Feb 17

About the data loss on February 16th 2020

Unfortunately, for the first time since 9 years and 3 months online, on February 16th 2020 AstroBin experienced a serious data loss.

The images uploaded to the website, belonging to a significant percentage of the user base, were permanently deleted from the remote storage server. The reason this happened is manyfold, and I will offer here a brief non-technical explanation, while taking full responsibility.

In the immediate aftermath of the disaster, I experienced shock and agitation that were second only to a time, some years ago, when a (since solved) health problem made me fear for my life.

Such emotional response was due in part to my feelings towards AstroBin as a defining part of 25% of my life on Earth, in part to the bubbling up of my overwhelming sense of responsibility towards this community, and in part to fear of what it would mean to my family, financially, if AstroBin were to be forfeit.

Luckily I managed to keep my cool, thanks to amazing emotional support from my wife, and the positive messages that over 200 of you wrote to me.

The chain of events that caused the data loss

1. AstroBin's images are stored on the Amazon “Simple Storage Service”, also known as S3. Amazon states the following in their S3 FAQ:

   Q: Where is my data stored?
   A: [..] your objects are automatically stored across multiple devices spanning a minimum of three Availability Zones, each separated by miles across an AWS Region. Objects stored in the S3 One Zone-IA storage class are stored redundantly within a single Availability Zone in the AWS Region you select.

   When I started AstroBin, I was inexperienced in these matters, and I wrongly assumed, from that sentence, that data was automatically duplicated across three availability zones, and in case of data loss in a zone, it could be recovered from any of the other

two. Apparently, this is incorrect.

2. AWS S3 has been working so well the past 9 years, with exactly zero issues, that I literally never had to visit the S3 console or worry about its settings. In retrospect, this was very naïve of me.

3. Because of the above, I never found out that S3 offers Versioning (so that deleted content can be restored) and automatic data replication (the latter since 2015).

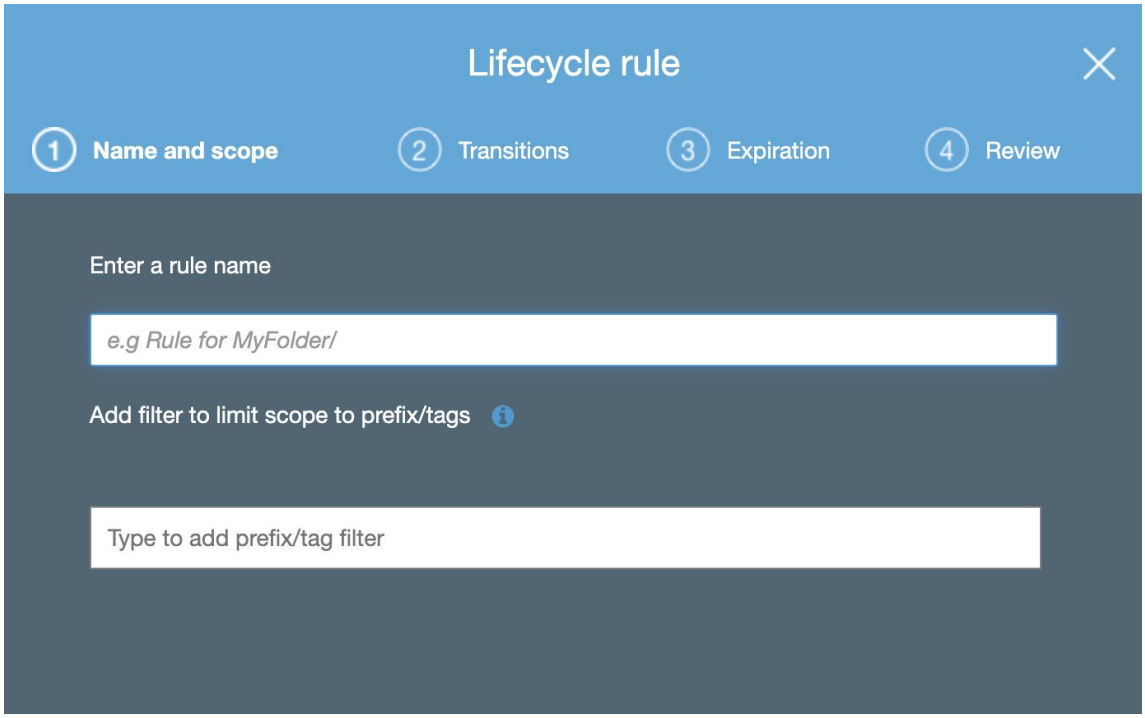
4. During the implementation of an exciting new feature on AstroBin, on February 15th I logged in to the S3 console and created a new data storage bucket for a “beta” AstroBin deployment that would allow me to test the new functionality on a production environment.

