Crowd-solving the file format problem

Paul Wheatley

SPRUCE Project Manager

University of Leeds

Twitter: @prwheatley

http://openplanetsfoundation.org/blogs/paul

- Understand the problem
- Examine the data
- Work out requirements
- Find solutions
  - Information, experience, research, best practice
  - Software tools
- Implement
- Evaluate

- **All of these tasks can be aided and enhanced by collaborative working and result sharing**

The collaborative initiatives that can help

- Atlas of Digital Damages
- OPF Format Corpus
- Datasets, Issues and Solutions
- Learning from practitioners needs - and then meeting them (poster)
- Format Risks
- Research challenges
- Tools

http://wiki.opf-labs.org/display/KB/DPC+Bring+out+your+dead+%28files%29

- Which files are useful to share?
    - Files, any files
    - Unusual formats
    - Files that exhibit specific characteristics
    - Broken, obsolete, or otherwise dead files
    - One constraint: must be CC0

- Why is this useful?
    - Helps others understand and address your challenges
    - Supports file format identification development
    - Supports broader tool development and testing

Collaboration and crowd sourcing best practice

- Don't reinvent the wheel
- Pool information, don't spread it around
- Use existing infrastructure, don't start from scratch
- Neutral locations are great
- Share, be open, share not so good results as well as good results
- Love CC

UNIVERSITY OF LEEDS

bit.ly/spruce-collaborate

Thanks for listening!

Paul Wheatley

SPRUCE Project Manager

University of Leeds

Twitter: @prwheatley

Email: p.r.wheatley@leeds.ac.uk

http://openplanetsfoundation.org/blogs/paul