

archive Transparency

Matthieu Daumas
archive.rip



Resilient
Internet
Preservation

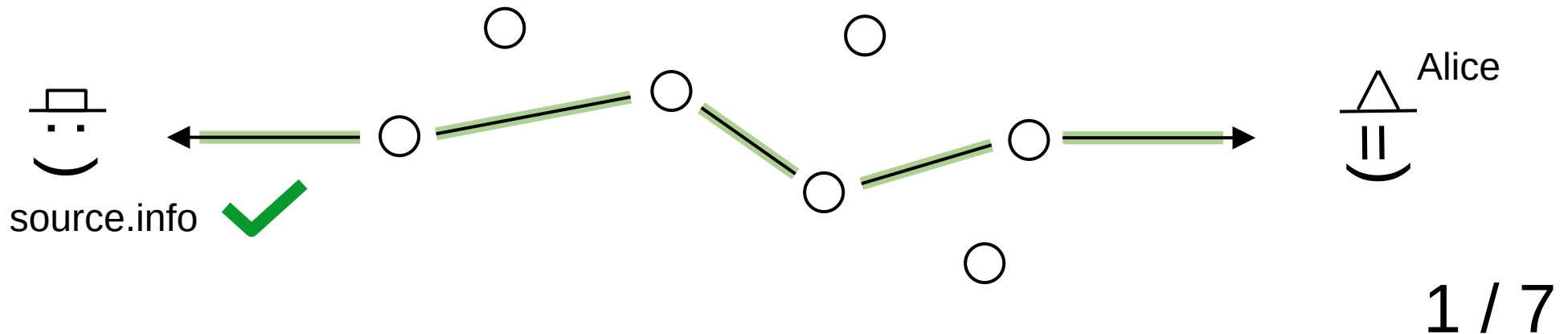
With funding from the:



Federal Ministry
of Research, Technology
and Space

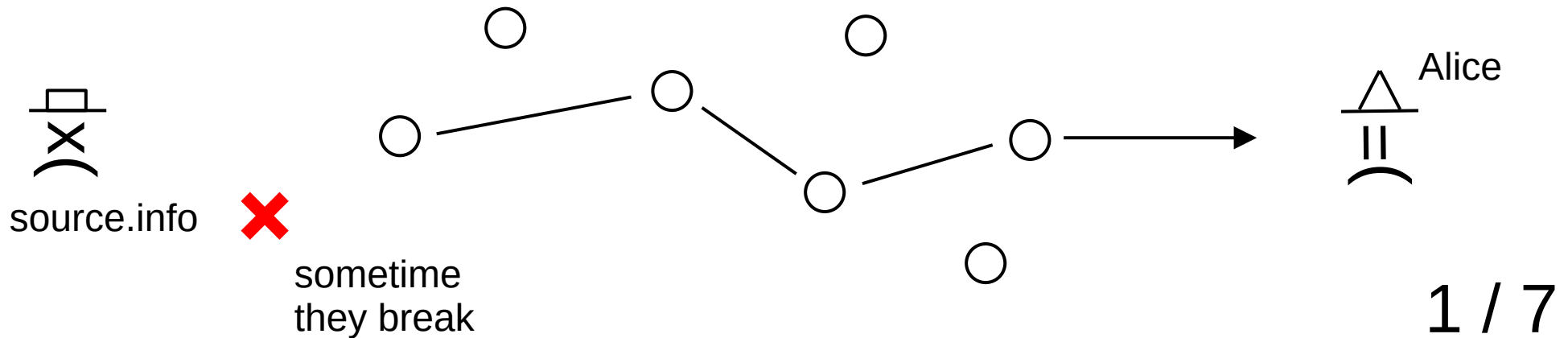
Introduction

We can have trust on the internet (thank you WebPKI) ;



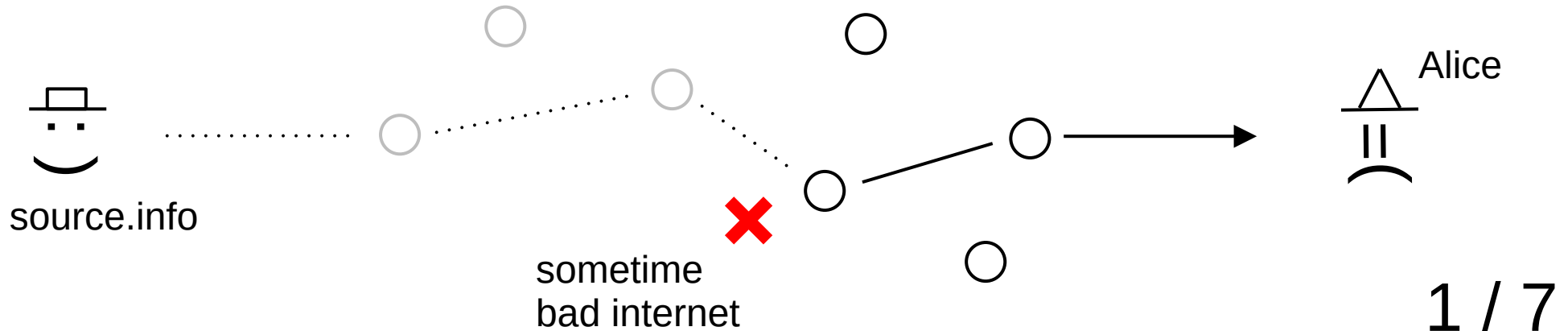
Introduction

We can have trust on the internet (thank you WebPKI) ;



