Manuscript: essd-2021-280
Authors: Xu, H.-H., Niu, Z.-B., Chen,Y.-S., et al.
Title: Ordovician to Silurian graptolite specimen images for global correlation and shale gas exploration.

Summary: It is a welcome development that more museums are placing photographs of their fossil collections onto the internet as publicly accessible galleries. This allows researchers to determine whether to travel to view the physical specimens – often an expensive undertaking – or to seek to loan the specimens, which can risk damage or loss of irreplaceable materials. This contribution will make available many high-resolution images of graptolite specimens from 154 important Chinese locations below and above the Ordovician Silurian boundary. Graptolites are one of three index fossil groups (graptolites, conodonts and chitinozoans) most widely used for global correlation in this time interval. The same interval includes a major extinction event across which the proportions of trilobite families change dramatically. The host shales are an important source rock for gas. The gallery of images has value for the local economic geologists and for academic research.

The fossil collection is connected to publications spanning many years. Fortunately, the authors have sought expert help to ensure that the graptolite taxonomy is up to date. Unfortunately, the experts are not all named. The images are accessible through five large ZIP files. A single Excel file provides detailed records for the 1550 high-resolution images of 113 species and subspecies. This file provides both the current expert name for the taxon and the originally "tagged" name in the collection. It includes higher taxonomic categories, collection locality, specimen number, image file name and publication. It also includes ages for the ends of a range. It does not make clear whether this is an estimate of the global range, the local range, or the age of the photographed specimen. The "mean age" column is rather puzzling, unless the values are attempts to calibrate the age uncertainty for the actual specimen (this seems unlikely). Of course, these calibrations are subject to change as new dates emerge, for example from the 2012 to the 2020 editions of Elsevier's Global Time Scale books.

Downloading the ZIP and xslx files is straightforward. Finding the ZIP file for a given taxon image is very cumbersome. Surely the ZIP file could have a column to show this.

The abstract claims to support an image-based automated classification model. I am unable to find justification for this claim, perhaps because it was not possible to access the necessary tools with the url provided. It was not clear to me how readers would determine the original dimensions of the specimens in the images that we have downloaded. No image that I examined has included a scale. This is a very basic requirement for illustrations of paleontological specimens. Line 165 mentions resizing the images. Perhaps all would be clear if it were possible to access the FSIDvis tool.

I was unable to access the FSIDvis tool from the information given. The text figures explained a lot, but lines 158-175 include some technical details about t-SNE that I could not understand.

The text includes numerous common failures of translation from Chinese to English. They are distracting, but almost none obscure the meaning.

Data and Tool Accessibility: I downloaded one of file large ZIP files that make up the gallery (272images_4.zip). It took one hour, using a fast internet connection at my university to download the 1.3GB. I examined a selection of the images. The photography is of high quality. The fossil preservation

is not the best, but that is presumably a true representation of the nature of the material at these localities. Some of the images included specimen numbers. None of the images included a scale. The images are valuable for anyone doing local or inter-regional research that includes the Chinese Lower Paleozoic. Expert users can make their own judgment about the species assignments, but few of the specimens could be recommended to those who wish to learn the characteristics of a taxon. For that, the original holotypes are essential and better-preserved material is desirable.

The xlsx file seems not to indicate which of the five ZIP files contains a given image. Instead, the "Preview" button on the host xenodo website opens a list of the image file names, for a given ZIP file. Most file names begin with the specimen number, followed by the species binomen. Some specimen numbers are alphanumeric. It is not easy to predict the range of specimen/image numbers from one ZIP file to the next. Surely the authors could add to their useful xlsx file a column indicating which ZIP file holds the image.

The caption to text-figure 7 refers to a "Fossil Specimen Image Dataset Visualiser" (FSIDvis) with a clickable link. The link is compromised by a line break. I tried entering many versions of the link to remove the page break. All yielded either an error message or an unsuccessful Google search. Another link appears at the end of the text under "Data Availability." It did not generate an error, but leads to a blank pale-blue screen. Consequently, it is not possible to comment on the quality or utility of this tool. Access to it might have overcome some short comings of the text and images.

Text Style: The usual Chinese-to-English grammatical errors are present. Two are most common: mismatches between singular are plural for verbs and subjects; and the misuse/omission of the definite and indefinite articles (the / a / an / some). The opening line (39) has examples of both flaws. I understand that Mandarin has no grammatical equivalents. Any English-speaking editor could easily fix all this. A spell-checker would find 'Geobiodiverisy' (line 103). More difficult would be the apparent contraction of 'shale gas' to 'gale.'

Text Content and Omissions: There may be very few readers who will need the introductory generalizations about graptolite paleontology. Readers from some institutions and museums may be upset by the gratuitous boast about the largest paleontology research center in the world.

Figure 4 claims that the "scientific species name of every specimen is given on each image." This is true for the fifteen images in figure 4, but it is not true for any of the images that I downloaded and viewed. The taxon names appear in the file names, NOT on the images themselves.

What is meant by a "big mixed, small settlements" posture (line175)?

Some readers might like to know the names of the "several distinguished palaeontologists" (lines 148-9) who provided the current taxonomic assignments for the specimens. Did they simply provide current names for older synonyms or did they make some taxonomic revisions for specimens that were initially mis-identified? The experts who performed this service were not named in the xlsx file. The acknowledgements paragraph mentions only one (X.M.). Should readers assume others are among the coauthors? The statement that authors contributed equally to this work (line 13) is not helpful and seems at odds with lines 200-204. Perhaps we should think about it this way: if a figured specimen were to be included in a synonymy list, should all these co-authors be named? Should this contribution to be understood to include any formal re-identifications?