

E-Specimina project funded by swissuniversities

Leveraging ORD and FAIR principles in management 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) principles 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**

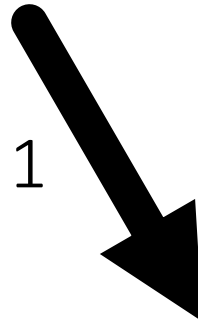
Interoperable: I1. (meta)data use a **formal, accessible, shared, and broadly applicable language** for knowledge representation

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)

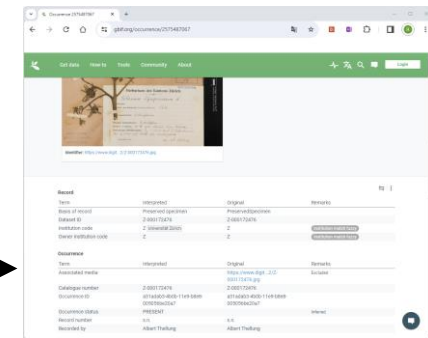
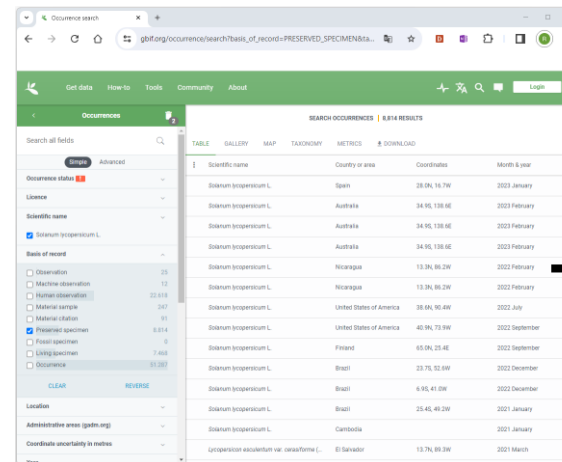
Collection of Natural History Objects



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

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)



(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

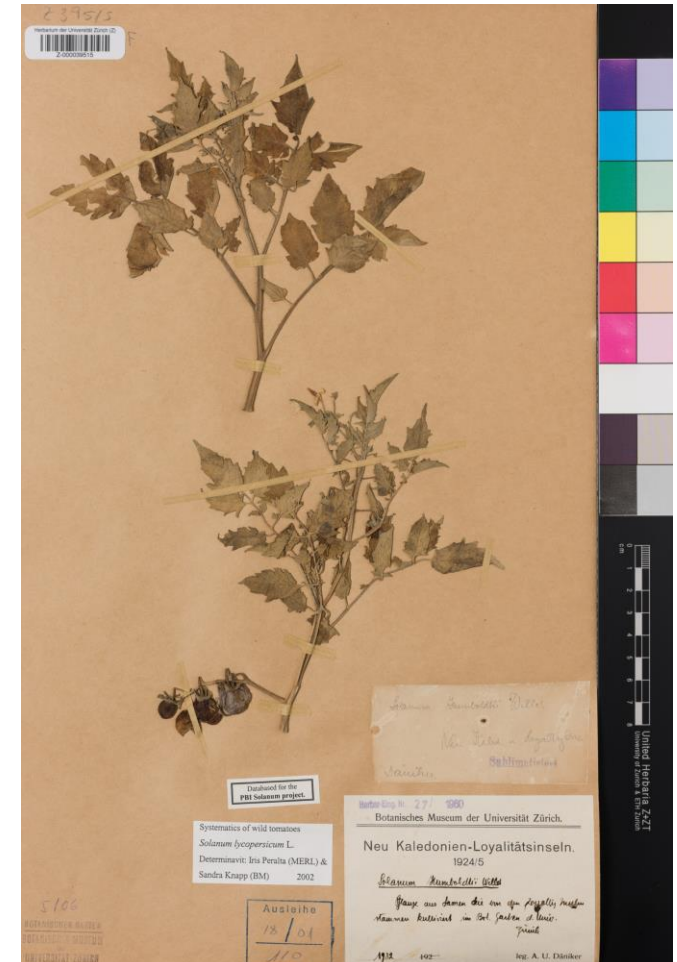
Study approach:

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.

