



AUSSDA
AUSTRIAN
SOCIAL SCIENCE
DATA ARCHIVE

Building a new social science data archive – covering the first year

Lars Kaczmirek & Veronika Heider

@TheAUSSDA

Strengthening and widening of the European infrastructure of social science data archives, 14.-15.11.2018, Belgrade, Serbia

Have data? Need data? | www.ausdda.at

Outline

- 🌐 Decisions, decisions, decisions
- 🌐 AUSSDA mission and the Austrian context
- 🌐 Considerations about funding, legal entity, staff expertise
- 🌐 Administration: high level decisions and getting down to the nitty-gritty
- 🌐 Business development: strategies and other documents
- 🌐 Developing the core services
- 🌐 Communication: invest more than expected
- 🌐 Acquisition strategy: building a pipeline for continuous deposits
- 🌐 Ingest: your core process and how to get it right

Decisions, decisions, decisions

- Know your OAIS functions and terminology
- Know the CESSDA ERIC roadmap and projects
- Are your decisions bringing you closer to the CoreTrustSeal?

- Establish procedures for fast daily decisions
- Establish procedures for decisions with larger impact such as policies

- Phases: drafting, testing, executing, reviewing

The Austrian context

- Population: 8.8 million
- Area: 83.879 km²
- 9 federal states
- 22 public and 11 private universities
- Spendings in research & development in humanities, social and cultural sciences in 2013: 450 million Euro (18,2% of the total)
- AUSSDA is employing 10 people in 3 cities and has external legal support from two companies specializing on GDPR, copyright, licensing and contracts

Sources: Brustmann, U., M. Reiter-Pázmándy. (2017). Geisteswissenschaften, Sozialwissenschaften, Kulturwissenschaften: Beitrag zur strategischen Weiterentwicklung der Rahmenbedingungen der GSK. www.bmbwf.gv.at/gsk and Wikipedia

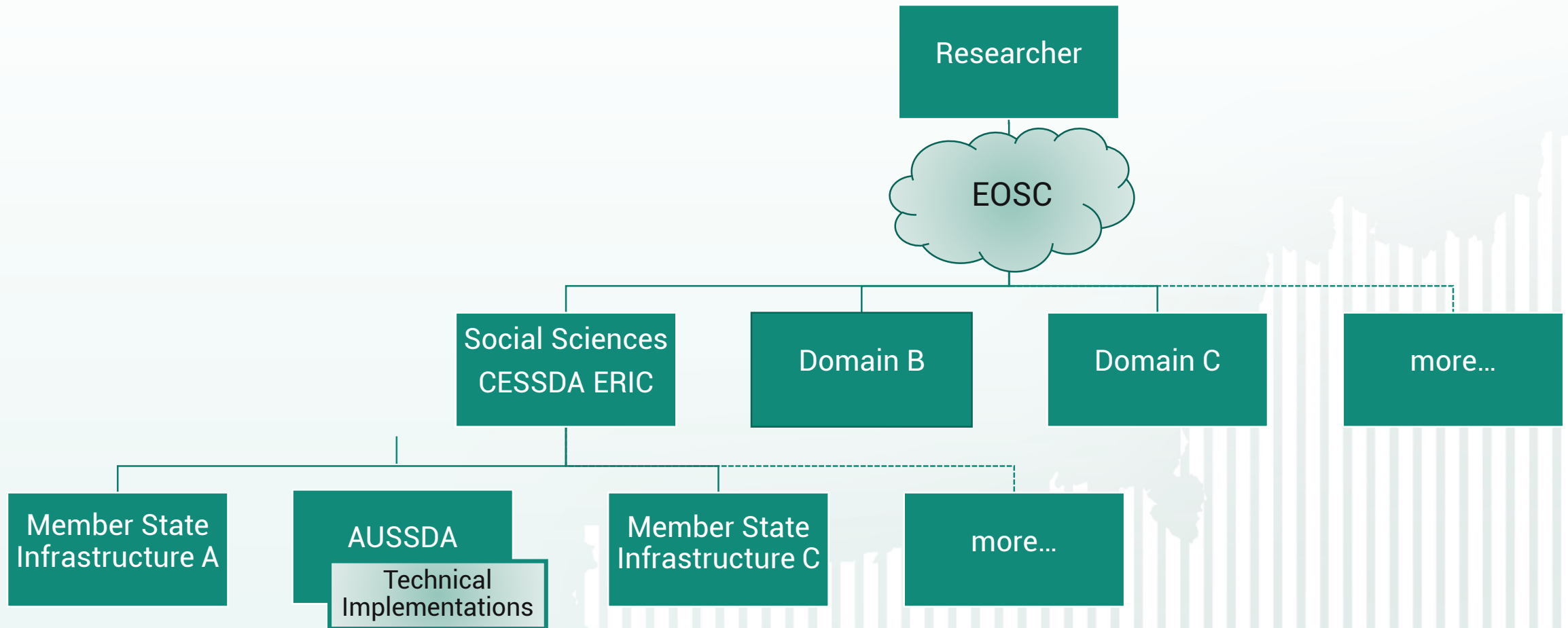
AUSSDA is a member of CESSDA ERIC and one of 6 ERICs in Austria

cessda eric



Connecting to the European Open Science Cloud

The EOSC federated model approach





AUSSDA

AUSTRIAN
SOCIAL SCIENCE
DATA ARCHIVE

“We make social science data accessible, creating opportunities for research and data reuse, benefitting science and society.”

Have data? Need data? | www.ausdda.at

AUSSDA Mission



AUSSDA ensures the sustainability of investments in research by establishing an infrastructure for social sciences in Austria.



AUSSDA archives data according to international standards and in line with CESSDA ERIC obligations and policies.



We offer sustainable solutions for archiving digital data and are prepared to adapt to future developments.



We support the open data movement and work towards maximizing the potential for data reuse.

