GRDC
GRAINS RESEARCH & DEVELOPMENT CORPORATION

Early Insights from the DataHarvest Project

Erin Elstermann

DataHarvest Data Management Officer

Curtin University

Outline

1. Research Data Management (RDM) in the Grains Industry
2. Introduction to DataHarvest
3. Activities and Approach
    - Developed resources
4. Early Insights from Support Services Centre
5. Summary

Why data matters: complexity of grains research

- Effective RDM in agriculture often dealing with diverse datasets

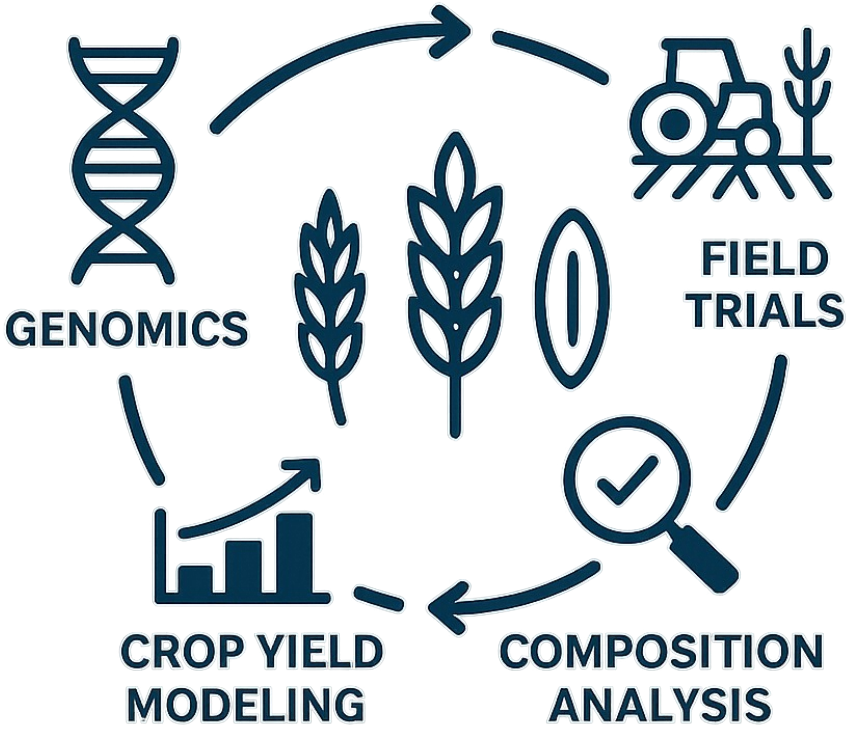
- Well executed RDM supports:

✓ reproducibility

✓ collaboration

✓ translation to industry outcomes

- Challenges in Australian grain industry include inconsistent practices, varying data literacy, limited organisational support



DataHarvest

A partnership to unlock the value of RD&E data for the Australian grains industry.

DataHarvest is an investment of the Grains Research and Development Corporation (GRDC). It is led by Curtin University through the Centre for Crop and Disease Management (CCDM).

