



File and Folder Structures

Garrett Speed and Maisam M. Dadkan

Data Stewards

GeoICT

Folder Structures

How and where will the data be stored

- A good folder structure helps you and others find files efficiently
 - Your research data will be ‘shared’ even when not published
 - Direct colleagues within a project
 - Indirect colleagues in the department / UU
 - Project leaders / supervisors when contract ends
- Collaboratively develop a folder structure
- Do it at the beginning, or your data will be everywhere

Folder Structures



- Navigating your folder by others or in the future you
 - Structure may be less logical to others
 - Structure may have some special occurrences

Folder Structures



- Describe your structure and files in a readme.txt or a Word file
 - Helps determine contents of folders
 - Directs readers to the right location to place and find data

File name conventions



REMEMBER

Your filename has to make sense:

- ✓ to you
- ✓ to you – after a holiday
- ✓ to your colleagues
- ✓ & preferably to a machine

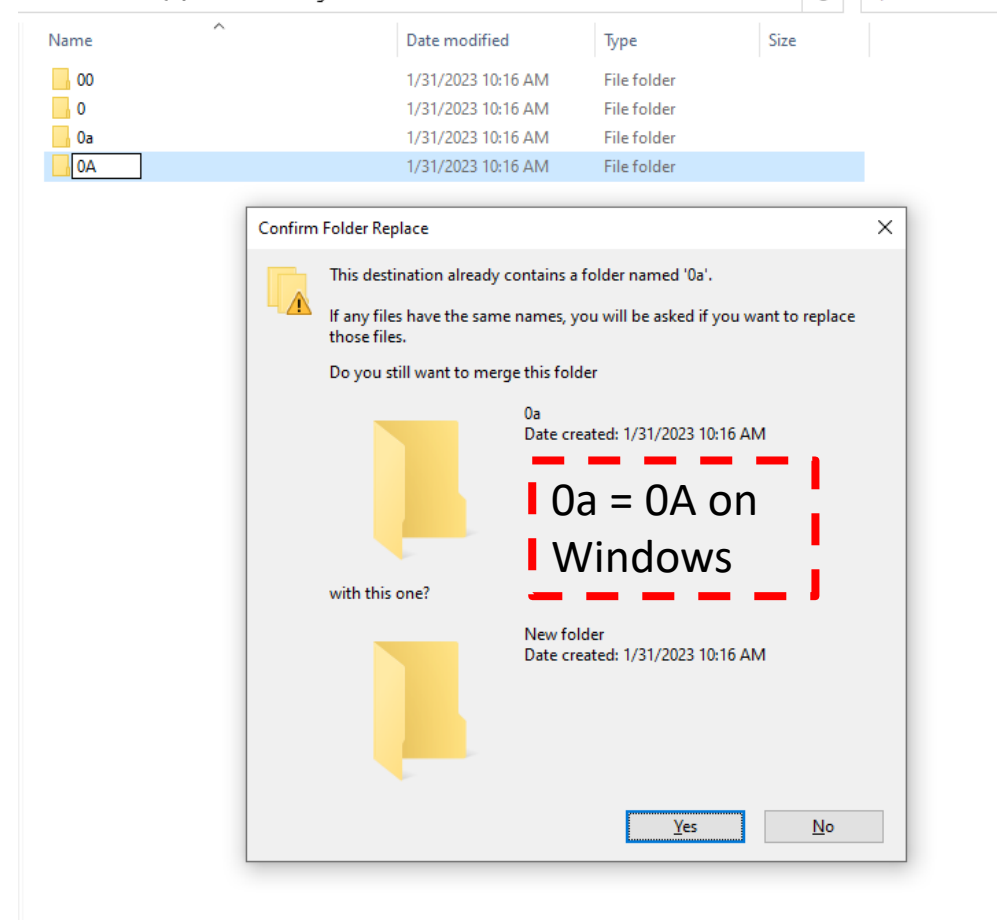
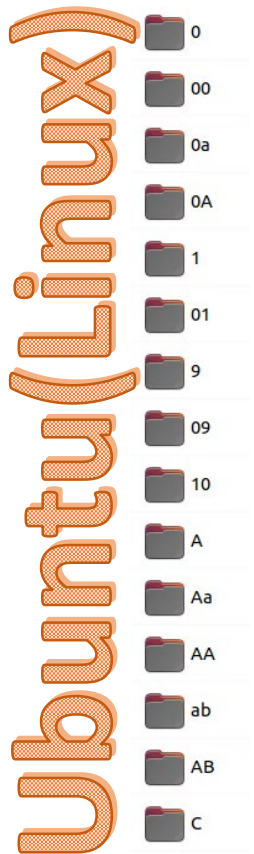
Universal Filesystem Sorting

Name files so they will sort in order on all file systems

- Windows
- MacOS
- Linux
- FreeBSD
- etc

Universal Sorting Order

1. Folders then Files
2. Numbers then Letters



Universal Sorting Order

Our suggestion: use consistent numbering with possible additions for letters.

- | | | | |
|----------|-----------|-----|---------|
| • 01.pdf | • 01g.pdf | v 1 | Garrett |
| • 02.pdf | • 01v.pdf | v 1 | Vincent |
| • ... | • 01m.pdf | v 1 | Maisam |
| • 55.pdf | • 02.pdf | v 2 | |

Hacking Sorting Order

If you have files that you want to always show at the top, prefix the name with numbers.