Legal entity and funding

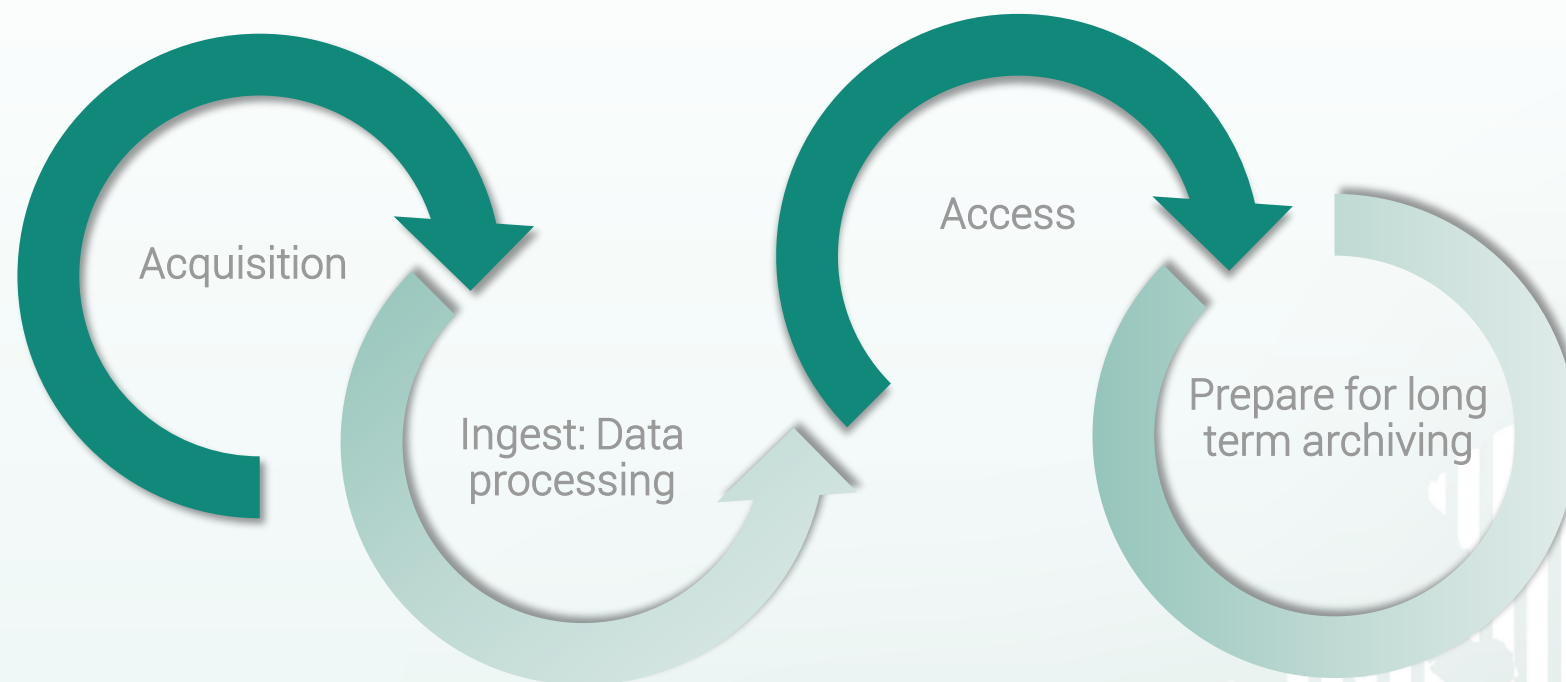
- Who is the legal entity?
- Who can sign contracts?

- At a university department, the library, research institution, association?
 - Consider reputation
 - Consider multiple locations
 - Consider flow of money