GRDC code: CUR2401-001BGX

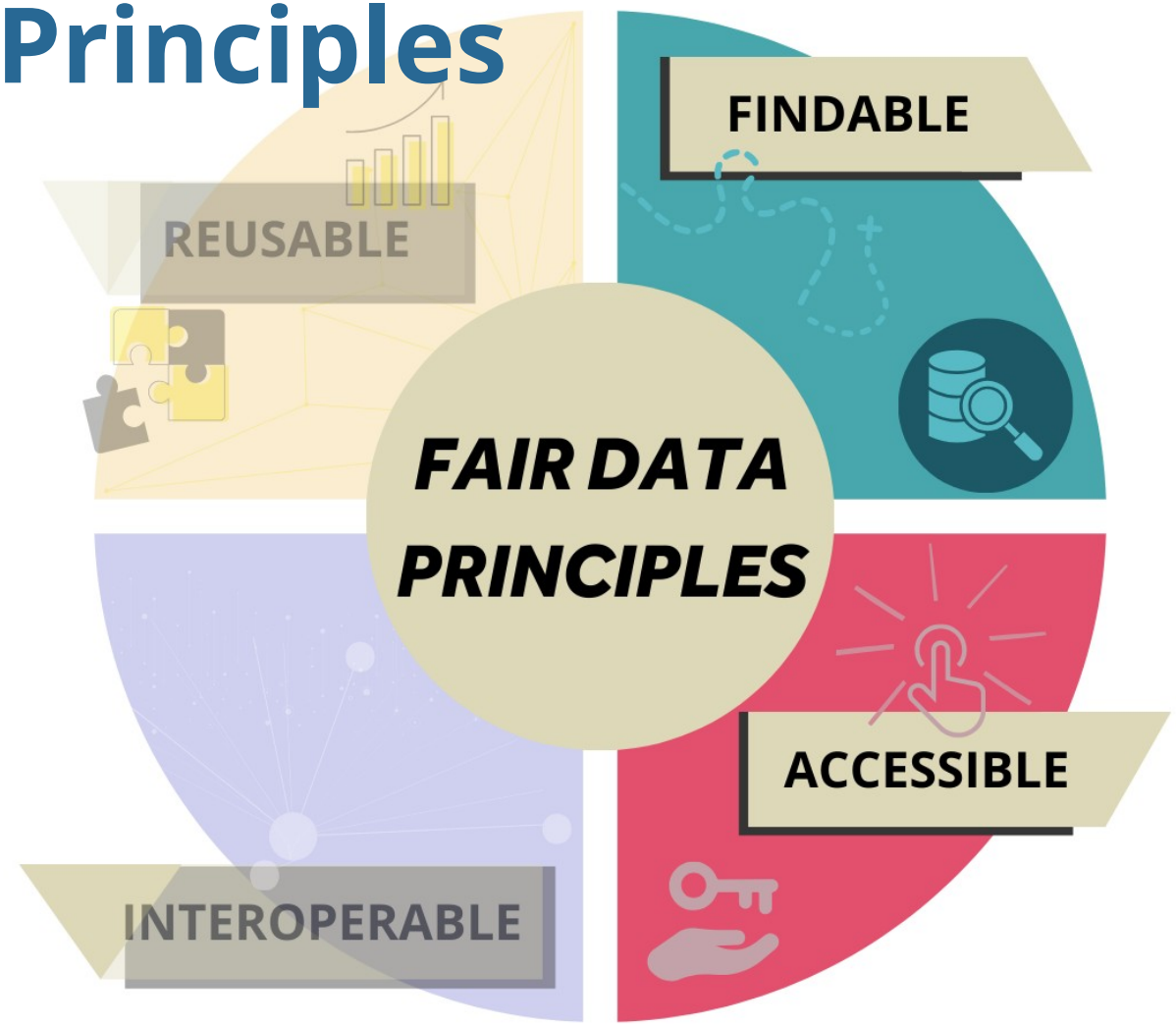
The purpose of DataHarvest

Develop capacity and capability in research partners to deliver against GRDC's RD&E data management policy.

This will be achieved by:

- Developing data management training materials suitable for GRDC's research partners
- Delivering tailored F2F workshops, courses and webinars
- Facilitating and providing opportunities for researchers to learn FAIR skills by doing

DataHarvest and the FAIR Principles



Increasing Findability through:

- Archiving data into an Approved repository
  - ✓ Metadata standards
  - ✓ Persistent Identifiers
  - ✓ Using the GRDC Data Catalogue

Increasing Accessibility through:

- Information on Data Access conditions
  - ✓ Online resources

Centre for Crop and Disease Management

GRDC

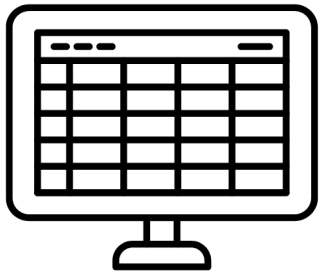
Activities and Approach

1. National workshops: face-to-face hands-on training

2. Resources: Step-by-step guides to complete GRDC RDM tasks
    - Data Management Plan template
    - Metadata standard guidelines

3. Support Services Centre: Dedicated ongoing support

RD&E Data Management Activities



During DMP completion:

- Dataset name and descriptions
- Access Conditions
- Identifying an Approved Repository for data storage

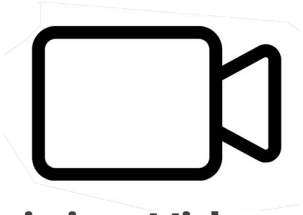
During data and/or metadata upload to repositories:

- Persistent Identifiers (ORCIDs, ROR etc.)
- Dataset title and description
- Keywords and subjects
- Data Reuse Licencing

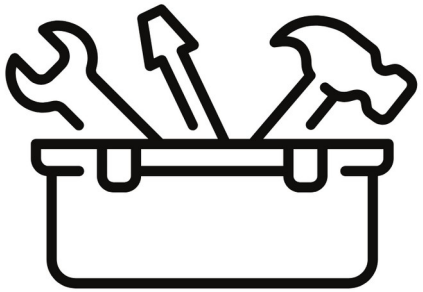
What have we developed in response?

