Researchers and their Data

Results of an Austria-wide survey

Executive Summary

e-infrastructures
austria
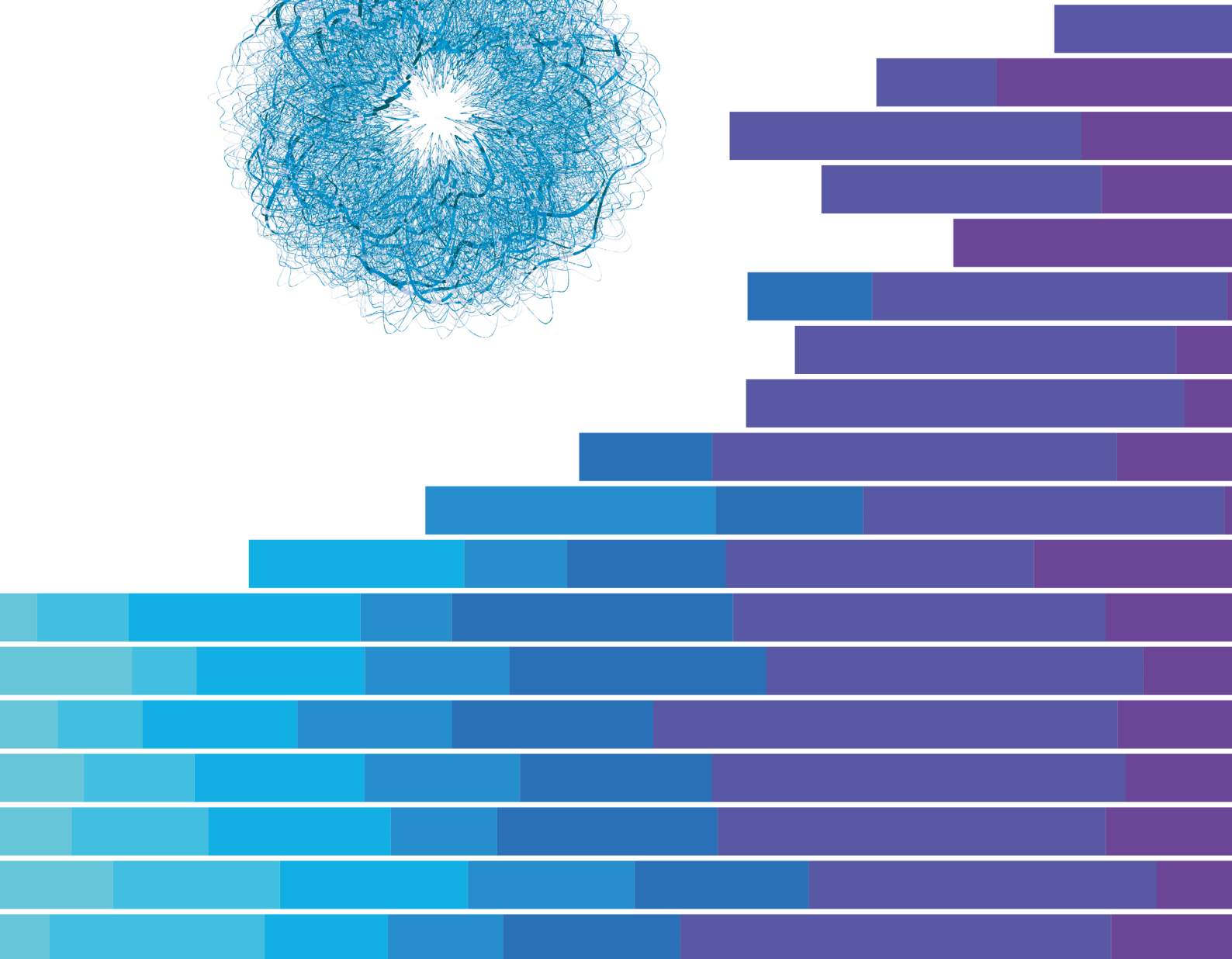
Introduction

This report provides an overview of the Austria-wide survey for research data, which was carried out within the framework of the project e-Infrastructures Austria[3] at the beginning of 2015. This survey was directed at the scientific and artistic-scientific personnel of all 21 public universities and three extramural research institutions in Austria.

The participants were asked about the following topics:
  – Data types and formats
  – Data archiving, backup and loss
  – Ethical and legal aspects
  – Accessibility and subsequent use
  – Infrastructure and services

This first inquiry conducted at a national level in this context facilitated the collection of methods for the practical handling of research data in Austria, and is therefore the basis: (1) for an on-going effort to optimize the infrastructure, (2) for an adaptation of the services provided, as well as (3) for a reorientation of the identification method for resources in this strategic area, which correspond to the expressed needs of people in the research process.

Background

A solid research data management system is the foundation of cooperative, open research and, thus, of its reproducibility and verifiability. Both, visibility as well as the reputation of the Austrian research landscape, also play an important role. This topic is relevant and up-to-date for researchers, funding bodies and senior posts of scientific institutions, which is currently demonstrated by the *Pilot for Research Data* of the European Commission[4].