Introduction

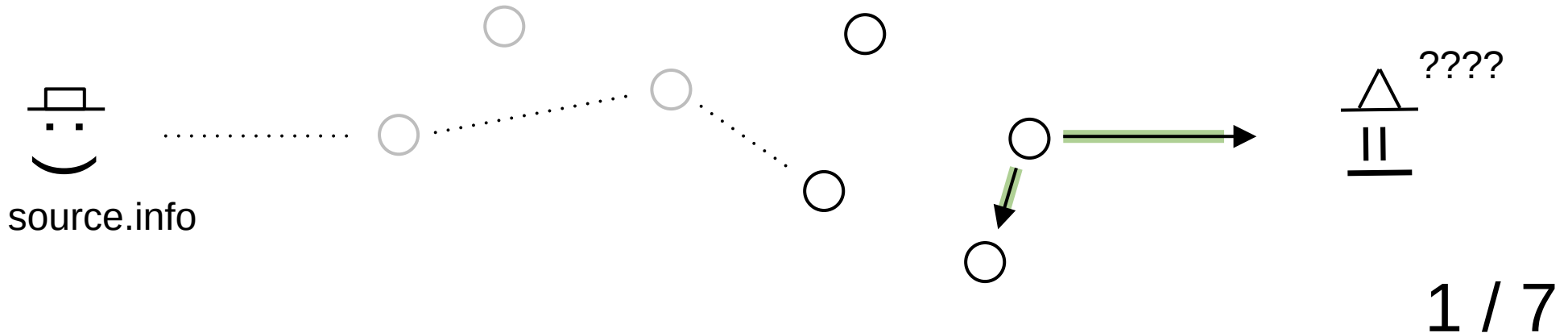
We can have trust on the internet (thank you WebPKI) ;



Introduction

We can have trust on the internet (thank you WebPKI) ;

We rely on third-party repositories all the time for availability ;

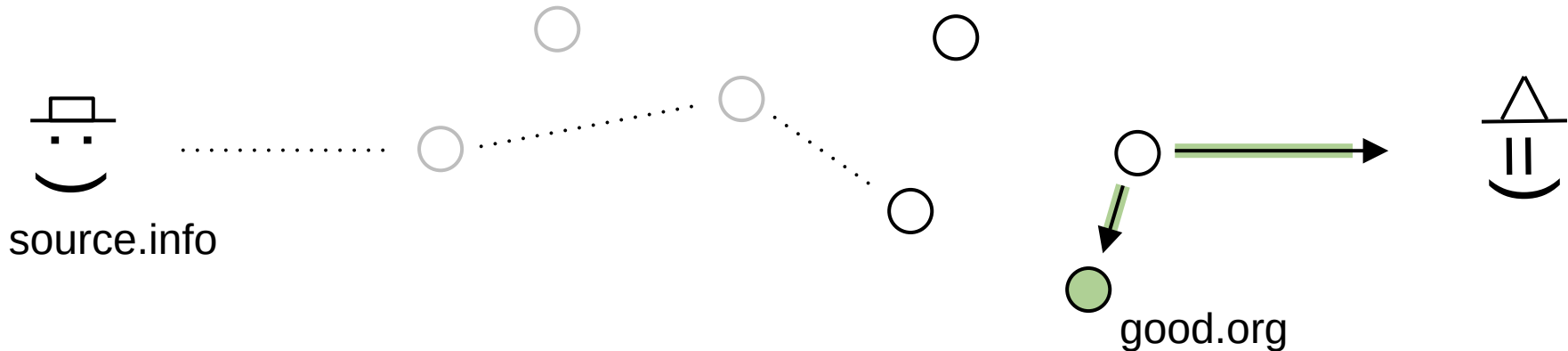


Introduction

We can have trust on the internet (thank you WebPKI) ;

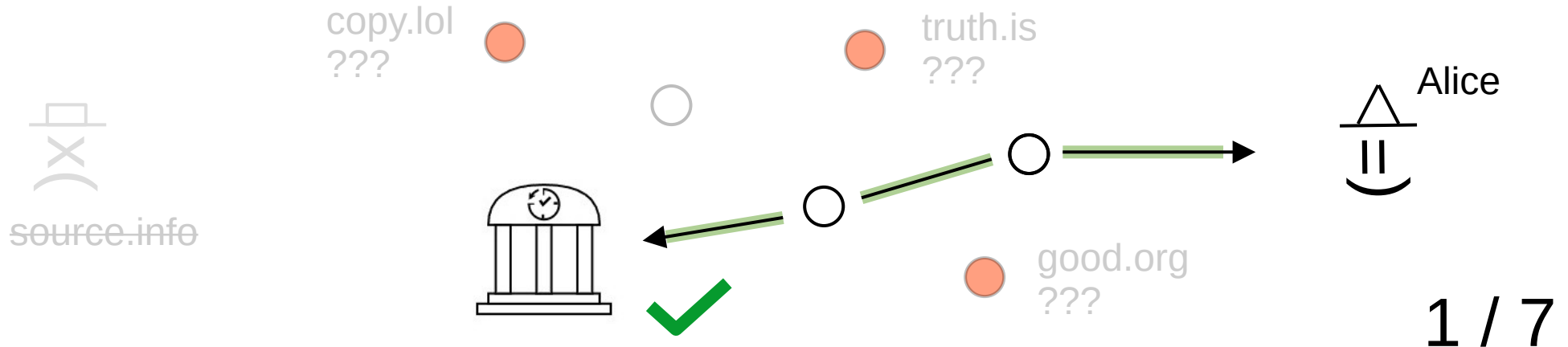
We rely on third-party repositories all the time for availability ;