Documents

- Step-by-step guide for:
  - GRDC DMP
  - GRDC metadata fields
  - Uploading data & metadata to Zenodo
- FAQ for GRDC partners
- Flyers on FAIR and Open Science



Training Videos



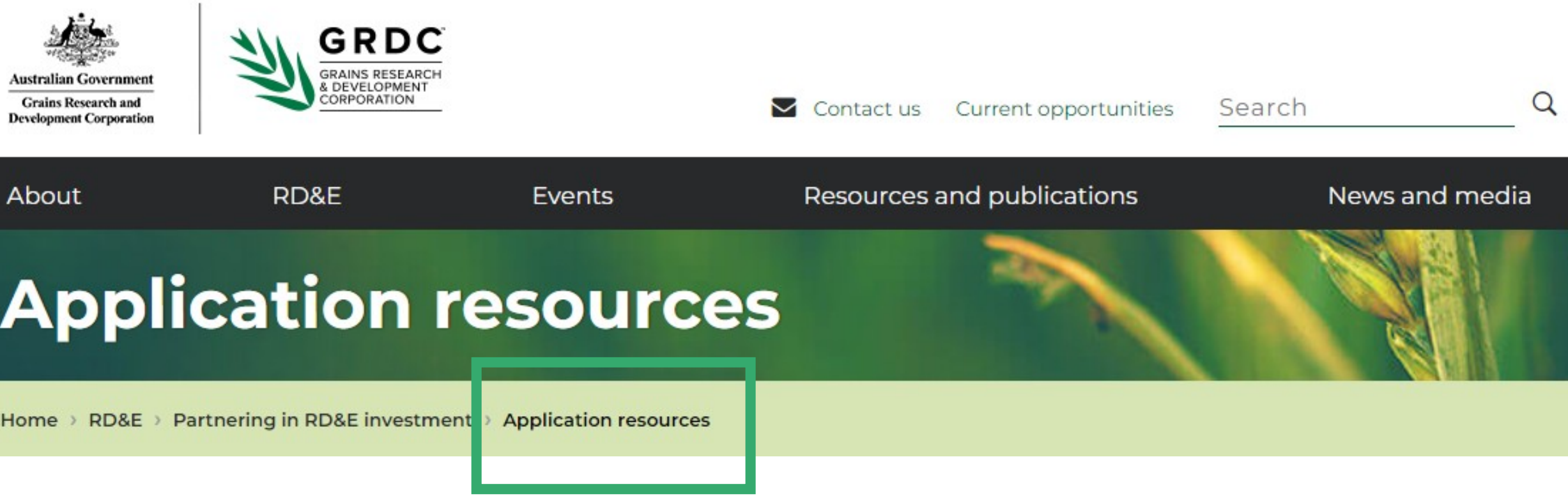
Together, these resources form a toolkit that makes RDM more accessible!

- Completing GRDC DMP
- Uploading data/metadata to Zenodo
- Using GRDC metadata forms
- Persistent Identifiers (PIDs)
- Controlled Vocabularies (AGROVOC)
- Data Access Conditions
- Data Licensing &

