

Advantages of dictionary-based software

- 1. Data name recognition
- 2. Alternative data name recognition (aliases)
- 3. Type and range validation
- 4. Referential integrity

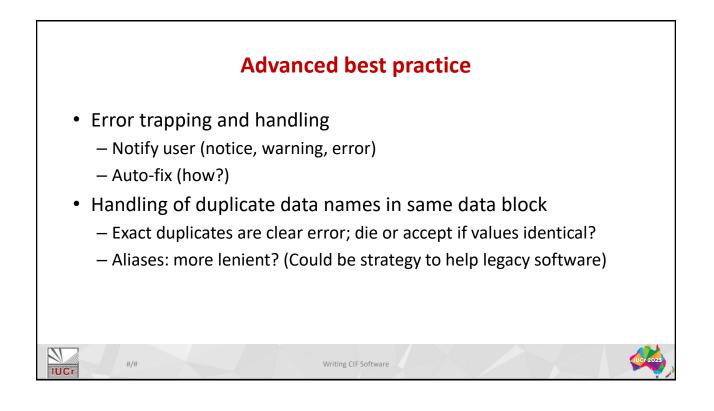
#/#

IUCr

5. Application of dREL methods

Best practice Validate against appropriate dictionary version (_audit_conform.dict_name, _audit_conform.dict_version, _audit_conform.dict_location) Consult list of registered prefixes if unexpected data names found (https://www.iucr.org/cgi-bin/cifreserve.pl) Check _audit.schema

Writing CIF Software



Things to take particular care about

- Ambiguity of data values that look like numbers.
- How to handle quoted numeric strings (e.g. '12')?
- Quoting styles (single vs double quotes, triple quotes, semicolons).
- Handling of 'scalar' data [*e.g.* adopt the CIF API approach of storing these as single-packet loops?

- But how to present them as unlooped?

- What to do if a (normally) multi-packet loop only has one packet?].

Writing CIF Software

• Long line folding.

#/#

IUC

Handling CIF2 and CIF1 syntax

- NB The presence of 'dotted' data names (<u>cell.volume</u>) is not a robust indicator that a file is CIF2 (likewise all-underscore data names do not guarantee that it is CIF1).
- Initial 'magic' string #\#CIF_2.0 is mandatory for CIF2.
- Handling different character encodings (ASCII vs Unicode).
- Conversion between different quoting mechanisms.
- Semantic nuances considered in detail in Vaitkus, A., Merkys, A. & Gražulis, S. (2021). Validation of the Crystallography Open Database using the Crystallographic Information Framework. J. Appl. Cryst. **54**, 661–672.

