

Scanning and Data Extraction

from Crop Collecting Mission Documents

Hannes Gaisberger, Imke Thormann, Milko Skofic, Lorenzo Stabile, Tom Hazekamp, Aixa Del Greco, Elizabeth Arnaud
Bioversity International

Bioversity supported more than 560 crop collecting missions since 1976, which harvested a total of 221.077 germplasm samples that were subject to severe and acute threats of genetic erosion, documenting 4300 species in 137 countries, safely stored in genebanks to remain available to users worldwide.

The project

The goal of an ongoing project, “Collective Action for the Rehabilitation of Global Public Goods in the CGIAR Genetic Resources System” phase 2 (GPG2), funded by the World Bank, is to ensure the quality, security, accessibility, and sustainability of the public crop collections, including information systems.

The story

Part of the accessions stored in the CGIAR genebanks was collected during Bioversity supported collecting missions. Information on these missions is accessible through the System-wide Information Network for Genetic Resources



(SINGER, <http://www.singer.cgiar.org/>) that contains summary information on the species collected and on where the samples have been sent for long-term conservation.

Valuable additional information, called “passport data” (such as site information, collecting source, etc.), is recorded in mission reports and collecting forms in paper format, which restricts access to these sources.

The outcome

Accessibility and usability of crop collections requires that detailed data is available to the users. As the information is currently at species level only, passport data is being extracted and transferred from collecting mission documents to a sample level database.

Activities include:

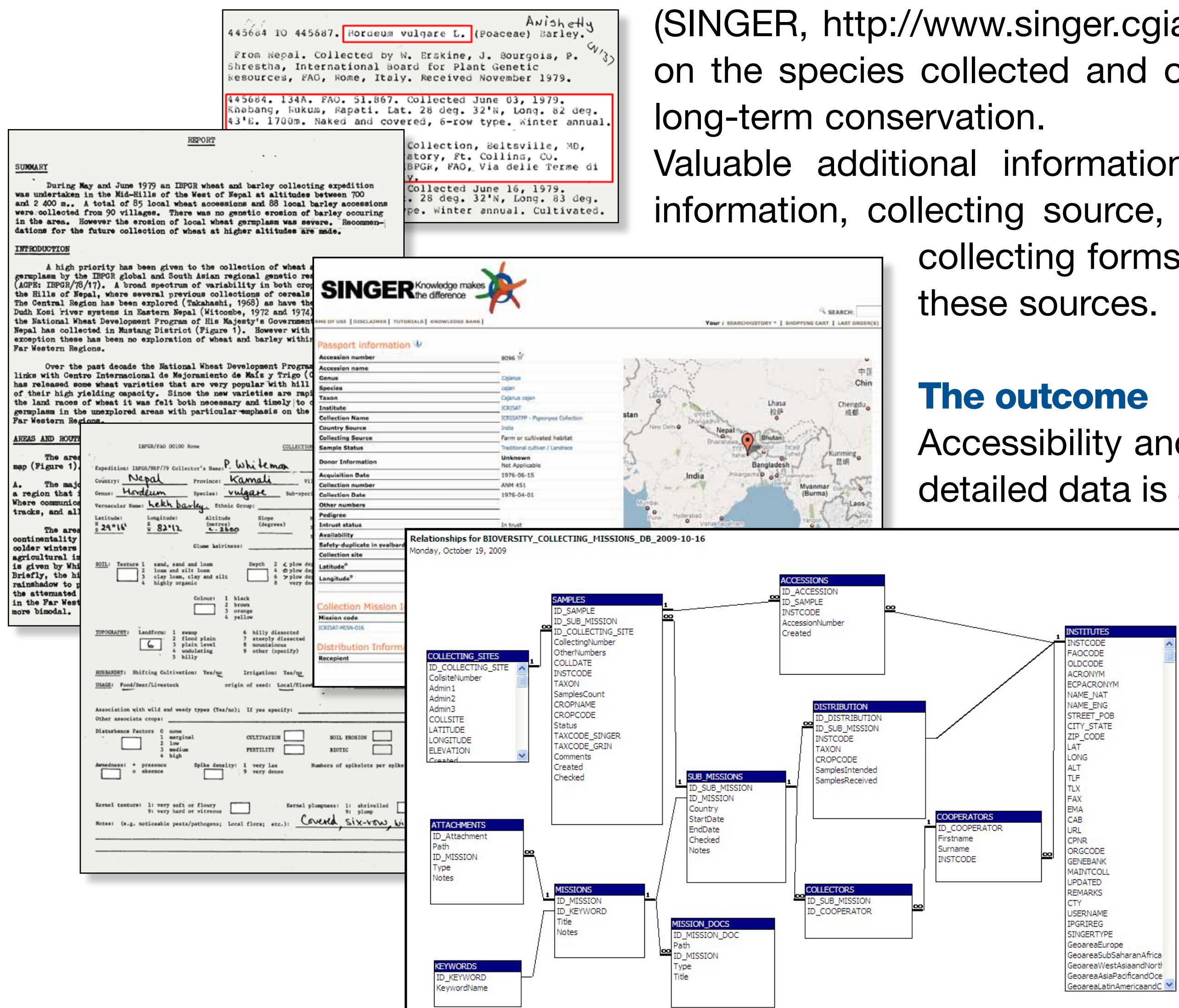
- Scanning of Bioversity’s collecting mission reports, collecting forms, and related documents;
- ⇒ *Mission documents of 109 collecting missions scanned*
- Storing of scanned reports into a

safe document repository with metadata;

⇒ *pdf-repository of 109 mission reports created*

- Extraction of passport information and input on sample level data in the Collecting Mission Database;
- ⇒ *Passport Information of 5400 samples (from 21 collecting missions) extracted and integrated*
- Integration of corresponding records in SINGER.

The new data extracted from the mission documents will be used to develop a method to assess gaps in diversity conserved due to loss of samples.



The screenshot displays the SINGER database interface. On the left, there's a 'SUMMARY' section with text about a 1979 wheat and barley collecting expedition in Nepal. Below it is a 'PASSPORT INFORMATION' section with fields for accession number, name, species, and collection date. A map shows the location in Nepal. On the right, a 'Data Model' diagram shows relationships between tables like 'COLLECTING_SITES', 'SUB_MISSIONS', 'ACCESSIONS', 'DISTRIBUTION', 'MISSIONS', 'MISSION_DOCS', 'ATTACHMENTS', and 'KEYWORDS'. The 'ACCESSIONS' table is highlighted.