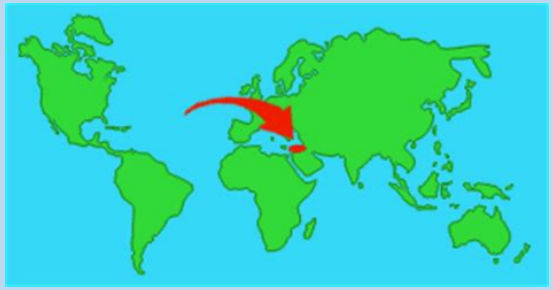
Vocabularies

from all | Any language

Vocabulary Categories

ACTIVITIES	ISICRev4
AID FLOWS	CRS 2016 05 CRS 2018 01
CROPS	ICC v1.0 ICC v1.1 WCACrops
FORESTRY	Forest Products Class. 2016
GEOGRAPHY	M49 FAO current M49 FAO Dec 2019 M49 FAO Jul 2019 M49 UNSD SDG grouping
NUTRITION	FoodEx2 2016
PRODUCTS	CPC v2.0 CPC v2.1 FCL HS
SOIL	WRB 1998

#1.Standard modelling and formats (URIs)



http://stats-class.fao.uniroma2.it/geo/m49/792

http://stats-class.fao.uniroma2.it/FCL/v2019/1079

http://stats-class.fao.uniroma2.it/ISIC/rev4/1079

#2. Common metadata models

Having in mind:

- machine interoperability
  - Dublin Core, DCAT, VoID, OWL procenance
- human credibility (!), to distinguish:
  - Author, version, etc of the classifications
  - Caliper as publisher of specific distributions

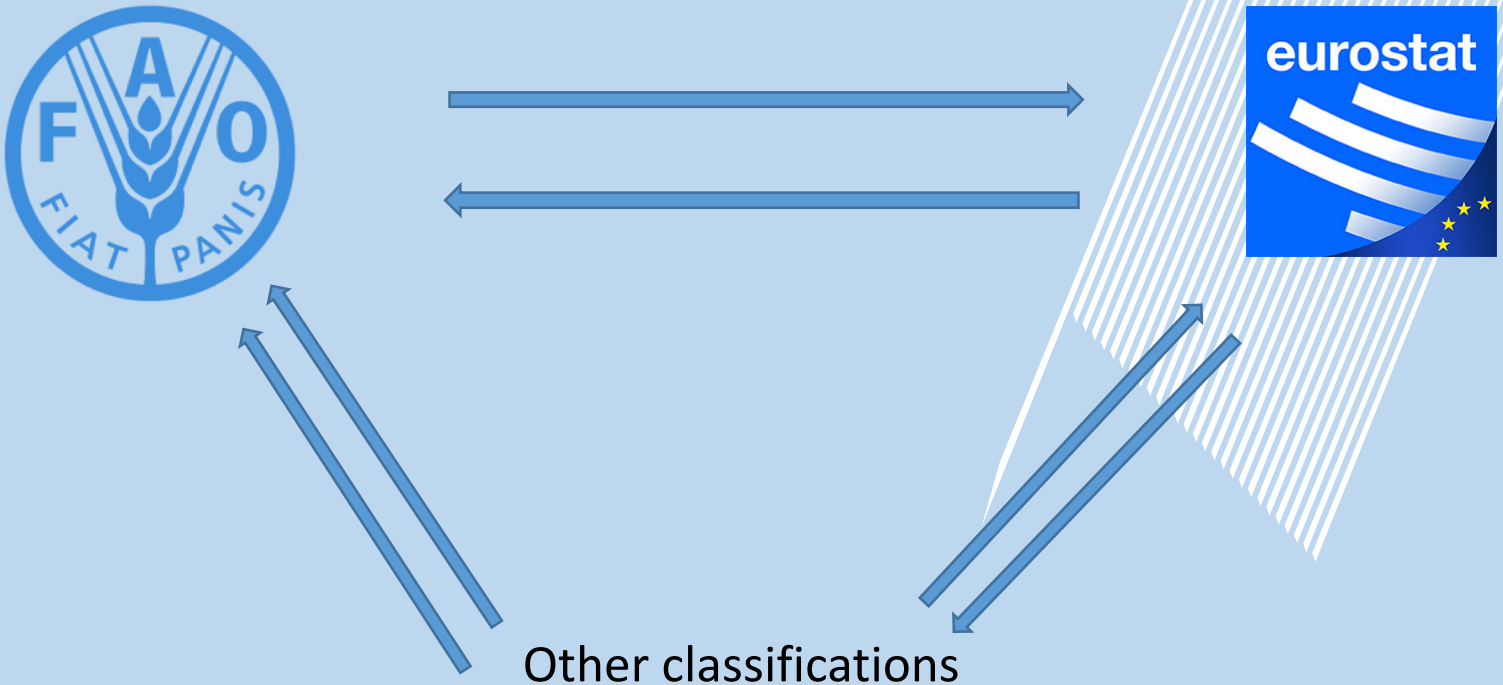
#3. Web platform and tools

All open source:

- Viz: SKOSMOS, Drupal, PMKI
- Editing: VocBench3
- Services: Apache Jena Fuseki, GraphDB, Yasgui

- Working towards a better integration of them, wrt
  - Data lifecycle
  - Users involved

Next: interoperability at a larger scale



http://stats-class.fao.uniroma2.it/caliper/

Thank you!

Agrisemantics @FAO

Caterina Caracciolo: caterina.caracciolo@fao.org

Caliper: http://stats-class.fao.uniroma2.it/caliper/

credentials: caliper, caliper

Extra! Survey on adoption of Agrisemantics

Prof. Chris Baker and **Dr. Brett Drury** are conducting a comprehensive survey of attitudes and opinions about the current state of the art and adoption of Agrisemantics. hey would like your opinions, and they would be grateful if you could complete a short questionnaire. See:

https://www.godan.info/news/new-study-agrisemantics-food-security

Contact: brett.drury@gmail.com

Tell your adoption story

- **Are you an adopter?** RDA is actively seeking new adoption stories to inspire the further uptake of RDA outputs.
- **Submit your story here:** https://www.rd-alliance.org/tell-your-rda-adoption-story

CODATA CfP Data Science Journal

- **RDA special collection themes:**
  - Results produced by an IG or WG;
  - Description of an Adoption Case outlining how a specific recommendation or output has been implemented;
  - Other types of work related to RDA activities.
- RDA Europe 4.0 still has funds available for the publication of articles in DSJ
- Open to all interested applicants regardless of their geographical provenance.
- **Deadline 17 July**

Recommendations and outputs catalogue

- RDA Outputs are classified as **RDA Recommendations** *(official, endorsed results of RDA Groups),* **Supporting Outputs** *(useful solutions from our RDA Working and Interest Groups)* or **other Outputs**
- **They can be searched according to their status, Data Life Cycle topics ??? or scientific domain**

rd-alliance.org/recommendations-and-outputs/catalogue

CC BY SA

WWW.RD-ALLIANCE.ORG/
@RESDATALL

RDA Global

Email - **enquiries@rd-alliance.org**
Web - **www.rd-alliance.org**
Twitter - **@resdatall**
LinkedIn - **www.linkedin.com/in/ResearchDataAlliance**
Slideshare - **http://www.slideshare.net/ResearchDataAlliance**

RDA Europe

Email - **info@europe.rd-alliance.org**
Twitter - **@RDA_Europe**

RDA US

Twitter - **@RDA_US**