 Presentations		1/31/2023 10:29 AM	File folder	
 Workshops_ics		1/16/2023 11:18 AM	File folder	
 0_List_of_topics.docx		12/19/2022 12:00 PM	Microsoft Word D...	24 KB
 1_WEBSITE_Workshops Spring 2023.docx		1/19/2023 1:40 PM	Microsoft Word D...	26 KB
 2_VideoAdvertisement.jpg		1/19/2023 2:54 PM	JPG File	887 KB
 2_VideoAdvertisement.pdf		1/19/2023 2:53 PM	Adobe Acrobat D...	144 KB
 2_VideoAdvertisement.pptx		1/19/2023 1:55 PM	Microsoft PowerP...	133 KB
 3_PrintPosters.pdf		1/19/2023 2:00 PM	Adobe Acrobat D...	126 KB
 3_PrintPosters.pptx		1/19/2023 1:58 PM	Microsoft PowerP...	137 KB
 3_PrintPosters_a5pair.pdf		1/19/2023 2:04 PM	Adobe Acrobat D...	139 KB
 File Organization Goals.docx		12/13/2022 3:43 PM	Microsoft Word D...	23 KB
 Getting Started with Git and GitHub.docx		12/13/2022 3:28 PM	Microsoft Word D...	22 KB
 Getting Started with Jupyter Notebooks f...		1/3/2023 5:24 PM	Microsoft Word D...	22 KB
 How to Organize a Python Package for S...		12/13/2022 11:38 AM	Microsoft Word D...	20 KB
 How to Write a README File Goals.docx		12/19/2022 10:56 AM	Microsoft Word D...	23 KB

File name conventions

Use fixed elements in your filename:

- Version number
 - _v01.docx
 - _v34.docx
- Avoid __final__final__final2.docx with release candidate
 - _rc01.pdf
 - _rc02.pdf

File name conventions

Use fixed elements in your filename:

- Content description
- Project number
- Research team
- Subjects, IDs, Codes
 - Participant_4021A.mp3, Participant_6134M.mp3
 - NL_UT_Utrecht.jp2, NL_UT_Bilthoven.jp2
 - Heuvelrug_Ijsvogels.csv, Heuvelrug_Schapen.csv

File name conventions

Use fixed elements in your filename:

- Variables
 - Buffer_10m.shp
Buffer_5cm.shp
Buffer_3ft.shp
 - Simulation_10x_5y.csv
Simulation_5x_10y.csv

File name conventions

DOs

- ✓ Separate README document
 - codes used in filename
- ✓ < **25** characters
- ✓ From generic → specific
- ✓ Filenames independent of folder names

DON'Ts

- ❑ Don't use
 - Special characters (&%\$#)
 - Periods
 - Whitespace (better '-' OR '_')
- ❑ Duplicate files in different folders
- ❑ new_version → newer_version → newest_version → final version → final ultimate...

ISO 8601



YYYY-MM-DDTHHH:MM:SS.ssssZ

YYYY-MM-DD^THH:MM:SS.sssssZ

Year Month Day Hours Minutes Seconds Fractions of a Second

- International Standard on Date & Time
- Best way to represent dates in
 - Data
 - Files
 - Folders

Why ISO 8601?



ISO 8601 (Y/M/D)

1970-01-01
1970-01-02
1970-01-03
1970-02-01
1999-01-01
2000-03-01
2023-01-23

Most of the World (D/M/Y)

01-01-1970
01-01-1999
01-02-1970
01-03-2000
02-01-1970
03-01-1970
23-01-2023

United States (M/D/Y)

01-01-1970
01-01-1999
01-02-1970
01-03-1970
01-23-2023
02-01-1970
03-01-2000

Using ISO 8601

- Format can be truncated as needed
- Need just the year? YYYY
- Need the year and month? YYYY-MM
- Need year, month, and day? YYYY-MM-DD
- Need the date and the hour? YYYY-MM-DDTHHZ
- Need the date, hour & minute?
YYYY-MM-DDTHH:MMZ

Using ISO 8601

- Can't use hyphens/streepjes?

YYYYMMDD or YYYY_MM_DD

- Can't use colons?

HHMM, HHMMSS, HH_MM_SS

ISO 8601 Z Time

- Time part of the standard has options for time zones or UTC
- General Advice is to store time in UTC wherever possible
- UTC Denoted by **Z**
 - Need the date and the hour? YYYY-MM-DDTHH**Z**
 - Need the date, hour & minute? YYYY-MM-DDTHH:MM**Z**

ISO 8601 Local Time

- Can store local times
 - 2023-01-31T13:00Z (UTC)
2023-01-31T14:00+01:00 (Amsterdam)
2023-01-31T07:00-05:00 (New York City)
- Not advised because
 - Longer representation
 - Time zones are hard
 - Computers sync to UTC and display offset
 - +/- are not usable in file/folder names

ISO 8601 Caveats

- Avoid specific dates before 1583AD unless your research has a standard for Gregorian or Julian
- YYYY-DDD is an alternative when day of the year is important

ISO 8601 Caveats

- For BCE dates, prefix dates with BC and CE.
BC = Before Current(era), CE = Current Era
BC0423 (Plato had a drink)
CE2023 (This year)
BCE dates will break chronological order

ISO 8601 Caveats

- Recording a lot of data on the scale of nano-/micro-seconds?

Use epoch seconds in the data to avoid the extra time to format time as ISO 8601.

Epoch seconds are the number of seconds since 1 January 1970

The background is split horizontally. The top half is orange with a white molecular or network pattern of circles and lines. The bottom half is black. A white, torn-paper-like border separates the two colors.