PeerJ

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

Tinde van Andel^{1,2}, Rutger A. Vos¹, Ewout Michels^{1,3} and Anastasia Stefanaki²

¹Bioinformatics Group, Wageningen University and Research, Wageningen, The Netherlands
²Naturalis Biodiversity Center, Leiden, the Netherlands
³Institute for Biology, Leiden University, Leiden, the Netherlands

ABSTRACT

Background: Soon after the Spanish conquest of the Americas, the first tomatoes were presented as curiosities to the European elite and drew the attention of sixteenth-century Italian naturalists. Despite of their scientific interest in this New World crop, most Renaissance botanists did not specify where these 'golden apples' or 'pomii d'oro' came from. The debate on the first European tomatoes and their origin is often hindered by erroneous dating, botanical misidentifications and inaccessible historical sources. The discovery of a tomato specimen in the sixteenth-century 'En Tibi herbarium' kept at Leiden, the Netherlands, triggered research on its geographical provenance and morphological comparison to other tomato specimens and illustrations from the same time period.

Methods: Recent digitization efforts greatly facilitate research on historic botanical sources. Here we provide an overview of the ten remaining sixteenth-century tomato specimens, early descriptions and 13 illustrations. Several were never published before, revealing what these tomatoes looked like, who saw them, and where they came from. We compare our historical findings with recent molecular research on the chloroplast and nuclear DNA of the 'En Tibi' specimen.

Results: Our survey shows that the earliest tomatoes in Europe came in a much wider variety of colors, shapes and sizes than previously thought, with both simple and faciated flowers, round and segmented fruits. Pietro Andrea Matthioli gave the first description of a tomato in 1544, and the oldest specimens were collected by Ulisse Aldrovandi and Francesco Petrolini in c. 1558, possibly from plants grown in the Pisa botanical garden by their teacher Luca Ghini. The oldest tomato illustrations were made in Germany and Switzerland in the early 1590s, but the Flemish Rembert Dodoens published the first image in 1553. The names of early tomatoes in contemporary manuscripts suggest both a Mexican and a Peruvian origin. The 'En Tibi' specimens were collected by Petrolini around 1558 and thus is not the oldest extant tomato. Recent molecular research on the ancient nuclear and chloroplast DNA of the 'En Tibi' specimen clearly shows that it was a fully domesticated tomato, and genetically close to three Mexican landraces and two Peruvian specimens that probably also had a Mesoamerican origin. Molecular research on the other sixteenth-century tomato specimens may reveal other patterns of genetic similarity, past selection processes, and geographic origins. Clues on the 'historic' taste and pest resistance of the sixteenth-century tomatoes will be difficult to predict from their degraded DNA, but should be rather sought in those landraces in Central and South America that are genetically close to them. The

Submitted 9 July 2021
Accepted 22 December 2021
Published 17 January 2022

Corresponding author: Tinde van Andel, tinde.vanandel@wur.nl

Academic editor: Vasilisa Galenkova

Additional Information and Declarations can be found on page 23

DOI: 10.7717/peerj.12799

© Copyright 2022 van Andel et al. Distributed under Creative Commons CC-BY 4.0

OPEN ACCESS

[How to cite this article: van Andel T, Vos RA, Michels E, Stefanaki A \(2022\) Sixteenth-century tomatoes in Europe: who saw them, what they looked like, and where they came from. PeerJ 10\(1\):12799. <https://doi.org/10.7717/peerj.12799>](https://doi.org/10.7717/peerj.12799)