...and these are often honest with a good reputation! ✓



Introduction

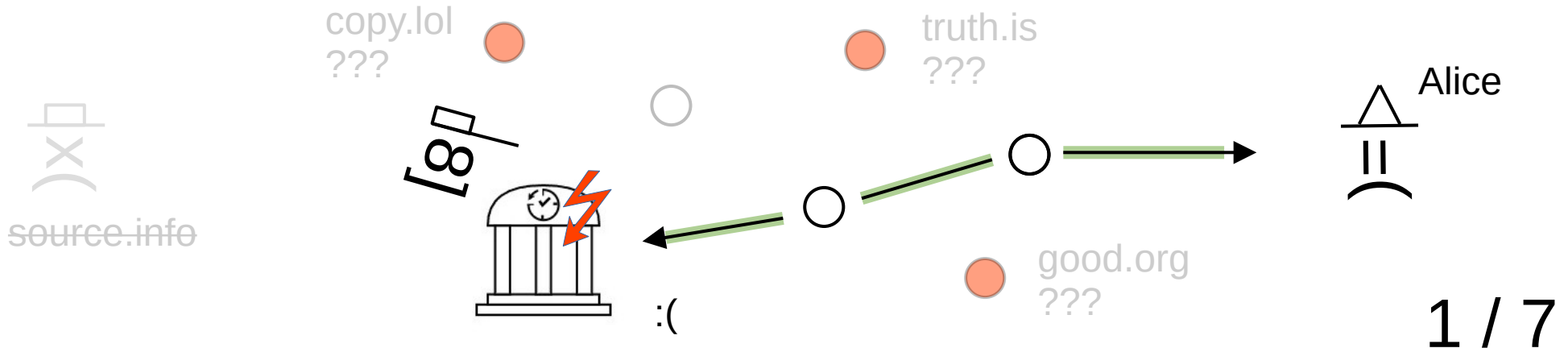
We use archives to distinguish the past from the present ;



Introduction

We use archives to distinguish the past from the present ;

Using archives for trust means bad actors will target them!

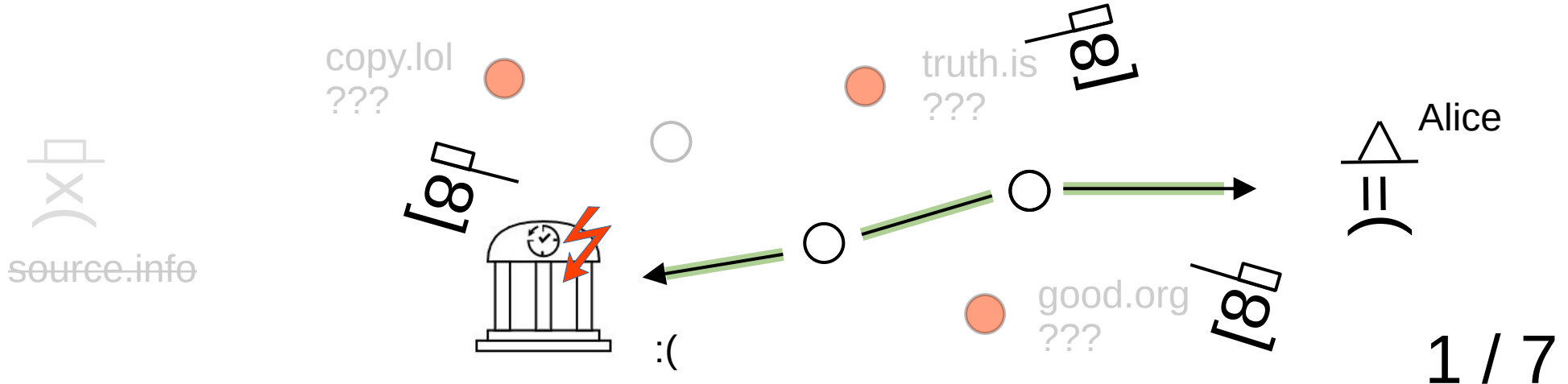


Introduction

We use archives to distinguish the past from the present ;

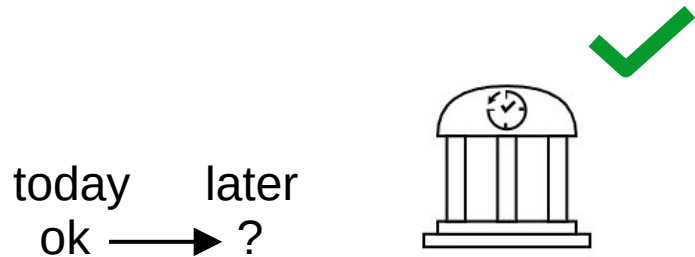
Using archives for trust means bad actors will target them!