Structured Files and Folders in Action

ESA Sentinels

- Sentinel is an ESA Earth Observation program
- Sentinel 1–6 produce terabytes of data every day
- This data needs to be structured for humans & machines



© ESA/ATG medialab

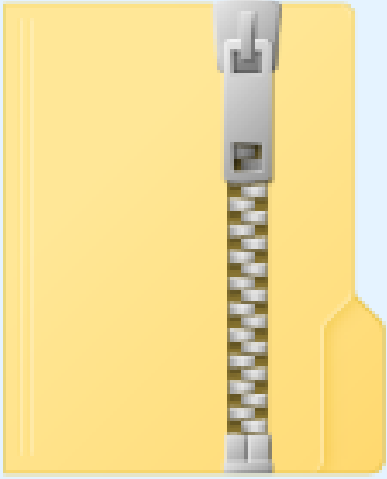
https://www.esa.int/ESA_Multimedia/Images/2014/07/Sentinel-2_brings_land_into_focus

Sentinel SAFE Format

- Structured folder layout so everyone knows where to find the data inside
- Structured file name so everyone can find important product information quickly
- Specification found on ESA website

<https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/data-formats>

Sentinel 2 SAFE Filename



S2A_MSIL1C_20221113T10431
1_N0400_R008_T31UFT_20221
113T141014.zip

S2A – Mission Sentinel 2, spacecraft A

MSI – Multispectral Imager

L1C – Level 1 C Processing

20221113T104311 – Time of Capture

N0400 – Processing Baseline 400

R008 – Orbit 8

T31 – UTM Zone 31

UFT – 100km grid assigned UFT

20221113T141014 – Time of Processing

Sentinel 2 SAFE Folder Structure

C:\USERS\SPEED001\DOWNLOADS\S2A_MSIL1C_20221113T104311_N0400_R008_T31UFT_20221113T141014.SAFE

INSPIRE.xml
manifest.safe
MTD_MSIL1C.xml

AUX_DATA
DATASTRIP
DS_ATOS_20221113T141014_MTD_DS.xml
QI_DATA
FORMAT_CORRECTNESS
GENERAL_QUALITY
GEOMETRIC_QUALITY
RADIOMETRIC_QUALITY
SENSOR_QUALITY

GRANULE
L1C_T31UFT_A038615_20221113T141014_MTD_TL.xml
AUX_DATA
AUX_CANSFO
AUX_ECHWFT

IMG_DATA
T31UFT_20221113T104311_B00.jp2
T31UFT_20221113T104311_B01.jp2
T31UFT_20221113T104311_B02.jp2
T31UFT_20221113T104311_B03.jp2
T31UFT_20221113T104311_B04.jp2
T31UFT_20221113T104311_B05.jp2
T31UFT_20221113T104311_B06.jp2
T31UFT_20221113T104311_B07.jp2
T31UFT_20221113T104311_B08.jp2
T31UFT_20221113T104311_B09.jp2
T31UFT_20221113T104311_B10.jp2
T31UFT_20221113T104311_B11.jp2
T31UFT_20221113T104311_B12.jp2
T31UFT_20221113T104311_B8A.jp2
T31UFT_20221113T104311_TCI.jp2

QI_DATA
FORMAT_CORRECTNESS.xml
GENERAL_QUALITY.xml
GEOMETRIC_QUALITY.xml
MSK_CLASS1_B00.jp2
MSK_DETFOO_B01.jp2
MSK_DETFOO_B02.jp2
MSK_DETFOO_B03.jp2
MSK_DETFOO_B04.jp2
MSK_DETFOO_B05.jp2
MSK_DETFOO_B06.jp2

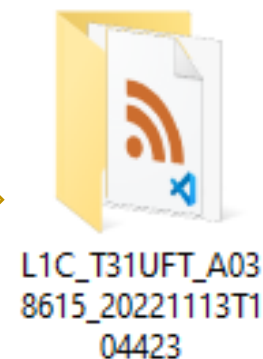
AUX_DATA
IMG_DATA
QI_DATA
MTD_TL.xml

MSK_QUALIT_B08.jp2
MSK_QUALIT_B09.jp2
MSK_QUALIT_B10.jp2
MSK_QUALIT_B11.jp2
MSK_QUALIT_B12.jp2
MSK_QUALIT_B8A.jp2
SENSOR_QUALITY.xml
T31UFT_20221113T104311_PVI.jp2

HTML
banner_1.png
banner_2.png
banner_3.png
star_bg.jpg
UserProduct_index.html
UserProduct_index.xsl