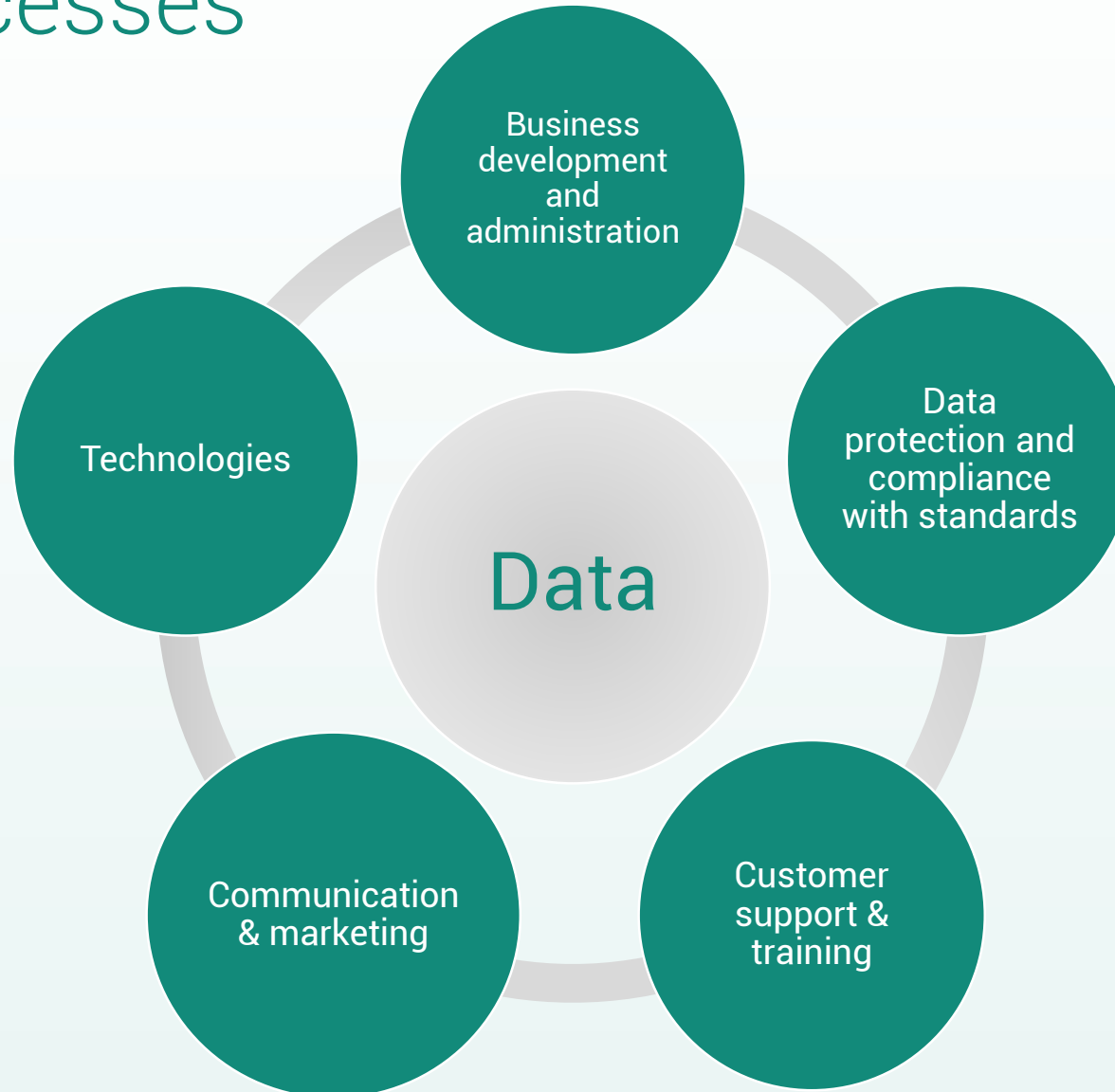
Your team

- Be prepared to invest substantially in training because it is unlikely to find the perfect match for all needed expertise and knowledge even in highly skilled staff
- Have a contingency plan: people go on parental leave, get promoted to somewhere else
- Deputy roles
- Dedicated expertise
- Dedicated topics to monitor (open science, technology, newspapers, etc.)
- Risk assessment and Plan B (e.g., service failure communication)
- Example IT infrastructure: How many points of single failure do you accept?

Key roles for staff: Expertise to manage data



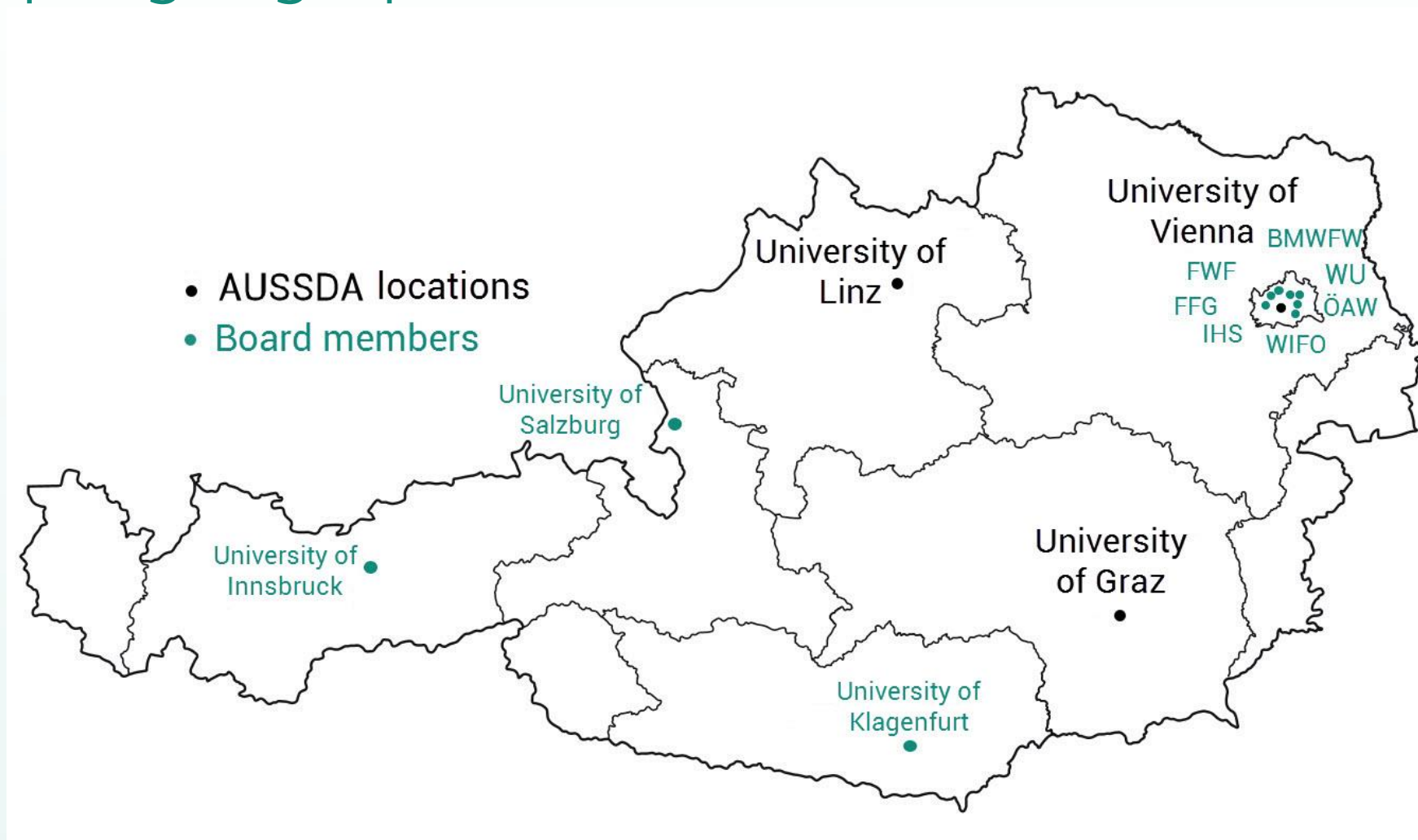
Key roles for staff: Expertise to support the core processes



Administration: high level decisions ...

- Which advisory/user boards do you need?
 - Meeting schedule, what do you expect them to do during the year?
- What is your primary and secondary language?
 - This affects documents, the repository, your metadata, communication material and your ability to share and use knowledge
- Will you be contributing to CESSDA projects?
- Will you cooperate with other archives, activities, organizations in your country? Assess the who is who

Example geographical outreach



Administration: ... and getting down to the nitty-gritty

- Plan with milestones
 - Makes your success visible
 - Helps in risk management
- Organize your documentation so that it builds and meets your reporting needs
 - E.g., weekly updates, monthly reports, quarterly review -> board meetings -> annual reports
- Plan folder structure for documents, naming, sending, saving procedures
- Use state-of-the-art cloud-based software solutions: project management and task management, Wiki, ticket system, website tracking

Why is it important to think about the order of what you develop?