(as well as other repositories, big or small) ❌



We need third-party attestation

Archives are still believed to be full of authentic files (for now) ;



We need third-party attestation

Archives are still believed to be full of authentic files (for now) ;

Let's make a list of the files, with their cryptographic digests ;

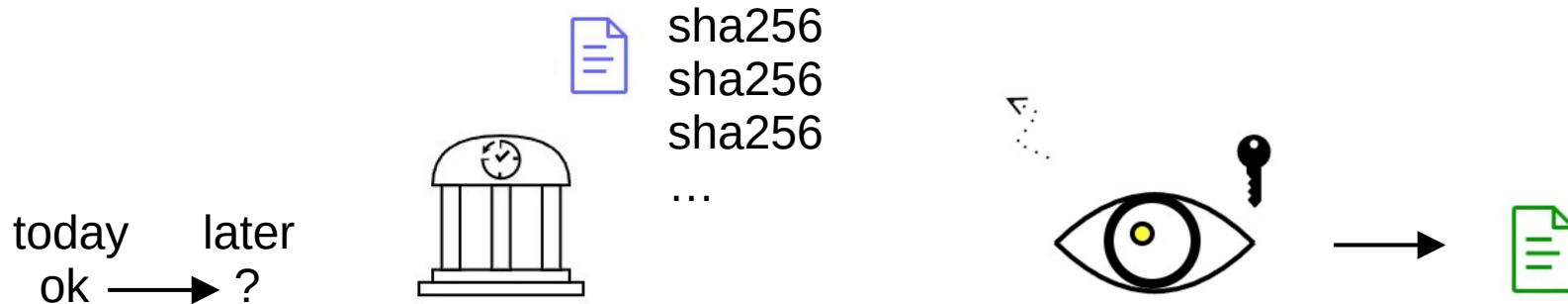


We need third-party attestation

Archives are still believed to be full of authentic files (for now) ;

Let's make a list of the files, with their cryptographic digests ;

...and let *anyone* come and authenticate that list!

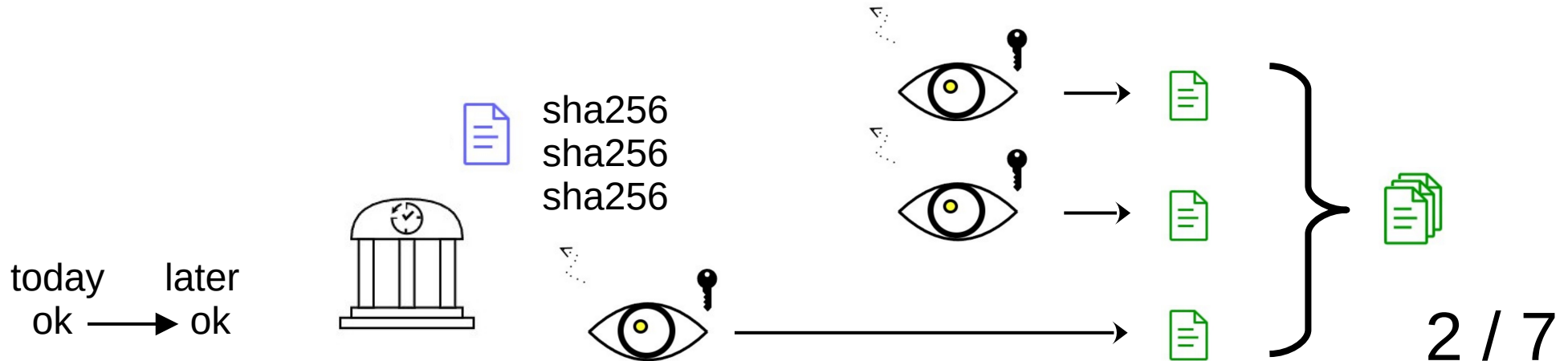


We need *third-party* attestation

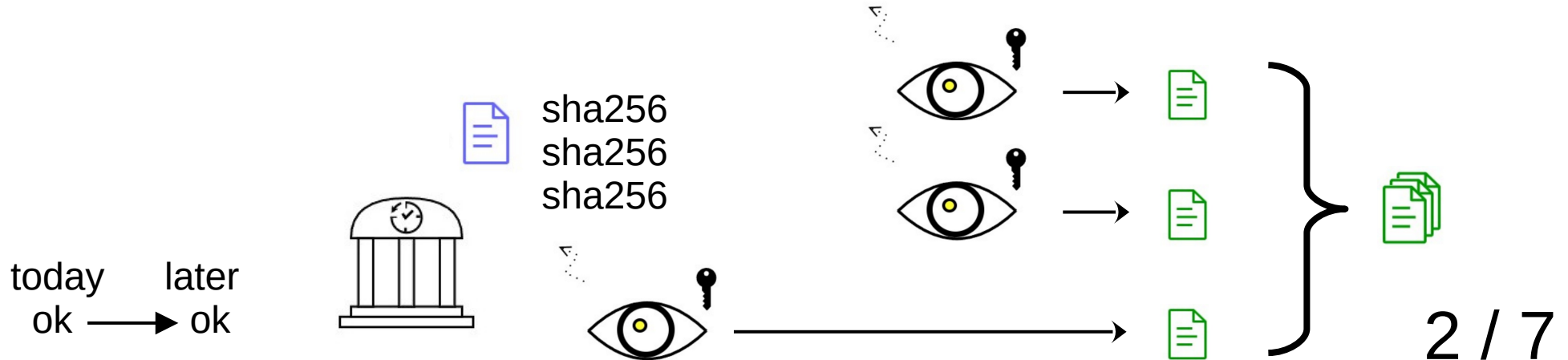
Archives are still believed to be full of authentic files (for now) ;