Research publication



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

NIH National Library of Medicine
National Center for Biotechnology Information

BioSample BioSample Advanced

Full -

Plant sample from *Solanum lycopersicum*

Identifiers BioSample: SAMN12785837; Sample name: En-Tibi; SRA: SRR5407108

Organism *Solanum lycopersicum* (tomato)
cellular organisms; Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophytes; Mesangiospermae; rosids; eudicots; Gunnerales; Pentapetalae; asterids; lamiales; Solanales; Solanaceae; Solanaceae; Solanum; Solanum

Package Plant_version.1.0

Attributes isolate obtained from the En Tibi herbarium specimen
cultivar possibly oerasilforme
development stage Fruiting stage
geographic location Netherlands
tissue leaf
biomaterial provider Naturalis Biodiversity Center
collected by Anonymous; En Tibi herbarium
specimen voucher L.2111092

Description This sample is from a herbarium specimen that was produced in Italy, to the best estimates, 1558.

BioProject PRJNA566320
Retrieve all samples from this project

Submission Naturalis Biodiversity Center, Rutger Vos; 2019-09-19

Accession: SAMN12785837 ID: 12785837
BioProject SRA

Record of specimen and genome sequence

NIH National Library of Medicine
National Center for Biotechnology Information

SRA SRA prjna566320
Create alert Advanced

Full -

SRX6871077: WGS of *Solanum lycopersicum*: adult plant leaf
1 ILLUMINA (Illumina HiSeq 2000) run: 246.9M spots, 59.8G bases, 23.6Gb downloads

Design: Extractions done in ancient DNA lab
Submitted by: Naturalis Biodiversity Center

Study: WGS analysis of the En Tibi tomato specimen
PRJNA566320 • SRP222460 • All experiments • All runs
show Abstract

Sample: This sample is from a herbarium specimen that was produced in Italy, to the best estimates, 1558.
SAMN12785837 • SRR5407108 • All experiments • All runs
Organism: *Solanum lycopersicum*

Library:
Name: En-Tibi
Instrument: Illumina HiSeq 2000
Strategy: WGS
Source: GENOMIC
Selection: RANDOM
Layout: PAIRED

Runs: 1 run, 246.9M spots, 59.8G bases, 23.6Gb

Run	# of Spots	# of Bases	Size	Published
SRR10143152	246,940,261	59.8G	23.6Gb	2020-10-01

ID: 9041813

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

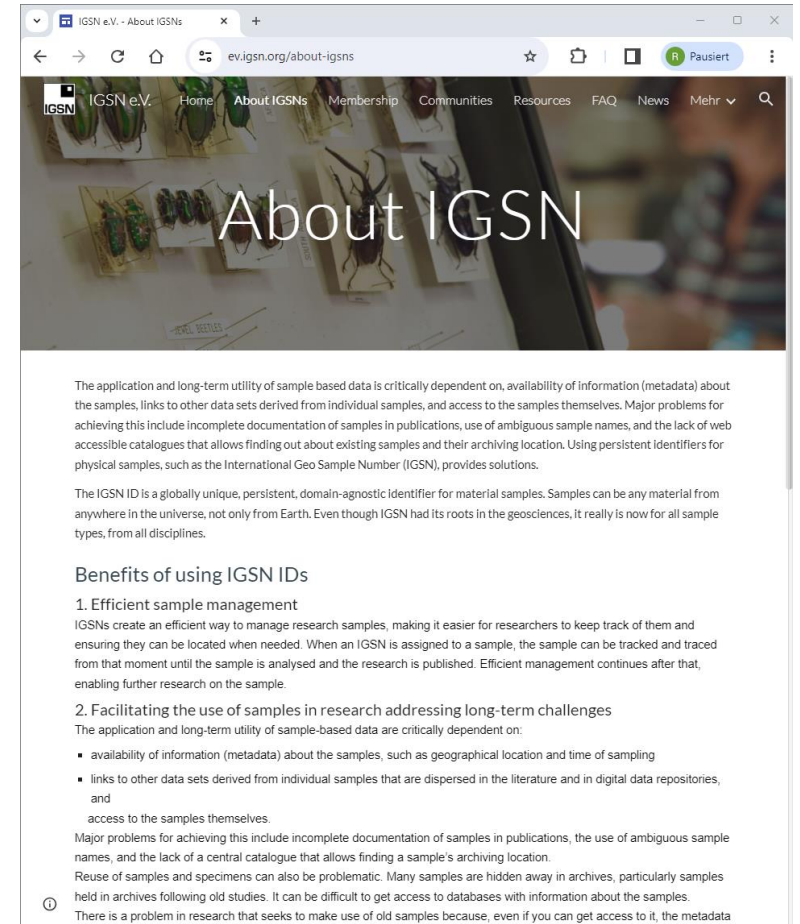
International Generic Sample Number (IGSN)