5. While I was waiting for some data to be copied, or a build to complete, I decided to have a look at AstroBin's main image storage bucket. Call it fate if you will.

6. I noticed that a folder called “tmpzips” had hundreds of GB of data that were sitting there absolutely needlessly: that folder is used to serve downloads of raw data in .zip format, and such downloads are only available for 7 days (on AstroBin’s side) but they were never getting expired on S3.

7. To save money on storage, I decided to add a rule to delete files in the “tmpzips” folder after they were older than 7 days. Hindsight is 20/20, but this can be universally regarded as utterly stupid.

8. The S3 console interface to add such lifecycle rules has **very poor UX** (see image below). Not making excuses, but it's true.



The "MyFolder/" placeholder, and the fact that the "Add filter to limit scope to prefix/tags" label seems to be associated to the first text input instead of the second (since it's closer) let me to put the name of the folder that I wanted affected by this rule in the wrong input. I don't want to blame anyone but me, but this is very poor UX on Amazon's side.

9. I set the rule to delete objects older than a week permanently, and at the end of the process Amazon did not warn me that this rule would apply to the whole storage bucket. Again, **very poor UX**.

10. I confirmed, then went to bed.

11. I woke up and 4.8 TB of data had been deleted from the storage. I stopped it as soon as I realized what had happened, so the entire archive wasn't lost.

That's it, no further warnings or confirmations necessary. Yes, I misread the form below, but in my opinion it was **just too easy** to delete a good part of 9 years worth of data.

How to prevent this from happening again

The silver lining here is that I learned something. I made assumptions and did not revisit them periodically. I did not have a process to review data safety conditions. I have now set up Versioning on the bucket, and Data Replication to a second Amazon data center. Of course this will make the AstroBin bills heftier, but apparently I was on too light a bill for many years.

Additionally, I will hire a consultant to review my software architecture and Amazon settings, and find potential holes that should be preventively closed.

Moving forward

I'm writing this two days after, after spending 16 hours on Sunday trying to figure out a way out and communicating the situation, all the while updating AstroBin to handle missing images (i.e. redirect you to a page where you can replace the image file), and not-quite-sleeping 4.5 hours. Rinse and repeat for Monday: 18 hours of work fire-fighting and keeping everyone updated (apologies if I didn't reply to all of the messages), followed by 4 hours of sleep.

I have received an overwhelming amount of positive messages, frankly beyond expectation. You all are an amazing bunch, with great empathy and understanding. Seriously, I'm amazed: so many of you have shown such a display of being damn fine human beings, that you make me want to be a better person myself!

Some people will leave the website to never return, of course. Some will never be able to trust it nor me again, and that's understandable. **In fact, I encourage that.** I should be the first one to trust myself less. As of today, tho, only 6 people have deleted their accounts, and one of them asked that it be restored.

Some people will have lost a lot of images. If you have the original files at hand in your computer, or an easily accessible backup, it should take you no more than 1-2 minutes per image, to fix the “sad face” thumbnails. This is because all the data associated with images (title, description, equipment used, acquisition details, comments, views, likes, bookmarks, and so on) are still intact and unaffected.

Some of you will not have backups. I understand that: AstroBin has partly advertised itself as a “safe storage for your images”. At the same time, AstroBin's FAQ has, since inception, stated that this is not a backup service and you should keep copies of your work. AstroBin has always been more oriented towards the communitarian and social aspect of astrophotography. I couldn't compete with Dropbox, or Google Drive, or a thousands other real data storage services out there. Anyway, the fact that I said that AstroBin is not a backup service is not an excuse. I still take full blame and responsibility.

Some of you will have lost no to few images, and hopefully you can restore them.

Many people have sent me messages telling me not to “give up on AstroBin”. I don't think that's up to me: it's up to you. Yes: I lost tens of thousands of images. Many of them, belonging to people with free accounts or who have since left the hobby, will probably never be recovered: this is definitely something I hate.

Anyway, at the time of writing, most of the active users have been replacing their images and fixing their galleries. Many have told me they intend to, and are just waiting out the storm.

Offering compensation

Some of you have told me that they want no compensation, just assurance that this won't happen again. They would even support AstroBin further, to allow me to dedicate more time to it.

I want to offer the option to receive compensation to everyone who is currently a paying member. As soon as possible, I will add a page on AstroBin where you can redeem an amount of free additional time of your choosing, on your existing subscription, between 0 and 12 months.