Let's make a list of the files, with their cryptographic digests ;

...and let *anyone* come and authenticate that list! ✓

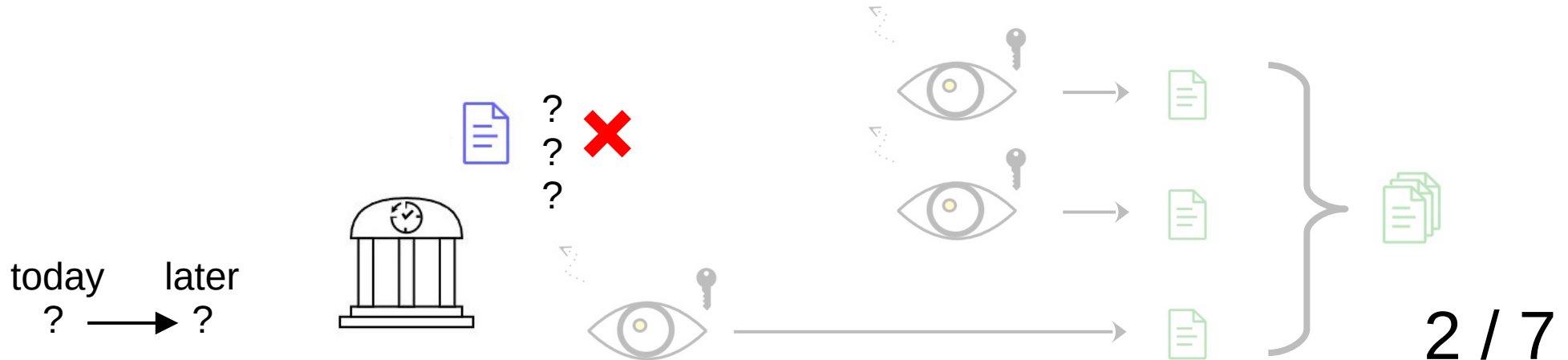


Transparency Logs could be very good for this!



Getting the digests?

Archives sometimes does not publish a list of their digests ;

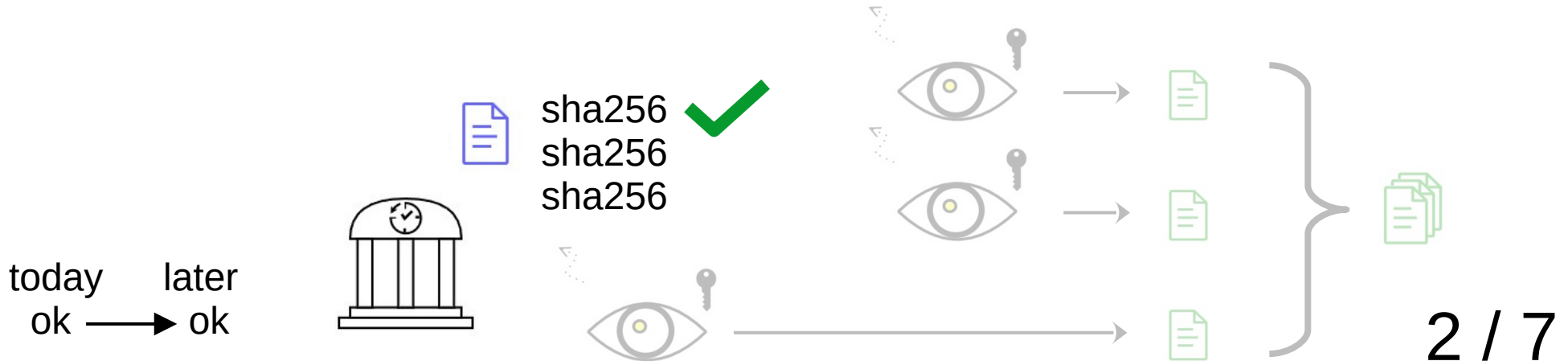


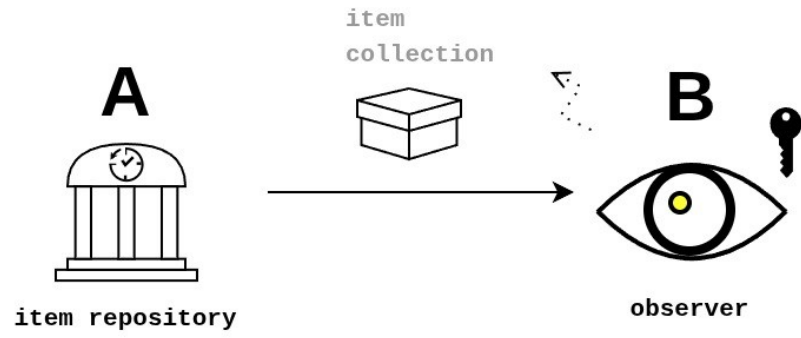
Getting the digests?

