



E-Specimina project funded by swissuniversities

Leveraging ORD and FAIR principles in managment and use of Natural History Collections (NHCs)

Reto Nyffeler, UZH Barbara Templ, UZH Simon Aeschbacher, UZH Thomas Grubinger, UZH Manuela Dahinden, PSC, ETHZ Jurriaan de Vos, UBasel

ORD – Open Research Data

ORD (Open Research Data) principes and practices:

- facilitate access to and reuse of research data
- promote better, more effective research
- support transparent, reproducible research findings
- foster collaboration by making it easier for researchers to share their knowledge across disciplines, legal systems, and borders, enabling creativity and innovation in the process

Swiss National Open Research Data Strategy and Action Plan (www.swissuniversities.ch)

FAIR – Findable, Accessible, Interoperable, Reusable

FAIR principles are targeted towards data-intensive science, in order to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of scientific data.

Findable: F1. (meta)data are assigned a globally unique and persistent identifier

Accessible: A1. (meta)data are retrievable by their identifier using a standardized communications protocol

Interoperable: I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

Reusable: R1. meta(data) are richly described with a plurality of accurate and relevant attributes

Wilkinson et al. 2016 (www.go-fair.org/fair-principles and force11.org/info/the-fair-data-principles)

Motivation to digitize Natural History Collections (NHCs)

Collection of Natural History Objects



Motivation (1)

- publicize digital inventory for public access
- preserve information in LTS archives
- generate research data linked to associated preserved material (RDM; research data management)

| ← → G ♡ | gbit.org/occu | rrence/search?basis_of_record=PRESERVED, | SPECIMEN8.ta | 🖈 🚺 🕼 | ti 🖬 🔘 | |
|--|---------------|--|--------------------------|---------------|----------------|--|
| | | | | | | |
| 🗶 Get data How | | | | | Q 📮 Login | |
| Cocurrencee | | SEARCH OCCURRENCES R, B14 RESULTS | | | | |
| Search all fields | ۹ أ | TABLE GALLERY MAP TAXONOMY | METRICS & DOWNLO | DAD | | |
| Simple Advance | ed | E Scientific name | Country or area | Coordinates | Month & year | |
| Occurrence status | ~ | Enlarum bronzeninum I | Seale | 28.4N 36.7W | 2021 January | |
| Licence | ~ | | | | | |
| Scientific name | × | Solarum lycopersioum L. | Australia | 34.95, 138.66 | 2023 February | |
| Solarum Inconstitute I | | Solarum (copersioum L. | Australia | 34.95, 138.6E | 2023 February | |
| Contracting Contracting | | Solatum lycopersicum L. | Australia | 34.96, 138.6E | 2023 February | |
| Basis of record | ^ | Solatum knotensioum i | Nicarana | 13.3N 86.2W | 2022 Entrolary | |
| Observation Machine characterise | | | | | | |
| Macrine coversion | | Solatum lycopersioum L. | Nicaragua | 13.3N, 86.2W | 2022 February | |
| Material sample | 247 | Solatum licopersioum L. | United States of America | 38.6N, 90.4W | 2022 July | |
| Material citation | 91 | | | | | |
| Preserved specimen | 8.814 | Solatum lycopersioum L. | United States of America | 40.9N, 73.9W | 2022 September | |
| Fossil specimen | 0 | Solarum Incoperatioum L. | Finland | 65.0N.25.4E | 2022 September | |
| Living specimen | 7.468 | | | | | |
| Occurrence | 51.287 | Solatum (copersioum L. | Brazil | 23.75, 52.6W | 2022 December | |
| CLEAR | REVERSE | Solatum lycopersicum L. | Brazil | 6.95, 41.0W | 2022 December | |
| Location | | Solatum lycopersicum L. | Brazil | 25.45, 49.2W | 2021 January | |
| A desired and the second from the second | | | | | | |

