GBIF
www.gbif.org

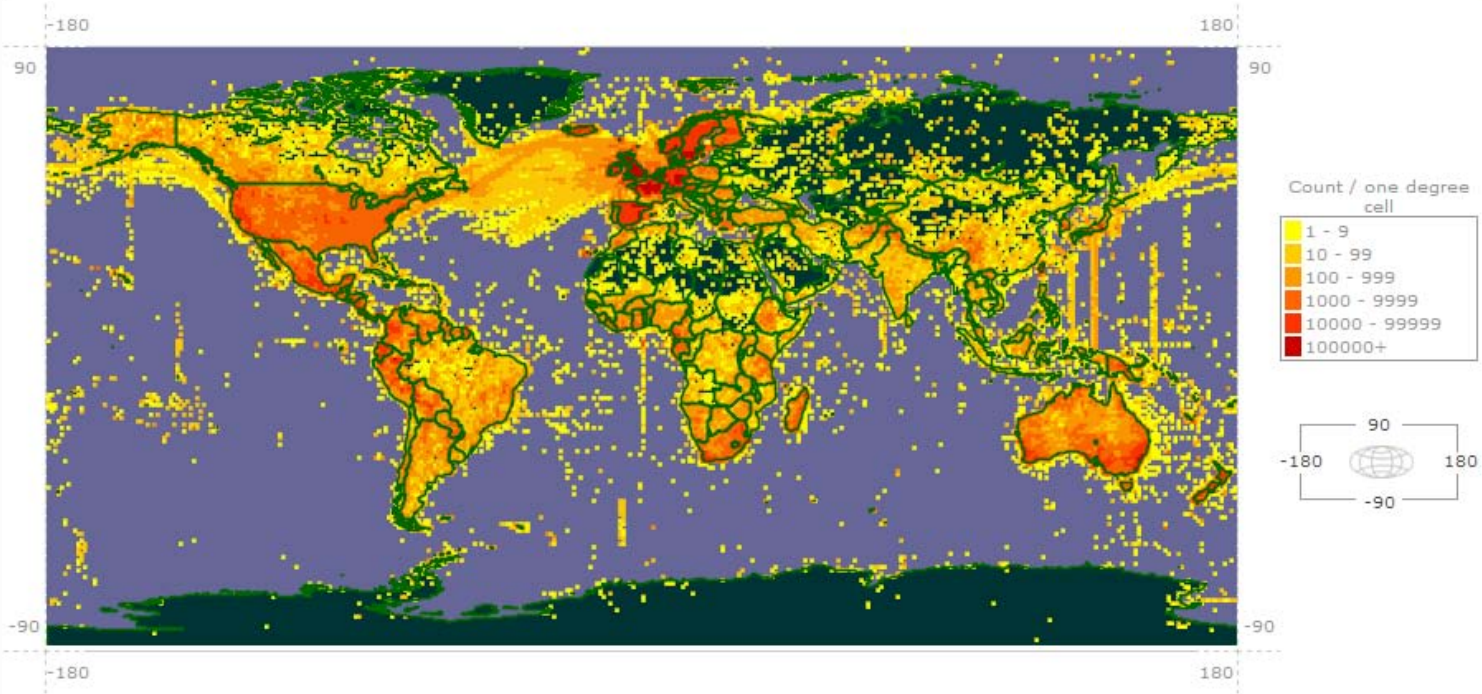
Mobilisation of botanical data and information to help build the WFO and support GSPC goals

Donald Hobern, GBIF Executive Secretary, dhobern@gbif.org
Global Biodiversity Information Facility (GBIF)

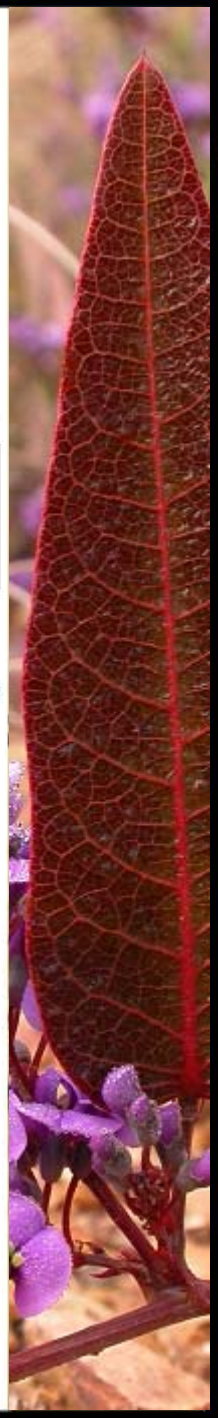
Hyderabad, CBD COP 11, 11 October 2012

Hardenbergia violacea (Schneev.) Stearn, Black Mountain, Canberra, ACT, 22 August 2011