FAQ

- Are you sure this will never ever happen again?

  If I were a salesman I would tell you that it's 100% sure. But I'm an engineer and an honest person. I have learned a tough lesson from this experience, and I have taken measures so make sure the same thing won't happen again. In life, nothing is ever sure. I will continue to do my best and I will rely more on help from people who know better than me.

  Having said that, yes, I have reviewed all of AstroBin data storage backends and policies, and they are currently all safe from reasonable harm. What happened this weekend, i.e. human error, is to be considered reasonable harm. Unreasonable harm would be two asteroids hitting both data centers where the data is.

- How do we know the rest of the data is safe?

  AstroBin uses 4 different subsystems to store data:

  1. The S3 storage mentioned above, for the image files: JPG/PNG/GIF/FITS files are stored there. This is the only subsystem that was affected.

  2. The database: it stores all the meta-data associated to images, the users, everything needed to make the website work, all text data. This is hosted on a separate server and backed up daily. My next step is to set up a task to periodically dump the snapshot off-site for increased safety.

  3. The search database: it is built on the fly from database data, and it exists only for performance reasons. It's not critical because all the data can be recreated from the

database, although it can take around 24 hours.

4. The main server: the server holds no data of its own, just the code and some ephemeral caching information. If the server were to suffer a failure, I could deploy another one in 15 minutes. The code is on GitHub and on two of my computers.

- Is my personal data safe?

  AstroBin doesn't store much about you, personally. Payment information, such as credit cards, are not stored on AstroBin at all: everything is outsourced to PayPal, so there's nothing to ever compromise on AstroBin, on that front.

- Is my password safe?

  The passwords are stored with encryption on AstroBin's database, which was unaffected and is unlikely to suffer from a catastrophic issue in the future.

- How do I replace my images?

  Go to your gallery: if you see a thumbnail with a sad :'(face, that image is most likely missing. Wait a couple of minutes, then refresh the page. Those that still have a sad face, are most likely not automatically repairable by AstroBin.

  Click on a sad face, that will redirect you to one of the following pages:

  1. The “Edit basic information” form: if your image doesn't have revisions, or it has revisions and none of them is marked as final, you got there because your main image file is missing. Use the Image file form field to upload the image again (the context on that page should help you find it on your computer), then click on the Save button.

  2. The “Edit revision” form: if your image has a revision that is marked as final, and that revision is indeed missing, you will get there. Please replace the file using the same instructions as the point above.

- One of my images has multiple revisions, some are broken, some not. How do I fix them individually?

  Click on the ✏ icon that is visible when you hover your mouse on a revision's thumbnail: you will get to the form where you can replace the missing file. If you don't care about that revision anymore, click on the trashcan button instead.

- Has everything been done, that could be done to attempt a data recovery?

> Yes, thanks to the help of an AstroBin user who works at Amazon and has connections, the issue was quickly escalated and Amazon went the extra mile to attempt and recover the deleted files. Unfortunately, despite their best effort, they were unable to.

Conclusion

Many of you (I don't know if “most” actually, but that's the feeling) have been replacing your missing images and are willing to give me another chance! I'm trying to put myself in your shoes right now, to think objectively. What would my reaction have been if I was a user and this happened? About 20 of my images are affected, and I have copies, so it's going to take me half an hour to fix my gallery. Obviously I have been too busy fire-fighting to do it so far, but I feel that while this sucks, it's no big deal *for an individual* like me.

What bums me the most, is the loss of images belonging to people who have grown out of the hobby and don't care anymore to come and replace them. Or people who have deceased.

This loss goes against my vision for AstroBin. My initial vision, which was to store and catalogue the output of the world wide astrophotography community. Over the years, AstroBin has shifted focus to more communitarian and social networking aspects, and the need to finance the project via paid memberships has anyway jeopardized that vision, by means of “excluding” people who didn't want to pay, or couldn't afford to.

I need some time to process this identity crisis and how I want to present AstroBin to newcomers to the hobby. I'm sure the help from the community will be invaluable here.

So far, judging from all the messages I've received, you all have taken this better than I have. So many of you have focused on rebuilding and comforting me, while my initial reaction was a deep shock and despair.

I've always felt a huge sense of responsibility towards AstroBin, as the sole guardian and benevolent dictator of this community. This responsibility, apparently, was not enough to prevent me from making a huge mistake, and I currently feel abysmally awful about it.

I just really want to thank everybody some more about your emotional and financial support.

Sincerely,
your deeply, deeply sorry Salvatore.