- Consider the workflow and dependencies of the data pipeline
- If you do not know what data to look out for, decisions to accept or decline offers will be arbitrary and time-consuming
- Think from your core process – ingest.
 - What needs to be working so that data comes in? Contracts
 - What needs to be working so that users can find data? Digital repository
 - What needs to be working so that data gets out to users? Access workflow
- You can tolerate to develop several aspects at once
- It is important to closely monitor and be aware of your bottlenecks in the workflows while you develop at all ends

What did we develop first?







- Mission statement
- Data collection policy
 - Defines dedicated community and disciplines
 - Defines preferred data types and formats
 - Outlines target groups
 - Outlines access options
 - Outlines licensing options
 - Outlines metadata scheme
 - If possible cover special cases that will likely occur at some point: qualitative data, sensitive data, big data, larger amounts of unstructured digital and non-digital data (e.g., from archives that are being closed or researchers preparing their legacy)

What about GDPR?

- The EU General Data Protection Regulation affects many of the workflows and rules in an archive
 - Administration (e.g., newsletter, website tracker, cookies)
 - Archive material (survey data, questionnaires, interviews)
- Ensure availability of legal advice
- Develop your legal framework
 - the data deposit contracts
 - how you give access to data
 - Among many others...

Example

Data deposit and license models

-  Open access offer:
 - Creative Commons Attribution 4.0 International Public License
 -  Additional option embargo: granting a license to AUSSDA that allows AUSSDA to publish under a CC license on a specified date
-  Scientific use
 -  Restricted safeguarded access (semi- or fully automatic)
 -  Restricted controlled access (additional conditions may apply, human gatekeeper)
-  Special conditions

What we covered in our deposit agreements

A selected list of addressed issues

- Transfer of archive materials
- Granting of non-exclusive, temporally and geographically unlimited rights to the archive
- Liabilities
- Succession planning
- Metadata is CC0
- Material can be published under CC BY or scientific use only or under special conditions
- German and English versions



AUSSDA

AUSTRIAN
SOCIAL SCIENCE
DATA ARCHIVE

Communication & Data Acquisition

Know your community!

Have data? Need data? | www.ausstda.at

What we did so people can find us...

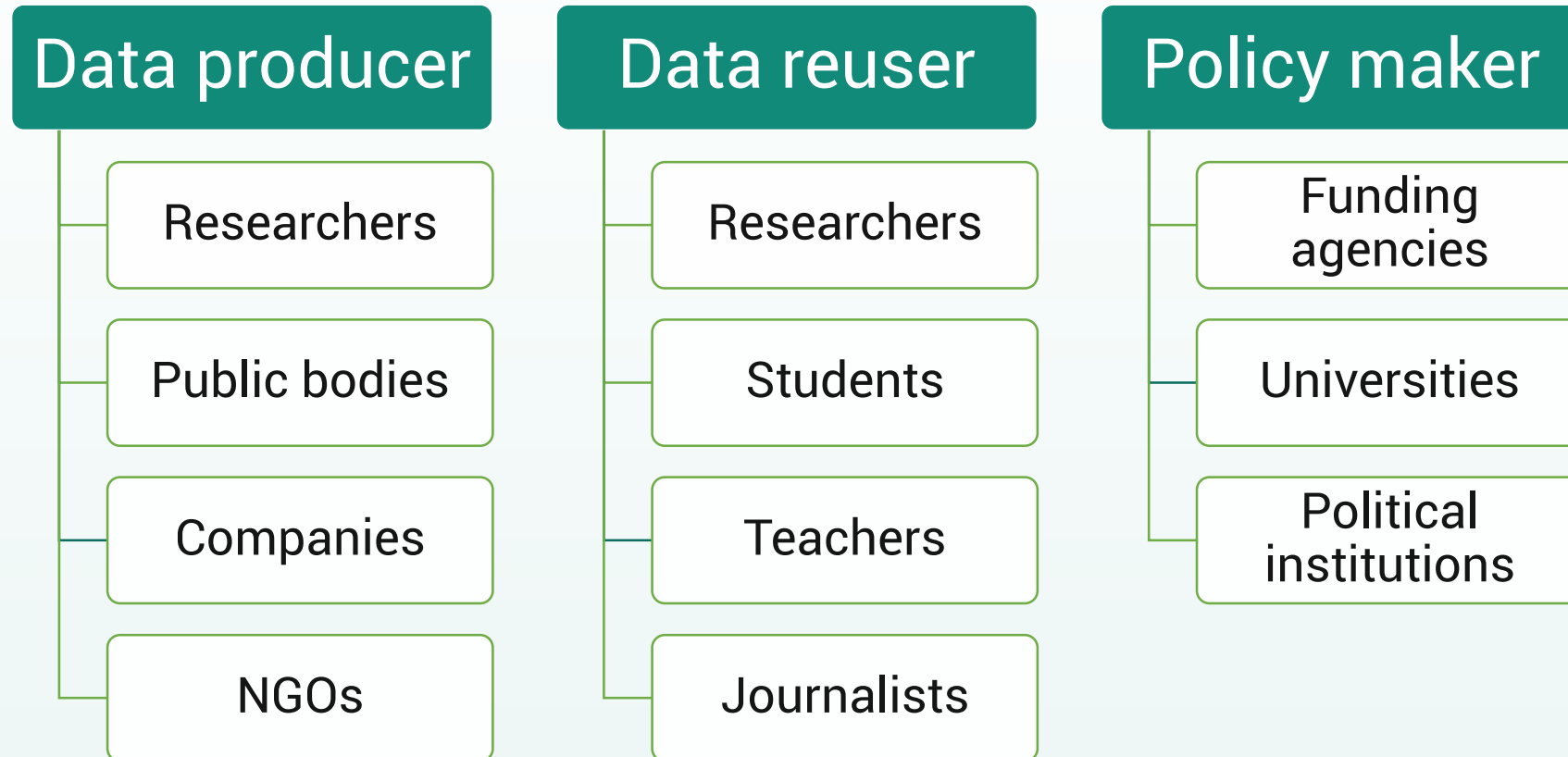
- Website: up and running as soon as possible
 - <https://www.aussda.at/>
 - We put 5 pages online about the organisation that was being established
 - In German and English
- Twitter: @theAUSSDA
- One central e-mail: info@aussda.at
- One central telephone number: +43 1 4277 15323



The ideas behind AUSSDA's communication

- Be approachable! Make the archive reachable and personal via
 - Website, E-mail & phone, Twitter
- Create a recognisable corporate design
- Create content
 - Material to be shared: handouts for researchers and funders, video
 - Material to support & inform: begin with weekly news pieces about services (e.g. data management plans)
- Understand and engage your community
 - To which group do we talk through which channel?
 - Twitter, News, E-mail, Community events, Newsletter

Who are key stakeholder groups?