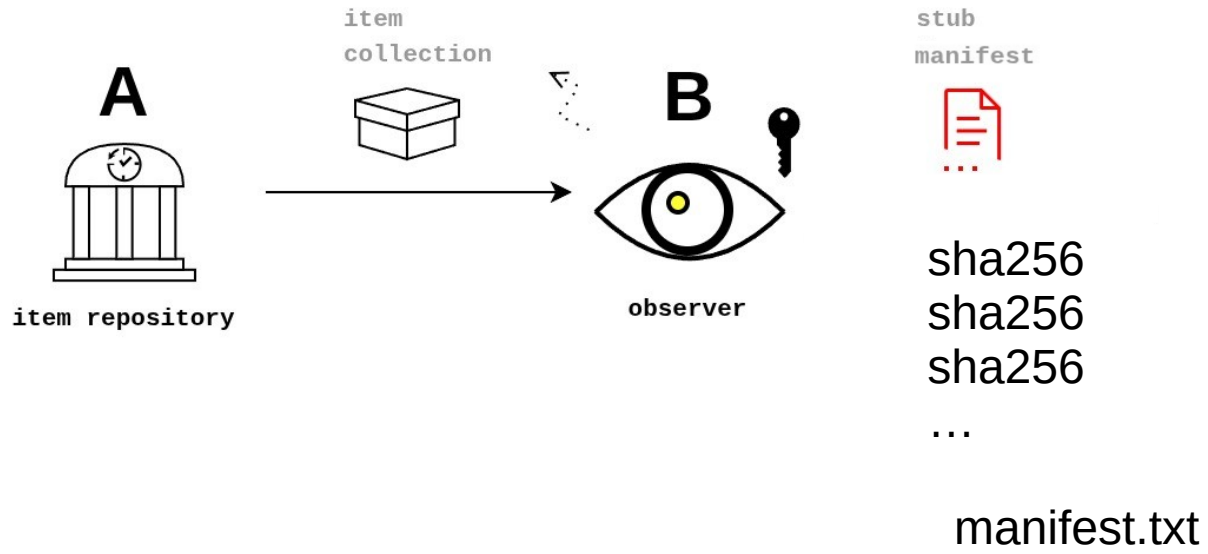
Archives sometimes does not publish a list of their digests ;

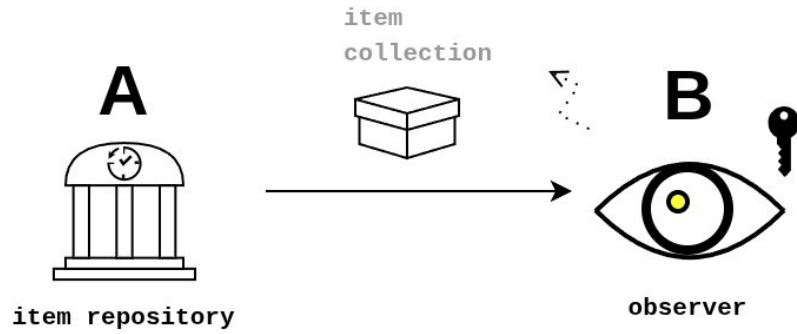
Tell your archivist friend to publish their integrity :)

(any format, unauthenticated, is fine)









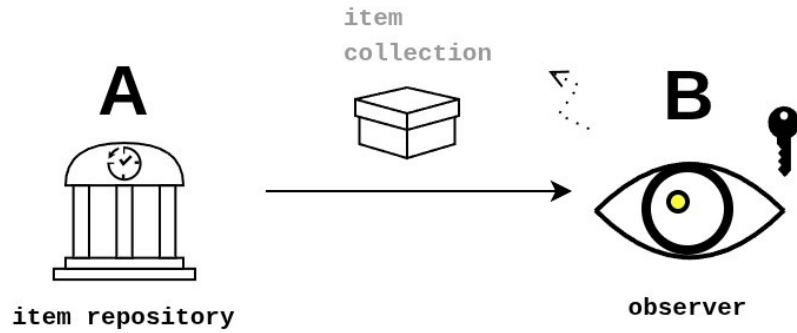
stub
manifest



sha256
sha256
sha256
...