Centre for Crop and Disease Management

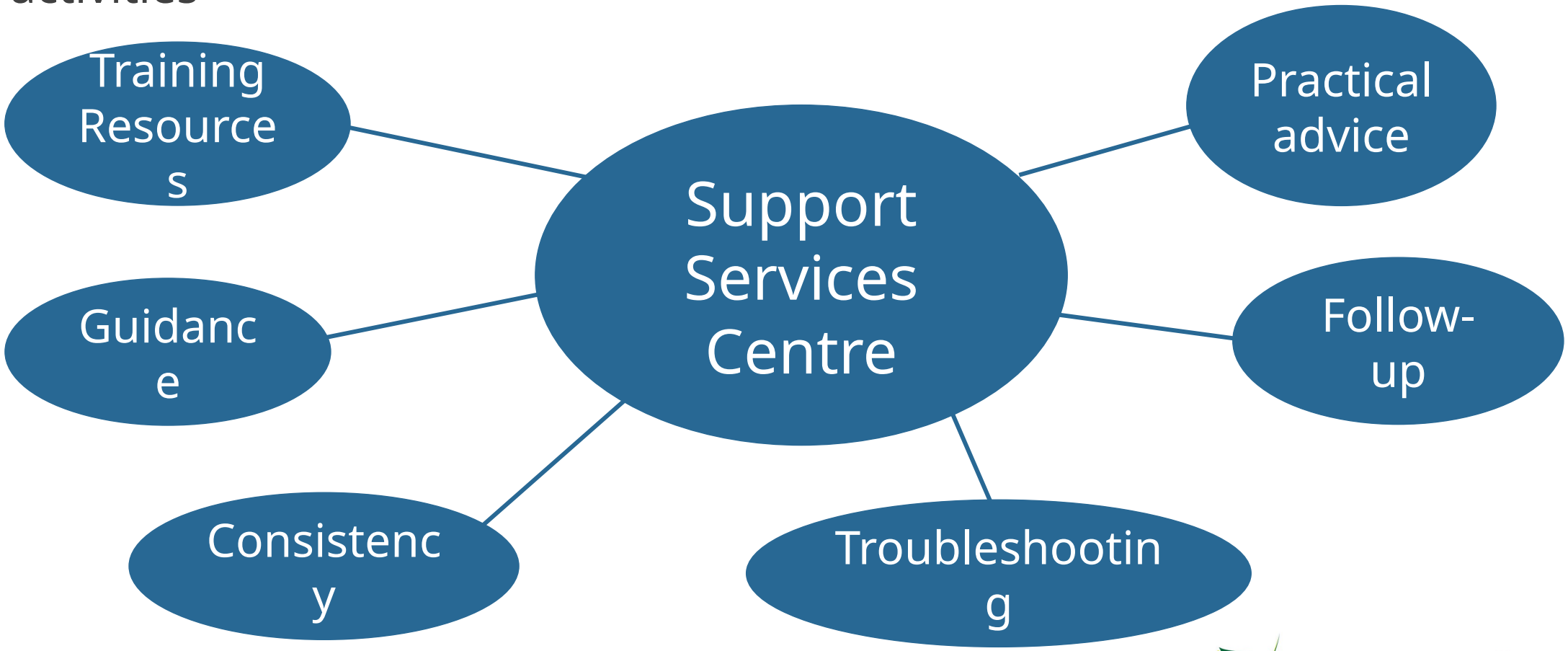
GRDC

Resources are available on the GRDC Website



Dedicated Support

- Central point for GRDC partners to request support for RDM activities



Improvements in DMP standards

- DMP standards are evolving
- Step-by-step guide developed in July 2025 to hand out during face-to-face workshops
- The handout includes various examples

DMP Example for Partners

Dataset Name	Dataset description	Storage location	Data format	Access conditions
EXAMPLE 1: Grower survey results from the West and East Australian regions in 2024-2025	This dataset will contain survey responses from approximately 80 growers in the Western and Eastern Australia regions, collected during the 2024-2025 growing season. The survey is focused on pest control practices. It will capture information on the types of treatments used, the timing and frequency of their application, the quantities applied, and growers' perceptions of their effectiveness. In addition to pest management data, the survey is also aimed at gathering information on broader farm management practices that may influence or relate to pest control decisions.	Zenodo	MS- WORD	Restricted. A de-identified version of this document can be made available upon request. Please refer to the ethics approval for this dataset for more information.
EXAMPLE 2: Assessment of heat tolerance in wheat in Western Australian systems from 2023-2025	Field wheat data at three locations at two times of sowing, both in season and post-harvest. Weather data collected at each site. Wheat phenotype traits, genotype based on 90K SNP Chip and site environmental data collected and at 2 different time of sowing (TOS1 and TOS2) from 2023 to 2025.	SAS data commons Zenodo	CSV	Conditional, embargoed for 12 months for the purposes of publication.
EXAMPLE 3: Herbicide resistance inheritance	Segregation of individuals from crosses after treatment with herbicides, including dose response data.	Figshare	XLSX, DOCX, JPEG, TXT	Open.

DataHarvest Support Services centre helps with Metadata Standards

Areas of Focus

- Dataset titles
- Descriptions
- Keywords
- Persistent Identifiers (PIDs)
  - Identifying creators and contributors of datasets
- Data Access Conditions and Data Reuse Licences

Support Solutions

- On-demand help with metadata entry
  - Teams/Zoom
  - Email
  - Phone
- Step-by-step guides
- Training videos for Zenodo upload process
- Clarifying metadata standards

Summary

- DataHarvest investment supports RDM practices for GRDC investments

- Online resources and hands-on support have shown improvement in metadata standards

Acknowledgements

Fatima Naim, CCDM
John Brown, Curtin University
Melanie Dixon, CCDM
Washington Gapare, GRDC
Jane Gibberd, Curtin University
Amy Taylor, Curtin University
Megan Jones, CCDM
Mark Gibberd, CCDM
Lisa Smith, CCDM
Janice Chan, Curtin University

DataHarvest

THANK YOU

Erin Elstermann

Data Management Officer, CCDM

erin.elstermann@curtin.edu.au

GRDC
GRAINS RESEARCH & DEVELOPMENT CORPORATION