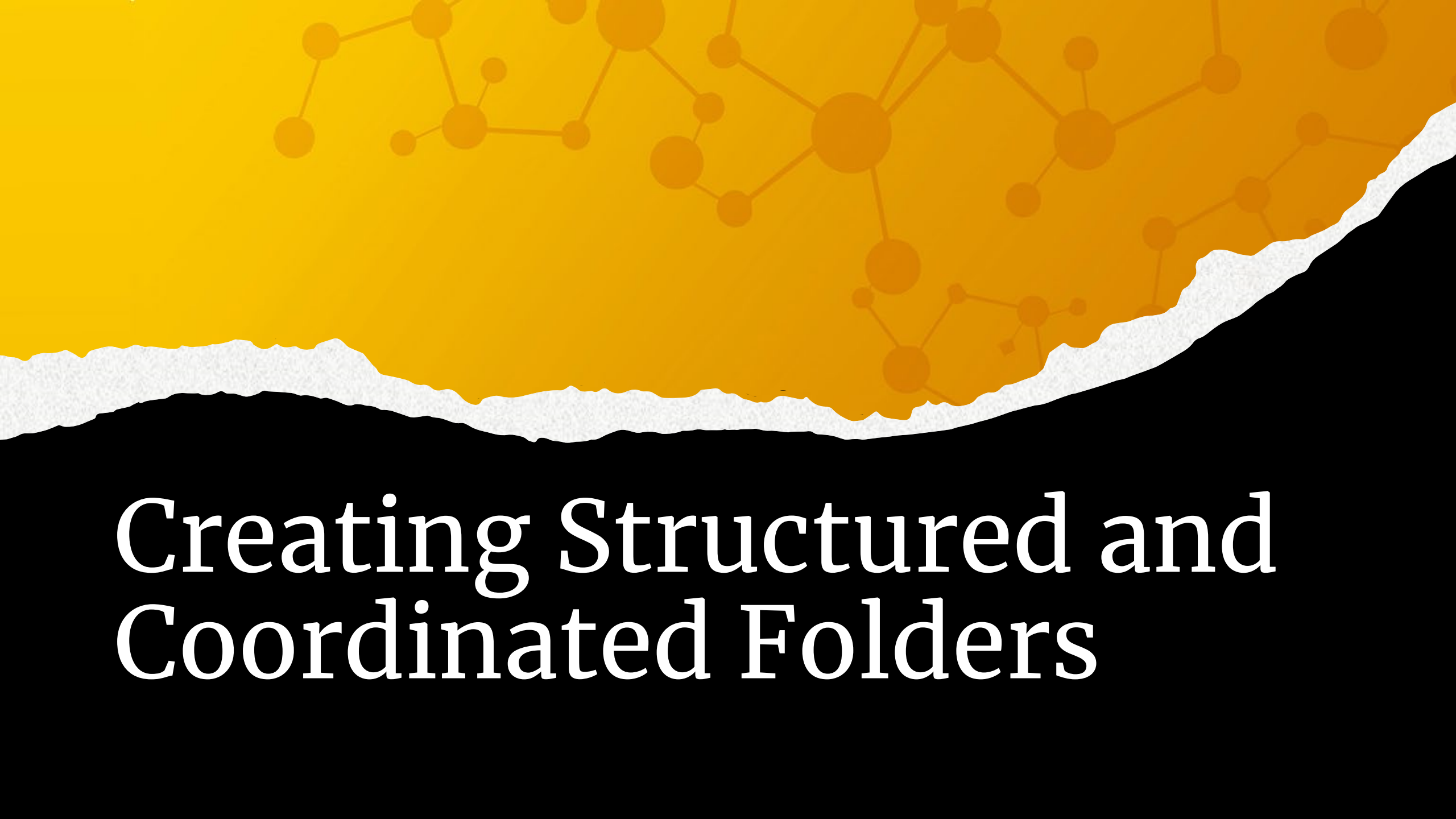
rep_info
S2_User_Product_Level-1C_Metadata.xsd

AUX_DATA	11/13/2022 4:03 PM	File folder
DATASTRIP	11/13/2022 4:03 PM	File folder
GRANULE	11/13/2022 4:03 PM	File folder
HTML	11/13/2022 4:03 PM	File folder
rep_info	11/13/2022 4:03 PM	File folder
INSPIRE.xml	11/13/2022 4:03 PM	XML Source File
manifest.safe	11/13/2022 4:03 PM	SAFE File
MTD_MSIL1C.xml	11/13/2022 4:03 PM	XML Source File



L1C_T31UFT_A03
8615_20221113T1
04423

T31UFT_20221113T104311_B01.jp2	11/13/2022 4:03 PM	JP2 File	3,194 KB
T31UFT_20221113T104311_B02.jp2	11/13/2022 4:03 PM	JP2 File	94,102 KB
T31UFT_20221113T104311_B03.jp2	11/13/2022 4:03 PM	JP2 File	99,437 KB
T31UFT_20221113T104311_B04.jp2	11/13/2022 4:03 PM	JP2 File	101,611 KB
T31UFT_20221113T104311_B05.jp2	11/13/2022 4:03 PM	JP2 File	30,492 KB
T31UFT_20221113T104311_B06.jp2	11/13/2022 4:03 PM	JP2 File	33,097 KB
T31UFT_20221113T104311_B07.jp2	11/13/2022 4:03 PM	JP2 File	33,001 KB
T31UFT_20221113T104311_B08.jp2	11/13/2022 4:03 PM	JP2 File	129,653 KB
T31UFT_20221113T104311_B8A.jp2	11/13/2022 4:03 PM	JP2 File	33,097 KB
T31UFT_20221113T104311_B09.jp2	11/13/2022 4:03 PM	JP2 File	3,671 KB
T31UFT_20221113T104311_B10.jp2	11/13/2022 4:03 PM	JP2 File	2,010 KB
T31UFT_20221113T104311_B11.jp2	11/13/2022 4:03 PM	JP2 File	31,441 KB
T31UFT_20221113T104311_B12.jp2	11/13/2022 4:03 PM	JP2 File	29,771 KB
T31UFT_20221113T104311_TCI.jp2	11/13/2022 4:02 PM	JP2 File	131,953 KB

The background is split horizontally. The top half is orange with a white molecular or network pattern of circles and lines. The bottom half is black. A white, torn-paper-like border separates the two colors.

Creating Structured and Coordinated Folders

Getting Started

- We'll make our folder structure in Word
- Using bullets to conceptualize the file system
- Shift + Enter to create a new line but no new bullet.

Getting Started

FolderStructure.docx

https://solisservices.sharepoint.com/:w:/r/sites/DatastewardsGeosciences/_layouts/15/...