Digital Data Repository (i.e., GBIF.org)

| | nes Community About | | + 7 | |
|--|---|--|--|----|
| 2 | for the second second | | | |
| Bentley This Provide 1 | | | | |
| Record | | | | 14 |
| | | | | |
| Tern | interpreted | Original. | menories | |
| Term Basis af record | Interpreted Preserved (pecificies) | Drighall Preserved (pecimies | menarko | |
| Term Basis af record Dataset 10 | meserved specimen 2 400172424 | Drighal Preservedtpecinies 2-668172476 | menarko | |
| Term Basis af record Dataset ID Futflation code | Interpreted Preserved Specimen 2 9991728298 2 Interestat Zance | origonil Preservedipecinte 2466173406 2 | teneto COLUMN | |
| Term Basis af lecost Dataset () Putituzze code Davier Kethalise code | Interpreted Protectived (perimen) 2 4001172/24 2 Universite Zinter 2 | Dripoal Preservedspecimies 2:000122476 2: 2: 2: | economic e o economic | |
| Term Trans al record Dataset () Franktice: code Davier institution code Dataset () | intersizeted Meserved (speciate) 2 dostrify/2/8 2 interestat Salety 2 | ongosil Preservedipecinie 2-060173476 2 2 2 | monto Contention | 3 |
| Term Basis al record Dataset 0 Futilizion code Dense Ketholine code Occurrence Term | ministered Preserved (perimen 2 dobr 75,0% 2 international 2 | origoval Preservedipectines 2 des 75476 2 2 2 2 2 0 origoval | Renarks | 3 |
| Term Boss of record Dataset ID Fundation mole Dever reptication code Conservance Term Associated media | Interpreted Preserved (persitien) 3 adapt 75204 2 bitwentie Zalety 2 2 | Original Presinvedipacimies 2 des 2 2 2 2 2 2 Original Bitgus / www. digit. 2/2/ 001112/2 in gip | Remarks Exclusion Remarks Exclusion | 5 |
| Term These all record Dataset (1) Futuration code Dever antibulien code Dever antibulien code Casarrando Term Associadad media Calabogae runther | Interpreted Preserved deciment 2 addit 19224 2 streamtal Zanch 2 Interpreted 2 addit 192424 | Original PreservedSpectrade 2 4065724/76 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 | Renardo Contractional Contractional Demarks Location | |
| Term Term al Protoid Dataset 0 Institute 0 Dever antibilition occo Dever antibilition occo Dever antibilition occo Dever antibilition Term Associated Profile California Profile Dataset of D | Interpreted Preserved operation 2 deal 752/s 2 interpreted 2 interpreted 2 deal 752/s interpreted 2 deal 752/s interpreted | 0 05044 24684 22476 2 2 0 05044 0 05044 00 | Remarks Exclamation Remarks Exclama | |
| Term Tenso al Ancoral Dataset 0 Institute 0 Dataset 0 Dataset 0 Dataset 0 Dataset 0 Term Associated metho Dataset on the Documence datas | Integrated Integrated operation 3 about 1920s 2 unevental Zalen 2 Integrated | 0 (sport) Preserved Specifierie 2 466172476 2 2 2 2 2 2 2 2 2 2 2 2 2 | Terrarko Contractorio Contractorio Formatio Locana | |
| Term Terms al Record Dataset 0 Institution colo Dever set55,000 color Term Associated methe Defailingue rountier Documence 00 Documence | Integrated Preserved partition 3 data 17626 2 similar Zuine 2 integrated 2 data 17640 2 data 176400 2 data 17640 2 data 17640 2 data 17 | 00004 Pressvelationnee 2 681/2406 2 2 01000 0000 0000 0000 0000 0000 0000 0000 0000 0000 00000 0000 00000 0000 0000 0000 00000 | Renarks | |

(not yet listed by SNF as registered repository)

E-Specimina: NHCs and research data management (RDM)

- Identification: reviewing ORD and FAIR protocols relevant to NHCs
- Integration: adapting ORD and FAIR standards to fit the specific needs of NHC data management
- Application: implementing standardized data practices in publishing and sharing NHC data
- Collaboration: exchanging with stakeholder groups to ensure seamless adoption of available and newly developed standards
 - NHC managers and curators
 - Researchers to use NHC data (and material)
 - Technicians implementing NHC data holding and managing infrastructure

Practical case: study of the Tomato Domestication History

Study approach:

Analysis of contemporary and historical tomato genomes, some more than four centuries old

Study material (research material and data from NHCs and literature):

- 21 herbarium specimens from the herbaria BAS and Z+ZT
- 1 publicly available genome sequence from a very old historical specimen (En Tibi, 1558; Van Andel et al. 2022)
- 166 publicly available modern wild and cultivated tomato sequences from Latin America (Razifard et al. 2020)

Sources are NHCs digital repositories, publications and research data portals [Sequence Read Archive (SRA), or the European Nucleotide Archive (ENA)]



TDH study: challenges encountered

Accessions for study material and records from research data portals are not clearly interlinked, research to locate data is laborious.