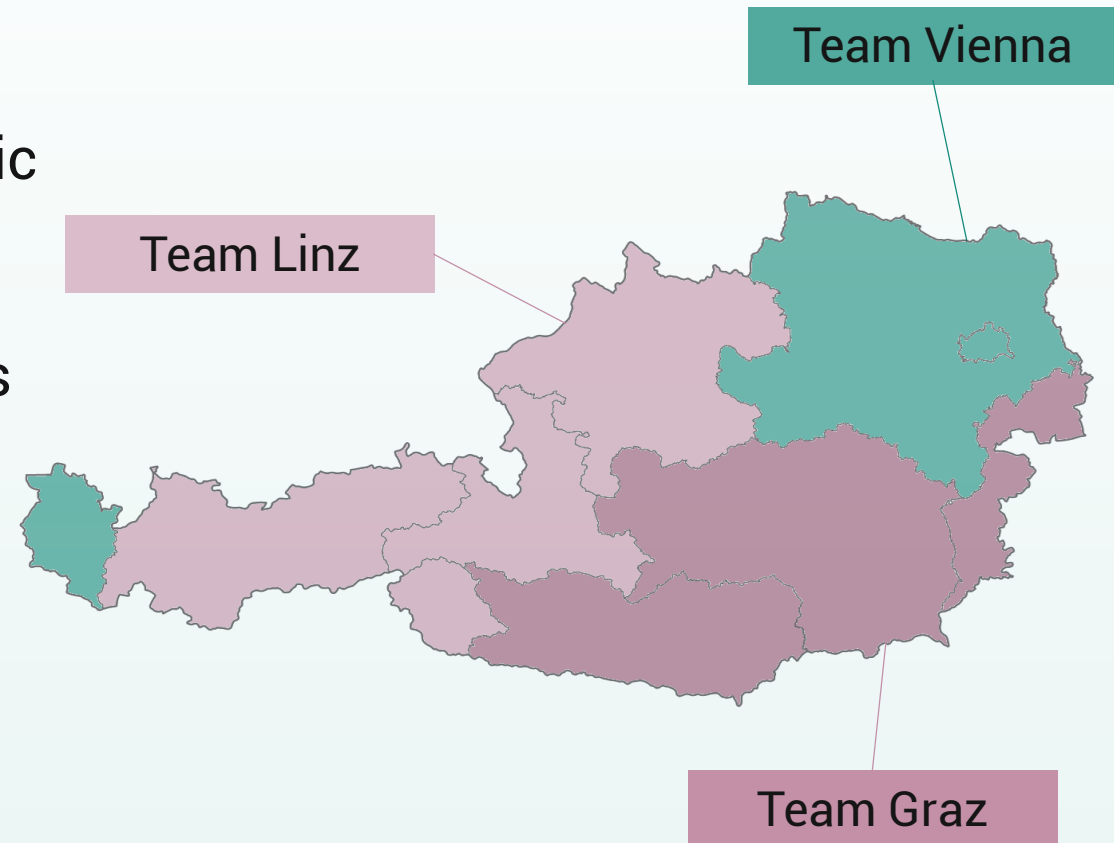


Data acquisition in the first year

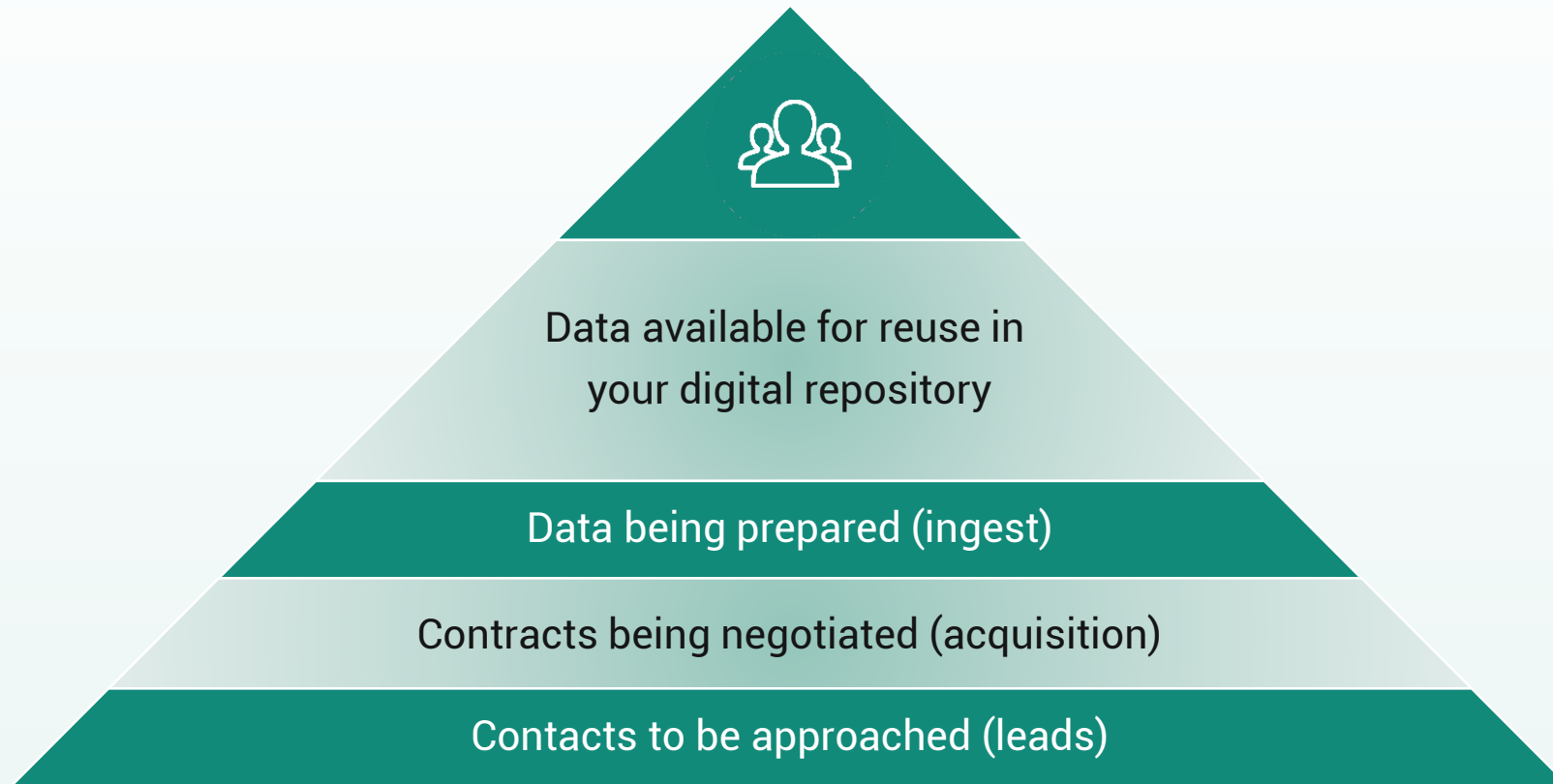
- The goal is to get your archive to grow continuously by
 - Selective acquisition: reaching out to specific data producers with very important/good data sets
 - Continuous, semi-automatic acquisition: agreements ensure steady input of data by journals, data collection agencies (e.g. survey institutes), funding agencies (data sharing policies of institutions), research management faculties, and research groups
 - Bulk acquisition: getting data from closures of an archive or organization
 - Licensing key datasets (e.g. census data)
- Find cooperation partners who are dedicated to build your institution with you
 - Data producers: who stay committed despite slow procedures
 - Data reusers: who give you feedback
 - Funding agencies & other archives: cooperation agreements, support networks...

Regional acquisition

- Acquisition is done by region or by expertise in an academic discipline
- We offer roadshows & workshops locally, we attend community specific events, we send out traditional PR, we write publications, ...
- The regional personnel has contacts in the regional institutions (“ambassadors”)
- Responsible for getting contracts signed, metadata sheet and data from data depositors



Building your data pipeline





AUSSDA

AUSTRIAN
SOCIAL SCIENCE
DATA ARCHIVE

Data Processing: Ingest

Have data? Need data? | www.ausoda.at

Acquisition is finished – what now?

- Ingest is about working with data and people
 - It comprises data checks, communicating with data depositors, publishing datasets, ...

- Many processes and workflows need to be developed in the first year
- get big dataset collector friends to be your testing partners