Word FolderStructure - Saved

File Home Insert **Layout** References Review View Help

Margins Orientation Size Spacing Before 0 pt Spacing After 8 pt

Collaborative Data in Riperian Umbrella Networks (CDATA RUN)

Project Name

Page 1 of 1 20 words English (U.S.) Text Predictions: On Editor Suggestions: Showing 142% Fit Give Feedback to Microsoft

Getting Started

FolderStructure.docx

https://solisservices.sharepoint.com/:w:/r/sites/DatastowardsGeosciences/_layouts/15/...

Word FolderStructure - Saved

File Home Insert **Layout** References Review View Help

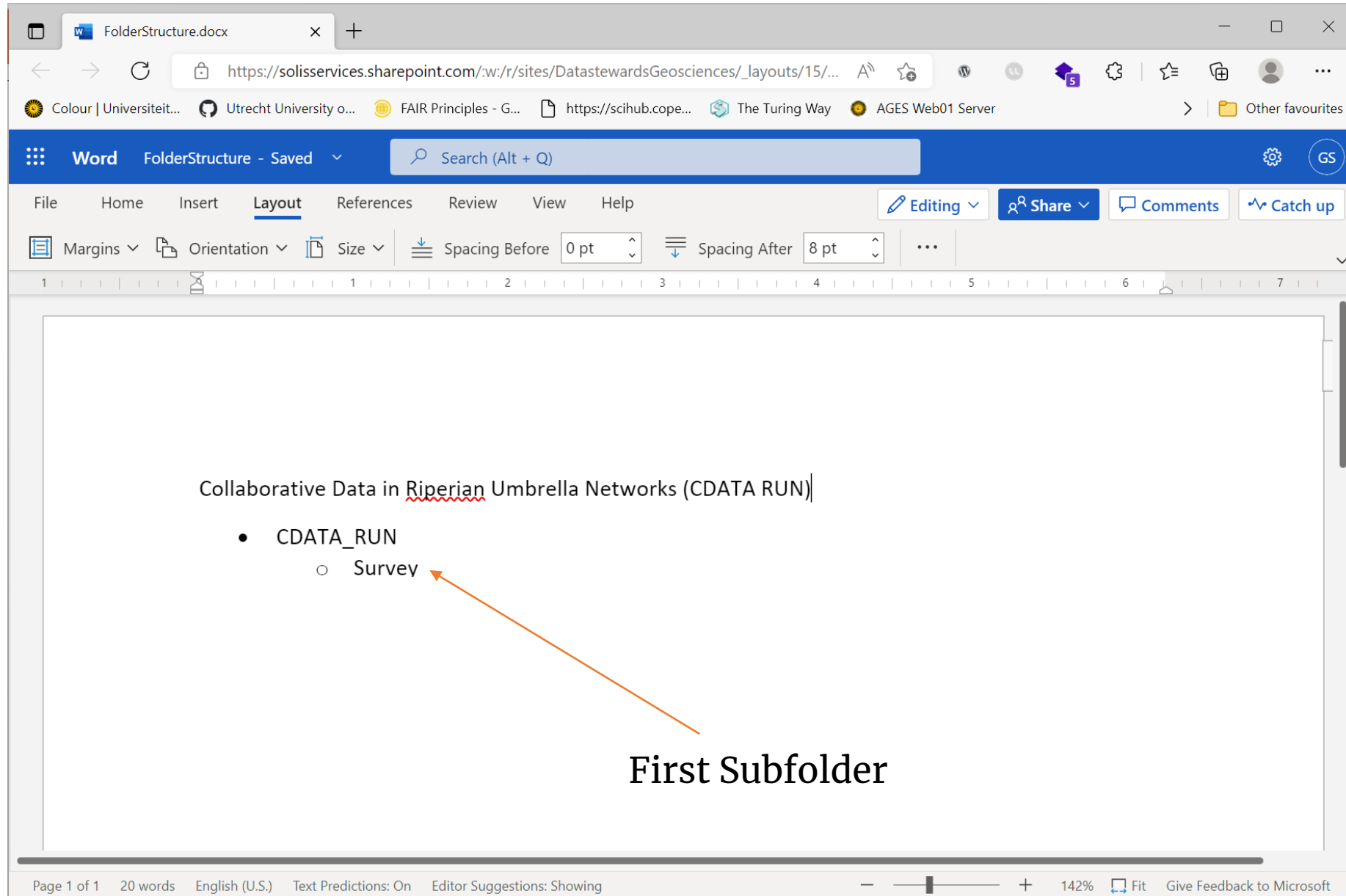
Margins Orientation Size Spacing Before 0 pt Spacing After 8 pt

Collaborative Data in Riperian Umbrella Networks (CDATA RUN)

- CDATA_RUN

Page 1 of 1 20 words English (U.S.) Text Predictions: On Editor Suggestions: Showing 142% Fit Give Feedback to Microsoft

Getting Started



The screenshot shows a Microsoft Word document titled "FolderStructure.docx" open in a web browser. The browser's address bar displays the URL: https://solisservices.sharepoint.com/:w:/r/sites/DatastewardsGeosciences/_layouts/15/.... The Word ribbon is set to the "Layout" tab, and the document content is as follows:

Collaborative Data in Riparian Umbrella Networks (CDATA RUN)

- CDATA_RUN
 - Survey

An orange arrow points from the text "First Subfolder" at the bottom of the page to the "Survey" subfolder in the list. The status bar at the bottom indicates "Page 1 of 1", "20 words", "English (U.S.)", "Text Predictions: On", "Editor Suggestions: Showing", and a zoom level of "142%".