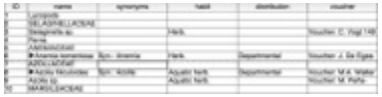
Existing occurrence data in GBIF



- Significant occurrence data already published
  - 76,319,933 records of Plantae with coordinates (October 2012)
  - Significant existing herbarium data – high digitisation levels
  - Many national vegetation data sets
- Identified as a priority focus for GBIF in 2013 and beyond
  - GBIF Governing Board meeting 2012
  - During 2013 explore steps for developing consistent and comprehensive plant occurrence data for all regions



GBIF recommended publishing model



Darwin Core spreadsheet

+

Metadata document

+

Creative Commons licence

→

Darwin Core Archive

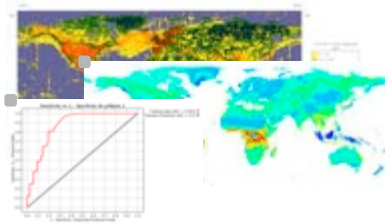
+

GBIF Regístry

+

Digital Object Identifier

→



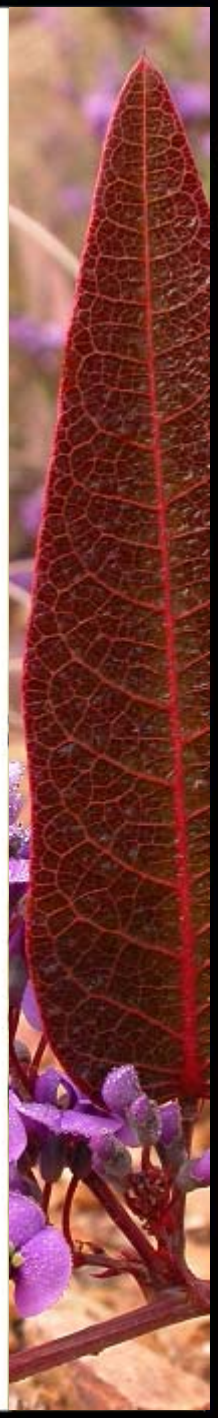
Discoverable, reusable data

+

Data Paper

+

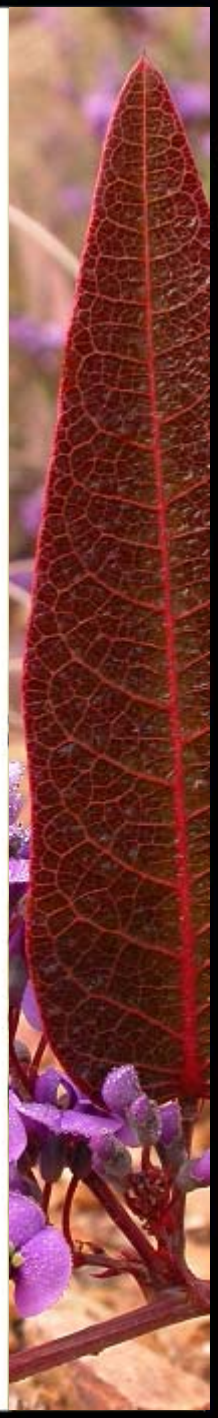
Data usage



GBIF
www.gbif.org

Darwin Core Archive (DwC-A)

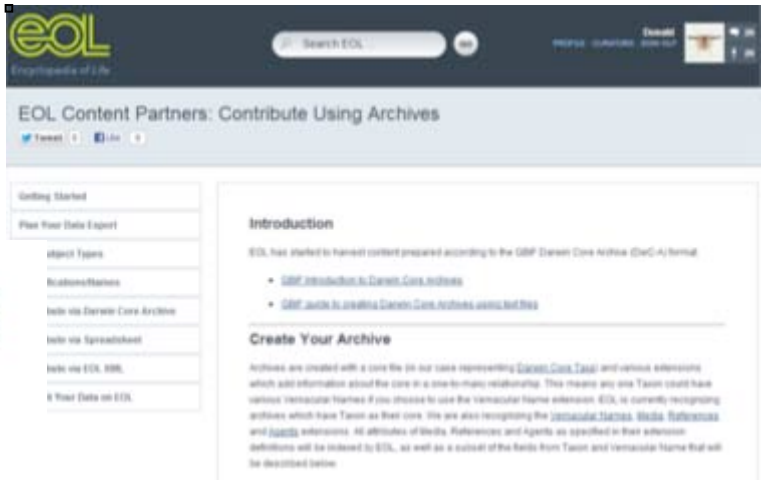
- Simple data publishing model
  - Data as one or more spreadsheets (comma-separated values)
  - Metadata describing data set
  - Licensing information
  - All in a single ZIP file
- Easily published
  - Upload to a web site or repository
  - Metadata extracted for discovery
  - Possibility to version data sets
- Supports citation through DOIs
  - Possibility to track usage and impact
  - Derive data paper
- **Free and open access is critical**
  - Ensures data can be reused in research and to develop new products
  - Responsibility of GBIF and others to support and promote good citation practices