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/>



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**.



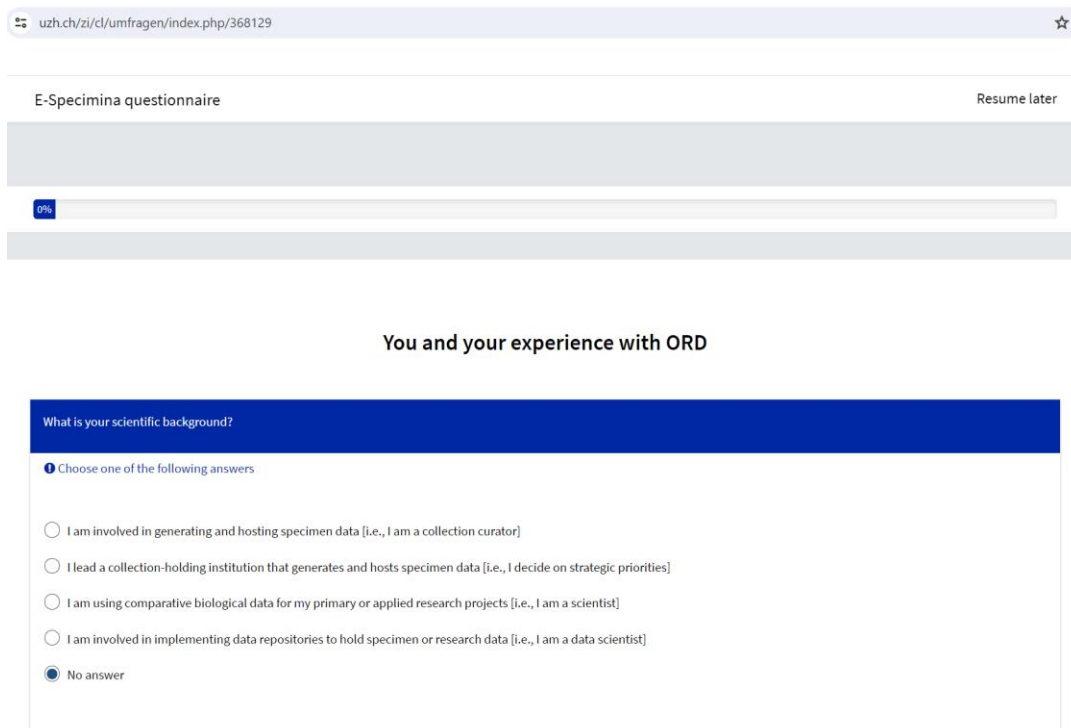
Aims of our E-Specimina project and next steps

- 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



The screenshot shows a web browser window with the URL uzh.ch/zi/cl/umfragen/index.php/368129. The page title is "E-Specimina questionnaire" and it includes a "Resume later" link. A progress bar indicates 0% completion. The main heading is "You and your experience with ORD". The first question is "What is your scientific background?" with the instruction "Choose one of the following answers". The options are:

- 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]
- 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

Objectives

The overall aim of E-Specimina is to support researchers engaging in and developing ORD practices with and for the research community using natural history collections and to assist these researchers in becoming Open 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