Sixteenth-century tomatoes in Europe: who saw them, what they looked like, and where they came from

Tinde van Andel^{1,23}, Rutger A. Vos², Ewout Michels^{1,3} and Anastasia Stefanaki^{1,3} Biosystematics Group, Wageringen University and Research, Wageringen. The Netherlands Naturalis Biodwrsity Cratter, Leiden, the Netherlands Institute for Biology. Leiden University. Leiden, the Netherlands

ABSTRACT

Background. Soon after the Spanial compared of the Americas, the first transress were presented as cursines to the European line and dreve the tartistion of sixteenthicentury Dalam naturalists. Despite of their scientific interest in this New World crop, most Tensiance behavior and the strangent transmission of the architecture of the strangent science of the strangent transmission and their origin is often handlered by premoves draining boundard middlenditions and microcrobic linearies lower at Lodon, the Netherlands, triggered research on the groupshild procemance and merphological comparison to other transmission from the same Methods. Recent digitization efforts gravity facilitate research on historic bastnaid.

Methods. Recent digitization entry greatly inclutate research on historic botanical sources. Here we provide an overview of the ter memiliany atteends-century tomato specimens, early descriptions and 13 illustrations. Several were never published before, We compare on thistorical influings with recent molecular research on the chloroplast and nuclear DNA of the 'Ta Thi's specimen. Results. Our answer) shows that the calcular total total cance in an ends wider

Submitted 9 July 2021 Accepted 21 December 2021 Accepted 21 December 2021 Corresponding submit Turke van Andre Under Stander Stander Under Stander Additional Hommation and Declarations can be found on page 23 DDI 16/2712/peer1,1279 Corp. 2022 Corp. 2022 DDI 16/2712/peer1 2022 are Andre 4 al. Distributed contents CC-BY 4. Distributed contents CC-BY 4. We compare our historical indiage with recent molecular research on the chlorophot and maler DNA of the Tin This spectrum. Suce in its marce size mices and the spectrum variety of colors, shopes and sizes than previously thought, with holds imple indifactuated lowers, resume and suggement fractions. Perture Andres Muthidia gree the first description of a tomato in 1544, and the oldest specimens were collected by Uliase Aldrownal and Francess Petrallini in : 1-153, non-bly from plant green in the Tiss bounnial garden by their tacket has a Chlin. The older tomaso illustrations were made published the first image in 1333. The mass of carbon strates in contemporary manuscripts suggest boths Alexican and a Perevision ergin. The Tal's speciment solecalced by Petralinia sound 1558 and this is no the dollest cratter tomason. However, the speciment molecular research on the ancient marker and also special DNA of the Ta Tal's speciment wave and the collection of the strate strates and the stratestically does to the strate to origin. Molecular research on the order stratesed to materia and grane and geographic corrigin. Class on the 'historyic' taste and peter stratest of graded DNAs that bandle branker sought both alternative and and alter DNA of the class and a peter stratest of the stratest sought in these is the dublest the research on the duries stratest of the stratest sought in the history's tastest and peter stratest of graded TDNAs of the alternative stratest.

How In cite this article van Andel T, Vas RA, Middel E, Stefanski A. 2022, Stateenth-century transities in Europe: who saw them, what they lasked like, and advery they came from . Peor DBo/12700 June (Marcon D. 7711 June) 13700

Research publication



Specimen 'En Tibi tomato' F. Petrollini (c. 1558), L.2111092, Photo credit: Naturalis, Leiden



ID: 9041813

www.ncbi.nlm.nih.gov/sra/?term=prjna566320

International Generic Sample Number (IGSN)

An IGSN ID is a globally unique and **persistent identifier for physical samples**. These samples are described with metadata following a standard schema. An IGSN ID cannot be used for an image of a sample or for digital data.

IGSNs are developed and maintained by International Generic Sample Number (IGSN) e.V. in collaboration with DataCite.org.

https://ev.igsn.org https://datacite.org/



The application and long-term utility of sample based data is critically dependent on, availability of information (metadata) about the samples, links to other data sets derived from individual samples, and access to the samples themselves. Major problems for achieving this include incomplete documentation of samples in publications, use of ambiguous sample names, and the lack of web accessible catalogues that allows finding out about existing samples and their archiving location. Using persistent identifiers for physical samples, such as the International Geo Sample Number (IGSN), provides solutions.