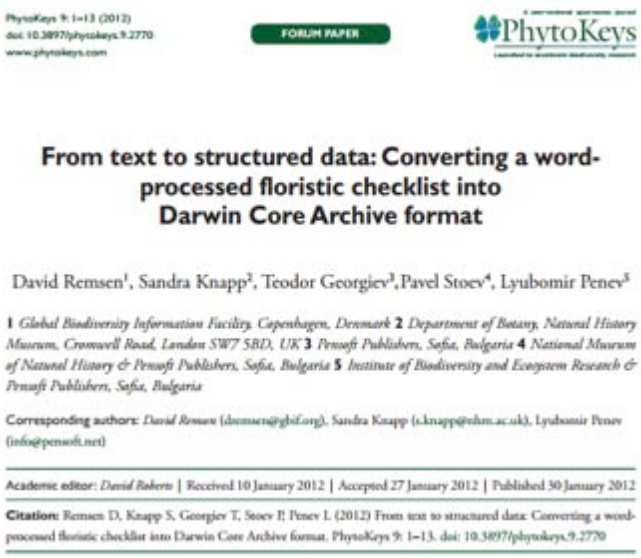


GBIF
www.gbif.org

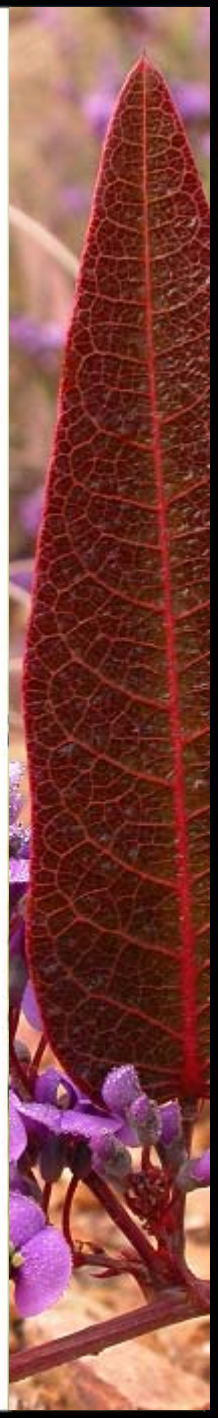
DwC-A for other data types

- Same model can be used to publish any structured data set
  - Species checklists
  - Species fact sheets
  - Trait databases
  - Etc.





- Metadata can specify:
  - Data types
  - Language
  - Geographic scope
  - Taxonomic scope
  - Origins of data set
  - Etc.



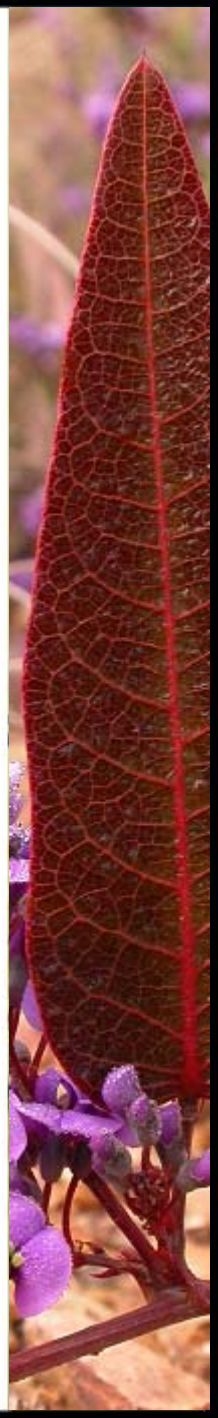
GBIF
www.gbif.org

Using DwC-A to mobilise flora data sets

- Preferred option: full access to data
  - All data for each species shared as spreadsheet under CC-BY licence
  - All content indexed and visible through WFO portal
  - All content available for use in national, regional and thematic portals and other applications, including portals in other languages
  - Focus on standardising most essential data elements
  - Process can transport non-standard elements for presentation to users

- Minimal option: access to links
  - Only species names and URLs, page references or other links under CC-BY licence
  - Users may be directed from WFO portal or elsewhere to content

- Exposing digitised literature (Biodiversity Heritage Library)
  - Develop annotation tools to locate species treatments in scanned literature
  - Expose species treatments via minimal DwC-A of links to literature pages
  - Annotation tools can be enhanced to support mark-up of text content into more structured forms



GBIF
www.gbif.org

A vision for a World Flora Online

- WFO should deliver:
  - Comprehensive set of reusable flora data sets
  - A WFO portal giving unified access to all data
  - A network of regional / national / thematic Flora Online portals
  - Tools and information products for specific user groups
- Free and open access data sets can support all of these

GBIF
www.gbif.org