Getting Started

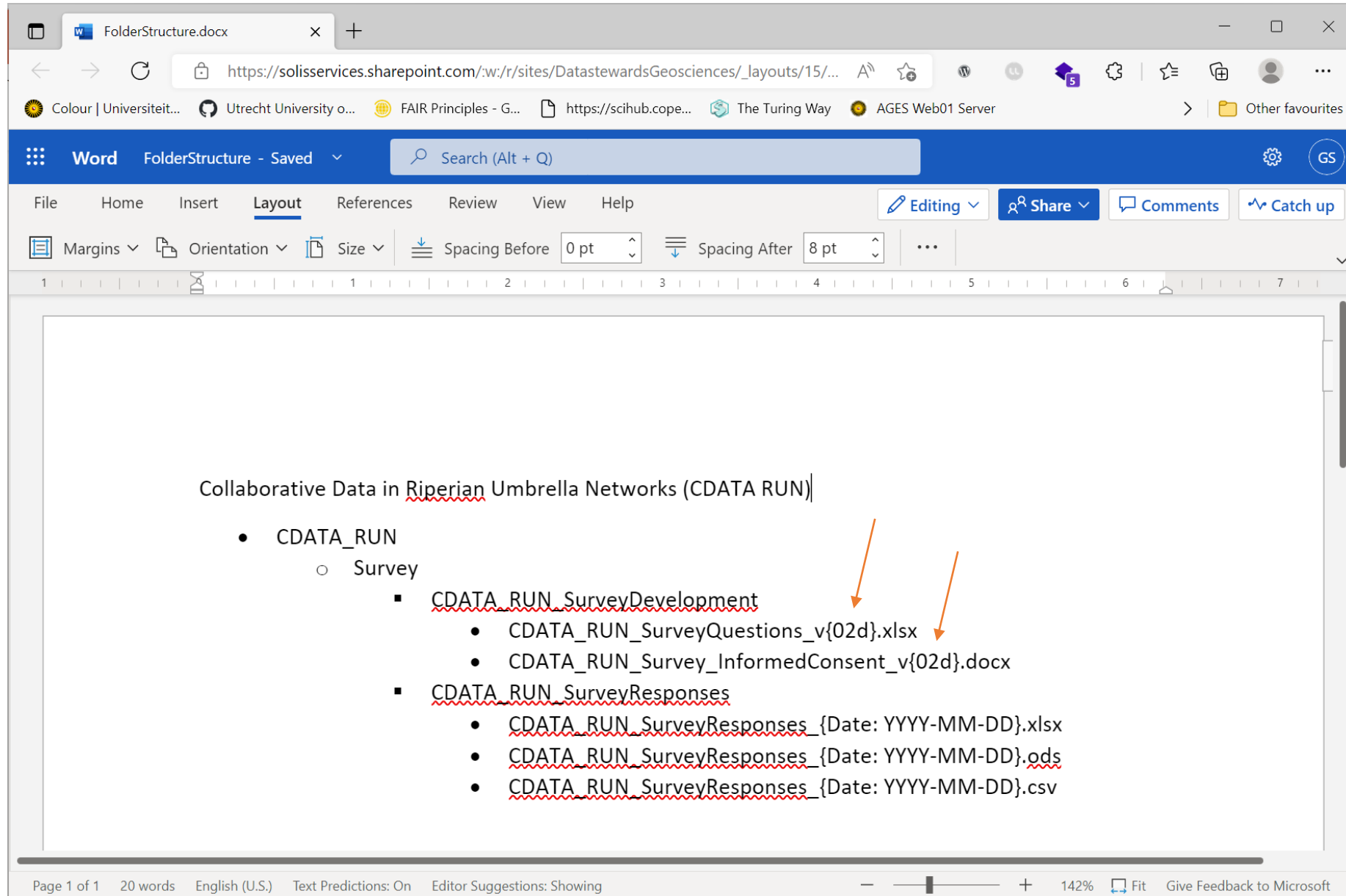
The screenshot shows a Microsoft Word document titled "FolderStructure.docx" open in a web browser. The browser's address bar shows the URL: https://solisservices.sharepoint.com/:w:/r/sites/DatastewardsGeosciences/_layouts/15/.... The Word ribbon is set to the "Layout" tab. The document content is as follows:

Collaborative Data in Riperian Umbrella Networks (CDATA RUN)

- CDATA_RUN
 - Survey
 - CDATA_RUN_SurveyDevelopment
 - CDATA_RUN_SurveyQuestions_v{02d}.xlsx
 - CDATA_RUN_Survey_InformedConsent_v{02d}.docx

The status bar at the bottom indicates "Page 1 of 1", "20 words", "English (U.S.)", "Text Predictions: On", "Editor Suggestions: Showing", and a zoom level of "142%".

Getting Started



FolderStructure.docx

https://solisservices.sharepoint.com/:w:/r/sites/DatastewardsGeosciences/_layouts/15/...

Word FolderStructure - Saved

File Home Insert **Layout** References Review View Help

Margins Orientation Size Spacing Before 0 pt Spacing After 8 pt

Collaborative Data in Riperian Umbrella Networks (CDATA RUN)

- CDATA_RUN
 - Survey
 - CDATA_RUN_SurveyDevelopment
 - CDATA_RUN_SurveyQuestions_v{02d}.xlsx
 - CDATA_RUN_Survey_InformedConsent_v{02d}.docx
 - CDATA_RUN_SurveyResponses
 - CDATA_RUN_SurveyResponses_{Date: YYYY-MM-DD}.xlsx
 - CDATA_RUN_SurveyResponses_{Date: YYYY-MM-DD}.ods
 - CDATA_RUN_SurveyResponses_{Date: YYYY-MM-DD}.csv

Page 1 of 1 20 words English (U.S.) Text Predictions: On Editor Suggestions: Showing 142% Fit Give Feedback to Microsoft

Adding Annotations

FolderStructure.docx

https://solisservices.sharepoint.com/:w:/r/sites/DatastewardsGeosciences/_layouts/15/...