Methodology

The survey is based on institutional, discipline-specific surveys that have already been performed at universities and research institutions in other countries. Emphasis was placed on the creation of a specially developed questionnaire which took into consideration the different cultures in science and art. This was created in German and English and was programmed by means of the open source-software, *LimeSurvey*. The implementation period was January 19 to March 31, 2015. The 3026 completed questionnaires (equivalent to an average response rate of 9 %) were statistically evaluated using the open source statistical software R and the open source spreadsheet OpenOffice.

Key Findings

The results of the study confirm the current expectations with respect to the handling of research data and settle them statistically. For each thematic area, both cross-curricular commonalities as well as discipline-specific features were determined where relevant.

---

3   Projekt e-Infrastructures Austria, Website. Online unter: http://e-infrastructures.at/startseite (Zugriff: 30.09.2015).

4   European Commission: HORIZON 2020, Open Science (Open Access). Online unter: http://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access (Zugriff: 30.09.2015).

Data types and formats

- From the majority of researchers, research data are generated in the form of unstructured text files, graphics and tables. One quarter of the participants use structured text, one quarter videos, databases and source code, and a fifth use audio and software. While in the technical disciplines – as expected – source code and configuration data are frequently generated, whereas the relatively frequent production of databases in the humanities and medicine is particularly striking.
- The majority of the researchers produced more than three quarters of their research data volume in digital form; whereupon analog data is often used only by slightly more than every twentieth participant (especially in the humanities).

Data archiving, backup and loss

- The majority of respondents use multiple storage options, where a clear preference was observed for the use of business and private computers as well as external hard drives and USB drives.
- Two-thirds of researchers require memory in the order of up to 100 GB per year. A higher memory requirement can be observed for both the medical and the artistic universities.
- More than two thirds of researchers indicate that they describe their research data individually and inconsistently, and more than nine out of ten of these indicate they are responsible for the archiving of research data themselves.
- More than one-third already had experiences with data loss.

Ethical and legal aspects

- While one-third of researchers claim to have never or rarely been confronted with legal ambiguities in foreign data usage, one fifth experienced legal uncertainties at least sometimes.
- When switching institutions, research data tends to remain at the entity concerned; from about half the researchers, this data was also transferred to the new institution.
- Sensitive data is used by every seventh researcher. Playing a major role here is medicine, which shows that four out of ten researchers often use such data.

Accessibility and re-use

- The use of external data is considered by many researchers to be a key aspect of their research, while a quarter of researchers use no external data.
- Access to self-generated research data by third parties is usually allowed by researchers to limited degree. While slightly more than half of the respondents allow access only on request, only one in ten provides their research data as open data for the public; the same number of researchers deny access altogether.
- Access to research data is made possible by the majority of researchers either by physical disks or via e-mail. More than two thirds of the researchers use cloud or website applications for this; data archives/repositories are used by every seventh researcher.
- Approximately one third of respondents allow the re-use of their own research data; this is occurring more often in geography, biology and chemistry, while this is occurring relatively less frequently in medicine, social sciences and humanities.
- User agreements are being entered into by more than one third of the researchers.
- For more than half of the researchers, the most attractive incentives for sharing their data were increased visibility and impact, new cooperation opportunities, recognition in professional circles as well as their consideration being referred to as scientific output.
- Alternatively, the impediments for such were mainly the increased time and costs, a possible misuse of data, legal uncertainties, potential data corruption, unwanted commercialization and increasing competitive pressure. Legal restrictions in particular were main obstacles in medicine, the social and behavioral sciences and the engineering sciences.

Infrastructure and Services

- With respect to the preferred data archive, the researchers show no clear preference. Mentioned quite often in this context are the international, specialized data archive, the institutional repository, the international multidisciplinary data archive and the nationwide specialized repository.
- The majority of researchers desire technical infrastructure and project-specific support for research data management. In addition, more than one-third show interest in legal advice, a general help desk, as well as training programs.
- More than half of the researchers expect the provision of additional qualified staff as well as the adoption of guidelines or policies for dealing with research data. One fifth desire that research data management be accepted as part of the educational curriculum and that it becomes anchored as a service requirement.

Recommendations

Based on the present survey results, the implementation of the following measures for the handling of research data in Austria is strongly recommended:
- Creation of a comprehensive, technological infrastructure in Austria, including any disciplinary needs
- Adoption of institutional policies
- Attainment of information professionals
- Implementation of supporting services for researchers
- Implementation of appropriate incentive systems
- Encouragement of international and interdisciplinary cooperation

These recommendations above aim to initiate highly efficient infrastructures for the proper handling of research data in the Austrian scientific landscape. In its concrete implementation as well as development of these infrastructures, not only the rapid changes in this area must be considered in an international context, but international cooperation must also be sought in order to develop synergies. The establishment of the e-infrastructure for research data has several advantages, including increases in the visibility and reputation of individual, participating Austrian research institutions, which is resulting in Austrian research benefitting as a whole.

Full report:
PI: phaidra.univie.ac.at/o:407513
DOI: 10.5281/zenodo.32043