- Pay attention to standards (CESSDA topic classification, ELSST/SSMT, DDI, ...)
- Taking OAIS and preparing for long term preservation serious:
 - SIP - AIP - DIP

Adding (high-quality) data to your archive: Ingest

- Ingest workflows are documented in a wiki system
- Tasks are tracked in our project management software
- Communication with data depositors is done in our ticket system to make hand over between steps easier
- Every dataset is checked by at least two persons
 - At some point one can think about a transition to checks by one person and supervision and quality assurance
- Work closely with the depositor: depending on time and resources in your archive the depositor needs to be an active contributor

Likely challenges in the first year

- Acquiring expertise & personnel
- Legal contracts
- Getting to know the digital archive software
- Deciding on formats and naming conventions
- Establishing workflows
- Creating templates for different steps
 - Data_checks.docx
 - Content.xlsx
 - History.xlsx
 - Data_checks.do
 - ...

Ingest

- Creating Submission Information Package (SIP) & Archival Information Package (AIP)
 - Did we receive all files? In correct formats? Is the contract signed? Is the metadata sheet filled out?
- Data processing – Compare data & documentation
 - Metadata checks
 - Data quality checks
 - Plausibility checks
 - Anonymization checks

Ingest (continued)

- Send feedback to depositor
 - Do final changes (PDF/A, add doi/version to dataset, create other data formats)
- Adding to the digital archive
 - Creating a dataset in AUSSDA Dataverse (metadata, terms, files, permissions)
 - Publish = Success!
- Post publishing
 - Get back to data depositor
 - Hand over Dissemination Information Package (DIP) to long-term archiving