The IGSN ID is a globally unique, persistent, domain-agnostic identifier for material samples. Samples can be any material from anywhere in the universe, not only from Earth. Even though IGSN had its roots in the geosciences, it really is now for all sample types, from all disciplines.

Benefits of using IGSN IDs

1. Efficient sample management

IGSNs create an efficient way to manage research samples, making it easier for researchers to keep track of them and ensuring they can be located when needed. When an IGSN is assigned to a sample, the sample can be tracked and traced from that moment until the sample is analysed and the research is published. Efficient management continues after that, enabling further research on the sample.

2. Facilitating the use of samples in research addressing long-term challenges The application and long-term utility of sample-based data are critically dependent on:

- availability of information (metadata) about the samples, such as geographical location and time of sampling
- links to other data sets derived from individual samples that are dispersed in the literature and in digital data repositories, and

access to the samples themselves

Major problems for achieving this include incomplete documentation of samples in publications, the use of ambiguous sample names, and the lack of a central catalogue that allows finding a sample's archiving location.

- Reuse of samples and specimens can also be problematic. Many samples are hidden away in archives, particularly samples held in archives following old studies. It can be difficult to get access to databases with information about the samples.
- held in archives following old studies. It can be difficult to get access to databases with information about the samples. There is a problem in research that seeks to make use of old samples because, even if you can get access to it, the metadata

TDH study: challenges encountered

A specimen may **consist of several individuals** or **part thereof**, but are not (clearly) identified during curatorial management processes and digitization.

The **genetic integrity** of distinct parts of a specimen **remains unresolved**.



Aims of our E-Specimina project and next steps

- Bring stakeholder groups together for discussing solutions and define interfaces among them
- Implement protocols and tools to apply ORD and FAIR principles with NHCs
- One-day workshop among representatives of stakeholder groups on Thursday, May 16th, 2024 (possibly Bern)
- Practical class in research data management (broadly defined and applied), with invited contributors, during the Swiss Systematics Day in November 2024 (Thursday, not yet confirmed; Zürich)

Questionnaire: request for your input and feed-back

We request **your opinion and fee-back** on the subject of ORD and FAIR principles applied to NHC data via our **questionnaire**:

www.uzh.ch/zi/cl/umfragen/index.php/368129

| | Resume later |
|----------------------------------|----------------------------------|
| | |
| | |
| | |
| | |
| You and your experience with ORD | |
| | |
| | You and your experience with ORD |

O I am involved in generating and hosting specimen data [i.e., I am a collection curator]

🔘 I lead a collection-holding institution that generates and hosts specimen data [i.e., I decide on strategic priorities]

🔘 I am using comparative biological data for my primary or applied research projects [i.e., I am a scientist]

 \bigcirc I am involved in implementing data repositories to hold specimen or research data [i.e., I am a data scientist]

No answer



Information resource and project collaborators

www.e-specimina.uzh.ch

| 😂 e-specimina.uzh.ch/en.html | | @ ☆ |
|--|----------------|----------|
| University of Zurich ^m E-Specimina | Home Contact | Q Search |
| Project overview People Events Research Publications | | |
| N. | | |
| | | |
| I-Checimina | | |
| Lopectructure | | |
| HERBARIA - NEEDS AND SOLUTIONS | | |
| E Constanting | | |
| E-Specimina | | |

Objectives

The overall aim of E-Specimina is to support researchers engaging in and developing ORD practices with and for the research committy using natural history collections and to assist these researchers in becoming Ope Science leaders in their field(s).

E-Specimina will address a specific gap in knowledge with an innovative approach; so far most research projects on biodiversity that rely on comparative datasets of selected organisms do not implement protocols to integrate the data compiled for the different repositories and, hence, the challenges to publish that information conform to ORD principles is not fulfilled. We seek to address this gap in the Open Science agenda with a needs analysis that brings together and in exchange researchers from different disciplines as well as repository managers. Please contact us outside of the lecture hall, at our information table!

Reto Nyffeler, UZH Barbara Templ, UZH Thomas Grubinger, UZH Jurriaan de Vos, U Basel Manuela Dahinden, PSC, ETHZ Simon Aeschbacher, UZH