Word FolderStructure - Saved

File Home Insert Layout References Review View Help

Calibri (Body) 11

Collaborative Data in Riperian Umbrella Networks (CDATA RUN)

- CDATA_RUN
 - Survey
 - CDATA_RUN_SurveyDevelopment
 - CDATA_RUN_SurveyQuestions_v{02d}.xlsx
 - CDATA_RUN_Survey_InformedConsent_v{02d}.docx
 - CDATA_RUN_SurveyResponses
 - CDATA_RUN_SurveyResponses_{Date: YYYY-MM-DD}.xlsx
Original download from Qualtrics, do not edit this file after downloading.
 - CDATA_RUN_SurveyResponses_{Date: YYYY-MM-DD}.ods
Convert the original xlsx download from Qualtrics to an open document spreadsheet, do not edit this file after downloading.
 - CDATA_RUN_SurveyResponses_{Date: YYYY-MM-DD}.csv
If we do download these, do not edit these files after downloading.

Page 1 of 1 62 words English (U.S.) Text Predictions: On Editor Suggestions: Showing 142% Fit Give Feedback to Microsoft

Code Example

The screenshot shows a Microsoft Word document titled "FolderStructure.docx" open in a web browser. The document content is as follows:

- CDATA_RUN_SurveyAnalysis
 - CDATA_RUN_Rpackage
 - .gitignore
Git ignore file, be sure to put in the data file template names so Git does not index our original data.
 - LICENSE
Initialize the repository with the MIT license.
 - README.md
 - CDATA_RUN_{package name}.r
 - CDATA_RUN_Outputs
Exports from our R scripts
 - *Tables*
 - CDATA_RUN_{package name}_{variable}_v{02d}.csv
 - *Graphics*
 - CDATA_RUN_{package name}_variable_v{02d}.csv

The status bar at the bottom indicates: Page 1 of 1, 109 words, English (U.S.), Text Predictions: On, Editor Suggestions: Showing, 142% zoom, and Fit view.

Interviews Example

The screenshot shows a Microsoft Word document titled "FolderStructure.docx" open in a web browser. The document displays a hierarchical folder structure for interviews, with the following content:

- CDATA_RUN_{package name}_variable_v{02d}.csv
- CDATA_RUN_Interviews
 - InterviewDevelopment
 - CDATA_RUN_InterviewQuestions_{v02d}.xlsx
 - CDATA_RUN_Interview_InformedConsent_v{02d}.docx
 - CDATA_RUN_InterviewsResults
 - Videos
 - CDATA_RUN_{PersonID}_{date YYYY-MM-DD}.mp4
 - Audio
 - CDATA_RUN_{PersonID}_{date YYYY-MM-DD}.mp3
 - Create this from the mp4, ffmpeg -i input.mp4 output.mp3*
 - Transcripts
 - CDATA_RUN_{PersonID}_{date YYYY-MM-DD}.docx
 - Use NVivo to generate the transcript from the original mp4 file.*
-

The document footer indicates: Page 1 of 1, 143 words, English (U.S.), Text Predictions: On, Editor Suggestions: Showing, 142% zoom, and Fit view.

GIS Example

FolderStructure.docx

https://solisservices.sharepoint.com/w:/r/sites/DatastewardsGeosciences/_layouts/15/...

Word FolderStructure - Saved

Search (Alt + Q)