Our digital archive

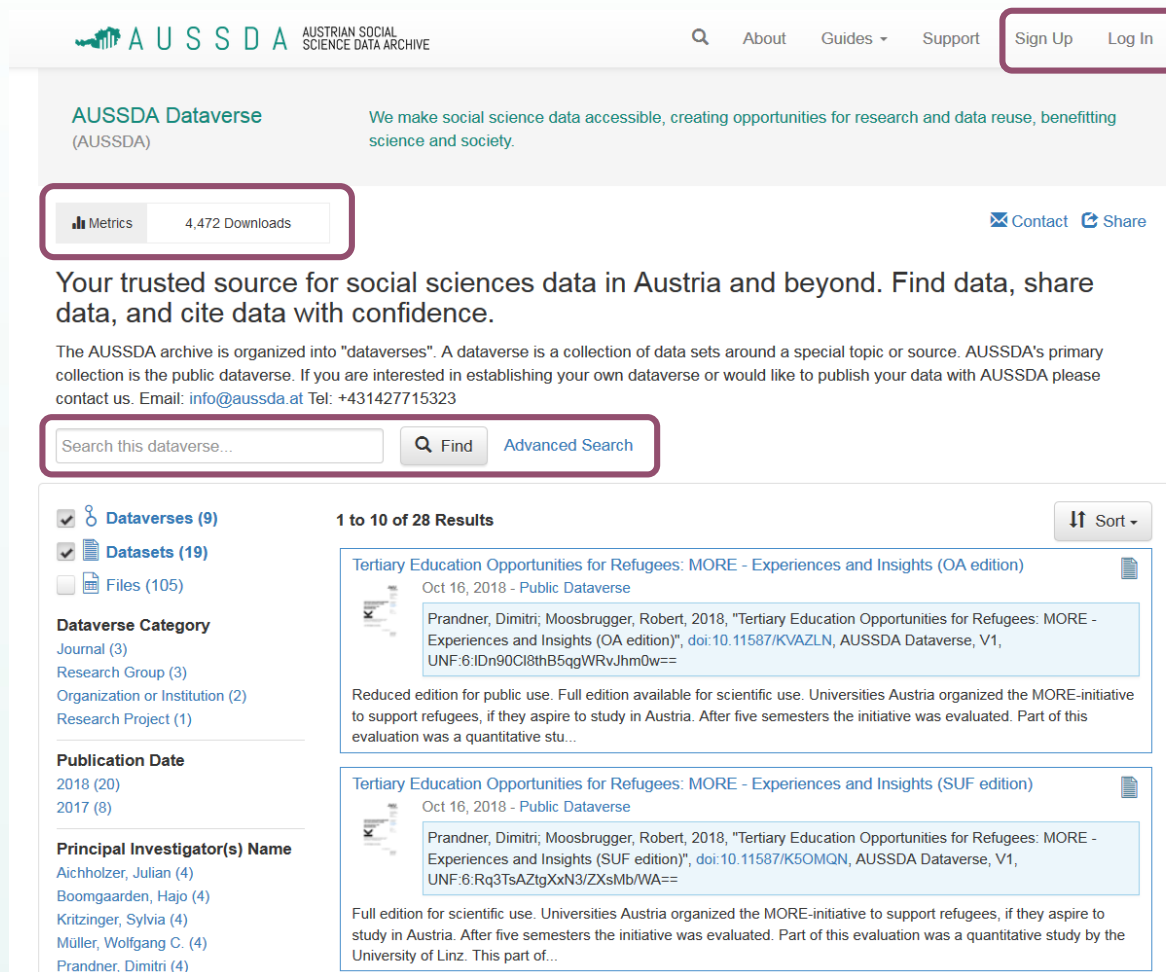
- We started with NESSTAR, knowing that we would transition soon to
- Parallel setup of Dataverse, <https://data.ausssda.at/>
- At the moment, data is entered by AUSSDA staff

- Every technical solution needs to be bought or developed, we are working on
 - Multilingual software (German and English)
 - CESSDA compatibility (implementation of controlled vocabularies)
 - Improving the usability and the user experience for our reusers
 - Data migration from NESSTAR to Dataverse

Our online archive: AUSSDA Dataverse

See download numbers at a glance

Search all fields



AUSSDA AUSTRIAN SOCIAL SCIENCE DATA ARCHIVE

AUSSDA Dataverse (AUSSDA) We make social science data accessible, creating opportunities for research and data reuse, benefiting science and society.

Metrics 4,472 Downloads

Search this dataverse... **Find** [Advanced Search](#)

Dataverses (9)
Datasets (19)
Files (105)

Dataverse Category
 Journal (3)
 Research Group (3)
 Organization or Institution (2)
 Research Project (1)

Publication Date
 2018 (20)
 2017 (8)

Principal Investigator(s) Name
 Aichholzer, Julian (4)
 Boomgaarden, Hajo (4)
 Kritzinger, Sylvia (4)
 Müller, Wolfgang C. (4)
 Prandner, Dimitri (4)

1 to 10 of 28 Results

Tertiary Education Opportunities for Refugees: MORE - Experiences and Insights (OA edition)
 Oct 16, 2018 - Public Dataverse
 Prandner, Dimitri; Moosbrugger, Robert, 2018, "Tertiary Education Opportunities for Refugees: MORE - Experiences and Insights (OA edition)", doi:10.11587/KVAZLN, AUSSDA Dataverse, V1, UNF:6:IDn90Cl8thB5qgWRvJhm0w==
 Reduced edition for public use. Full edition available for scientific use. Universities Austria organized the MORE-initiative to support refugees, if they aspire to study in Austria. After five semesters the initiative was evaluated. Part of this evaluation was a quantitative stu...

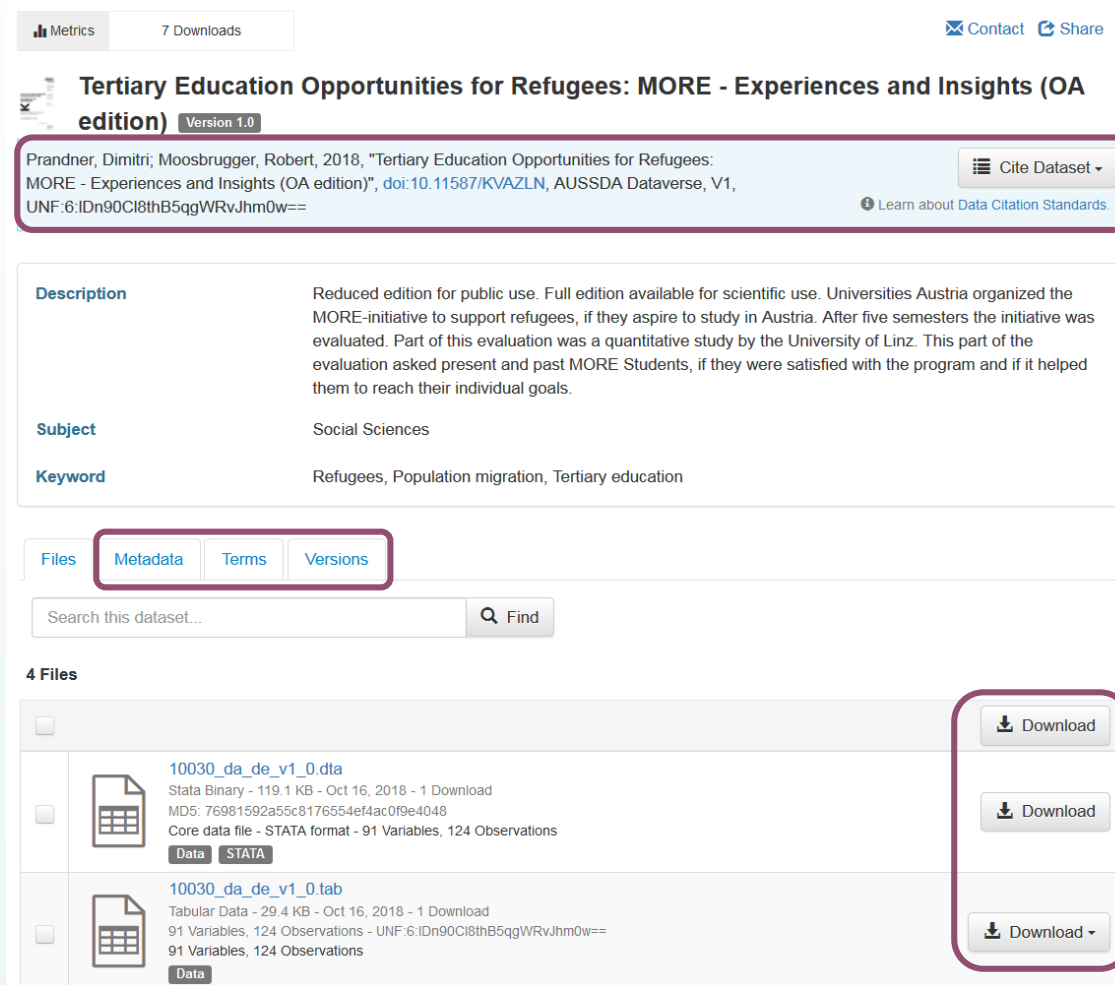
Tertiary Education Opportunities for Refugees: MORE - Experiences and Insights (SUF edition)
 Oct 16, 2018 - Public Dataverse
 Prandner, Dimitri; Moosbrugger, Robert, 2018, "Tertiary Education Opportunities for Refugees: MORE - Experiences and Insights (SUF edition)", doi:10.11587/K5OMQN, AUSSDA Dataverse, V1, UNF:6:Rq3TsAZtgXxN3/ZXsMbWA==
 Full edition for scientific use. Universities Austria organized the MORE-initiative to support refugees, if they aspire to study in Austria. After five semesters the initiative was evaluated. Part of this evaluation was a quantitative study by the University of Linz. This part of...

Create a new account under "Sign Up" or "Log In" with your institution's login

New released datasets are on top of the list

Accessing datasets in the AUSSDA Dataverse

Cite the dataset, including the doi



Metrics 7 Downloads Contact Share

Tertiary Education Opportunities for Refugees: MORE - Experiences and Insights (OA edition) Version 1.0

Prandner, Dimitri; Moosbrugger, Robert, 2018, "Tertiary Education Opportunities for Refugees: MORE - Experiences and Insights (OA edition)", doi:10.11587/KVAZLN, AUSSDA Dataverse, V1, UNF:6:IDn90CI8thB5qgWRvJhm0w== Cite Dataset Learn about Data Citation Standards

Description Reduced edition for public use. Full edition available for scientific use. Universities Austria organized the MORE-initiative to support refugees, if they aspire to study in Austria. After five semesters the initiative was evaluated. Part of this evaluation was a quantitative study by the University of Linz. This part of the evaluation asked present and past MORE Students, if they were satisfied with the program and if it helped them to reach their individual goals.



Subject Social Sciences

Keyword Refugees, Population migration, Tertiary education

Files Metadata Terms Versions

Search this dataset... Find

4 Files

| | | | |
|--------------------------|---|--|----------|
| <input type="checkbox"/> |  | 10030_da_de_v1_0.dta Stata Binary - 119.1 KB - Oct 16, 2018 - 1 Download MD5: 76981592a55c8176554ef4ac0f9e4048 Core data file - STATA format - 91 Variables, 124 Observations Data STATA | Download |
| <input type="checkbox"/> |  | 10030_da_de_v1_0.tab Tabular Data - 29.4 KB - Oct 16, 2018 - 1 Download 91 Variables, 124 Observations - UNF:6:IDn90CI8thB5qgWRvJhm0w== 91 Variables, 124 Observations Data | Download |

More information about metadata, terms of use and access, versioning of the data

Fast & easy download of Open Access files



AUSSDA

AUSTRIAN
SOCIAL SCIENCE
DATA ARCHIVE

The journey continues

Have data? Need data? | www.ausoda.at

How to plan for the future

- Run pilot projects or at least a reviewing project for everything important that you cannot cover during the build-up phase
 - Big data and geocodes
 - Citizen science
 - Qualitative data

Step up your game by identifying and adding relevant services for your community

- Work on defining your unique selling points
- What will no one else do?
- What can be achieved by teaming up with others?

- DOI
- Recommendation on how to cite a data set
- Individual branding of collections
- Data management plans
- Informed consent templates
- Trainings, train the trainers
- Team up with other services at the universities
- Twitter

Preparing for your second year

- Make a major event (e.g., launching your digital repository)
 - Inviting international and national well-known speakers
 - Let them praise your archiving efforts and the importance for the country
 - Invite the press
 - Consider one talk on a substantial topic, preferably showing that you have very valuable data available for reuse that benefits society
- Expand your services to other data types (if this is part of your strategy)
- Expand community engagement
- Strengthen stakeholder engagement



“Building a new social science data archive - covering the first year” by Lars Kaczmirek and Veronika Heider is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

We would like to acknowledge and thank the following people for their support and feedback: Lisa Hönegger, Julian Ausserhofer, Christian Bischof, Otto Bodi-Fernandez, Iris Butzlaff, Andreas Loibl, Dimitri Prandner, Lena Raffetseder, Barbara Sanchez Solis, Annika Schwabe, Frank Andreas Sposito.

The license covers only the text on the slides.

Images, the AUSSDA logo and design, the font and other aspects of the presentation may be subject to copyright, trademark and possible other rights.



AUSSDA

AUSTRIAN
SOCIAL SCIENCE
DATA ARCHIVE

Thank you!

Lars Kaczmirek

Veronika Heider

info@aussda.at

+43 01 4277 15323

www.aussda.at

Have data? Need data? | www.aussda.at