File Home Insert Layout References Review View Help

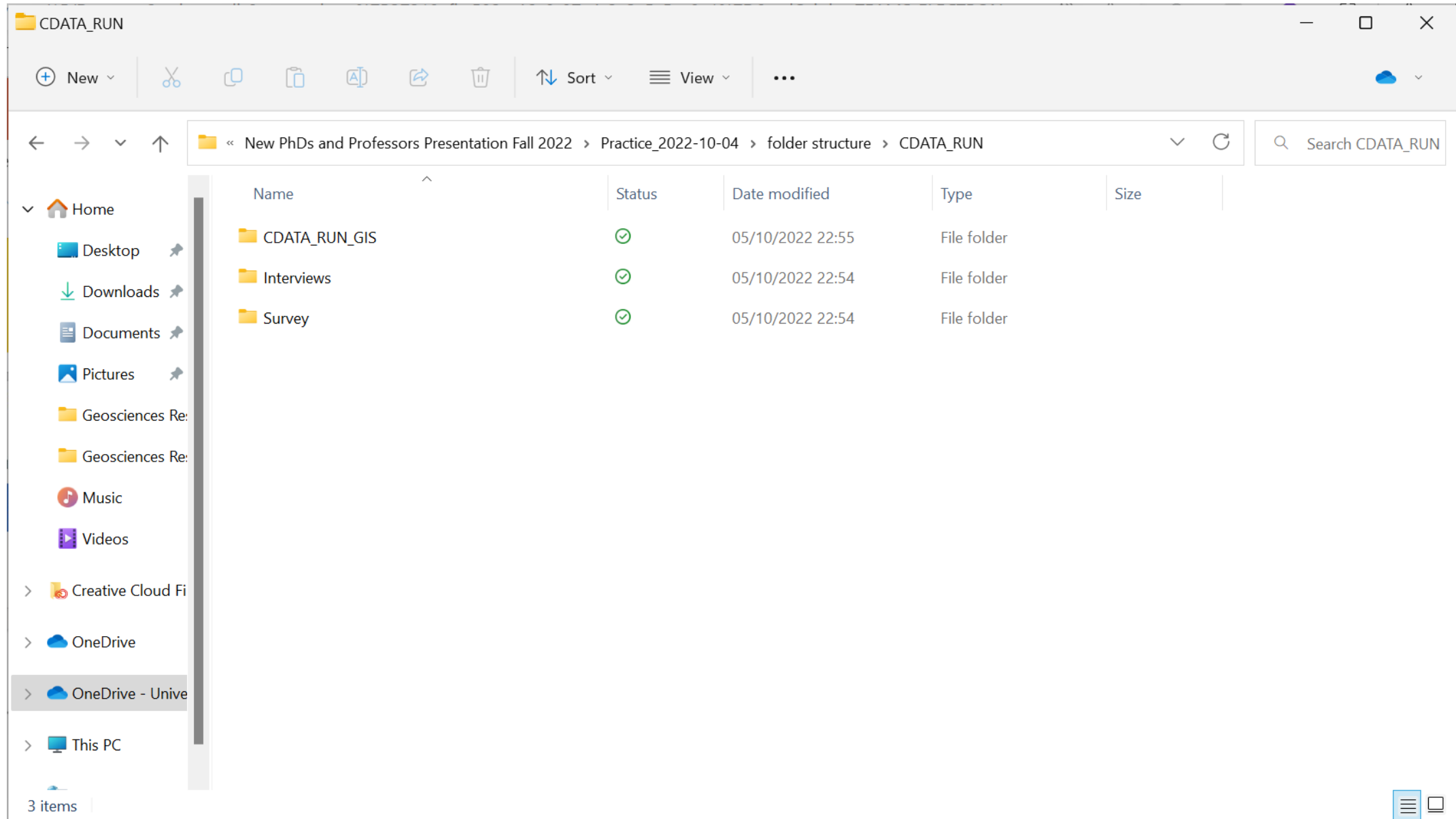
Editing Share Comments Catch up

Calibri (Body) 11

1 0 1 2 4 5 6 7

- CDATA_RUN_GIS
 - Imagery
 - Place all project imagery in this folder*
 - Sentinel_2
 - Zip
 - place the raw Sentinel 2 L1C zip files in this folder
 - 10m_jp2
 - Process the raw Sentinel 2 L1C data for just the 10m layers using the composite bands tool.
 - {Satellite: S2A}_{UTM Zone}_{Tile ID}_{date YYYY-MM-DD}.jp2
 - Elevation
 - LiDAR
 - las
 - Place the las files in this folder, we need las for ArcGIS Pro
 - DEM
 - Process the las ground data into a DEM and place in here
 - {CDATA_RUN}_{LiDAR Source}_{tile}.tif
 - We have to use tif because nothing else supports floating point raster types

How it translates



How to Create a Directory Tree from an Existing Folder

- Windows
 - tree /f
- Linux
 - tree -a
- MacOS
 - tree -a

```
C:\USERS\SPEED001\DOWNLOADS\S2A_MSIL1C_20221113T1043: gtspeed@gtspeed-ubuntu2204-vm:~/Documents/iBridges-Gui$ tree -a
INSPIRE.xml
manifest.safe
MTD_MSIL1C.xml

AUX_DATA
DATASTRIP
├── DS_ATOS_20221113T141014_S20221113T104423
│   └── MTD_DS.xml
└── QI_DATA
    ├── FORMAT_CORRECTNESS.xml
    ├── GENERAL_QUALITY.xml
    ├── GEOMETRIC_QUALITY.xml
    ├── RADIOMETRIC_QUALITY.xml
    └── SENSOR_QUALITY.xml

GRANULE
├── L1C_T31UFT_A038615_20221113T104423
│   └── MTD_TL.xml
├── AUX_DATA
│   ├── AUX_CAMSFO
│   └── AUX_ECMWFT
└── IMG_DATA
    ├── T31UFT_20221113T104311_B01.jp2
    ├── T31UFT_20221113T104311_B02.jp2
    ├── T31UFT_20221113T104311_B03.jp2
    ├── T31UFT_20221113T104311_B04.jp2
    ├── T31UFT_20221113T104311_B05.jp2
    ├── T31UFT_20221113T104311_B06.jp2
    ├── T31UFT_20221113T104311_B07.jp2
    ├── T31UFT_20221113T104311_B08.jp2
    ├── T31UFT_20221113T104311_B09.jp2
    ├── T31UFT_20221113T104311_B10.jp2
    ├── T31UFT_20221113T104311_B11.jp2
    ├── T31UFT_20221113T104311_B12.jp2
    ├── T31UFT_20221113T104311_B8A.jp2
    └── T31UFT_20221113T104311_TCI.jp2

docker
├── debbull
│   └── Dockerfile
├── Dockerfile
├── .gitkeep
├── README.md
├── ubu1804
├── ubu2004
│   └── Dockerfile
└── .git
    ├── branches
    ├── config
    ├── description
    ├── HEAD
    ├── hooks
    │   ├── applypatch-msg.sample
    │   ├── commit-msg.sample
    │   ├── fsmonitor-watchman.sample
    │   ├── post-update.sample
    │   ├── pre-applypatch.sample
    │   ├── pre-commit.sample
    │   ├── pre-merge-commit.sample
    │   ├── prepare-commit-msg.sample
    │   ├── pre-push.sample
    │   ├── pre-rebase.sample
    │   ├── pre-receive.sample
    │   ├── push-to-checkout.sample
    │   └── update.sample
    ├── index
    ├── info
    │   └── exclude
    ├── logs
    │   ├── HEAD
    │   └── refs
    │       ├── heads
    │       └── master
```