manifest.txt
+manifest.txt.sig

“I saw these existing items”



stub
manifest



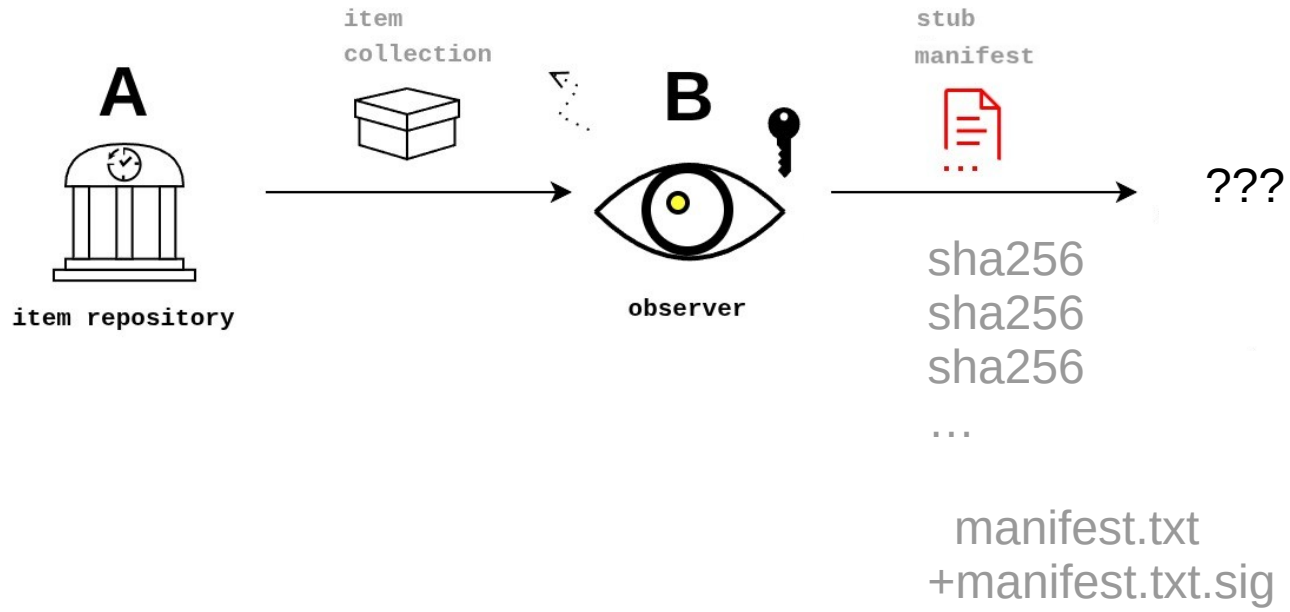
sha256
sha256
sha256
...

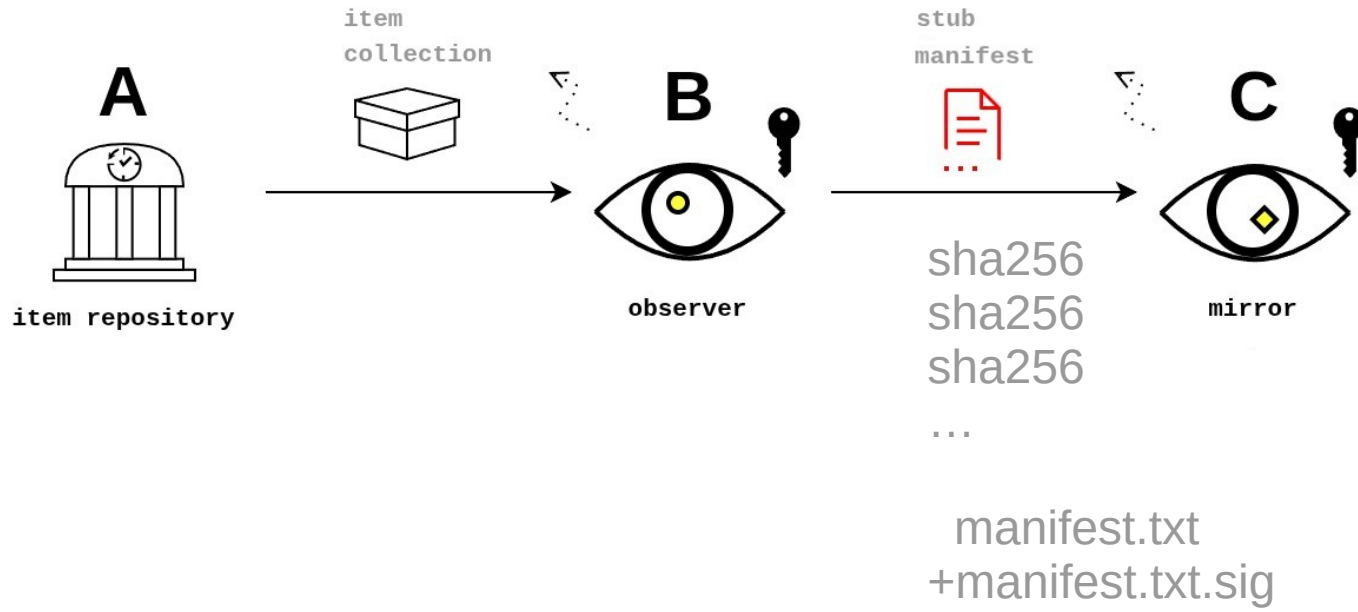
observer identity issuer:

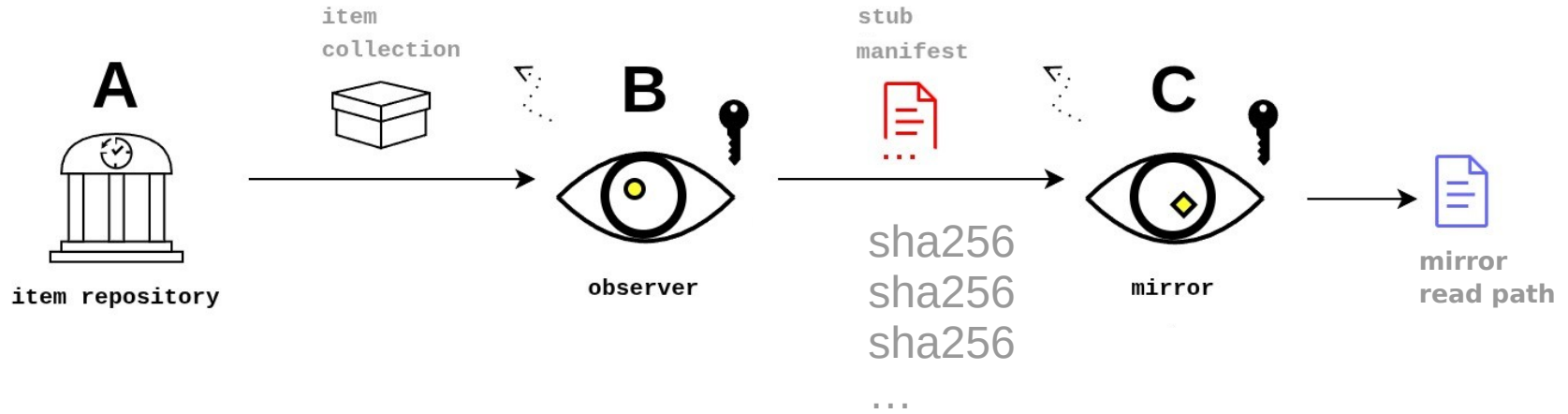
- OpenPGP (current)
- Fulcio (would like)
- ...

manifest.txt
+manifest.txt.sig

"I saw these existing items"

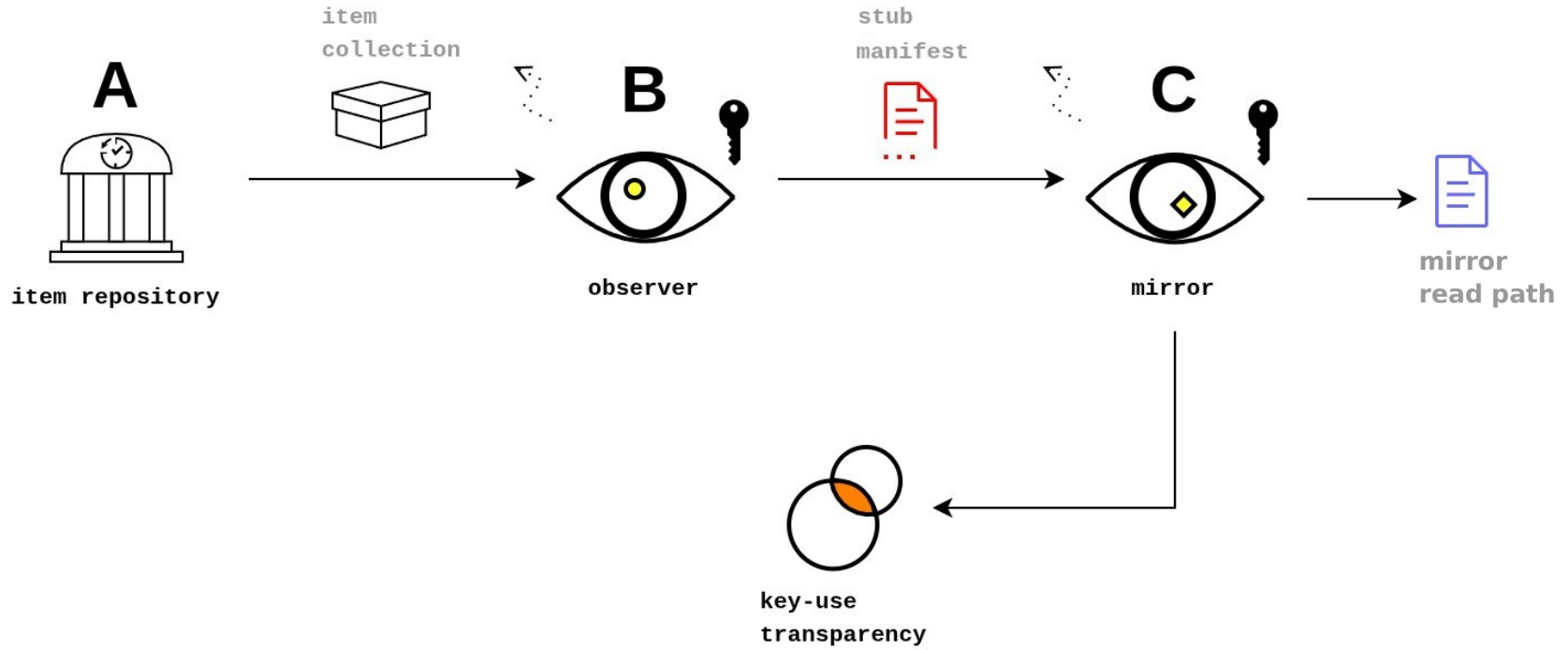






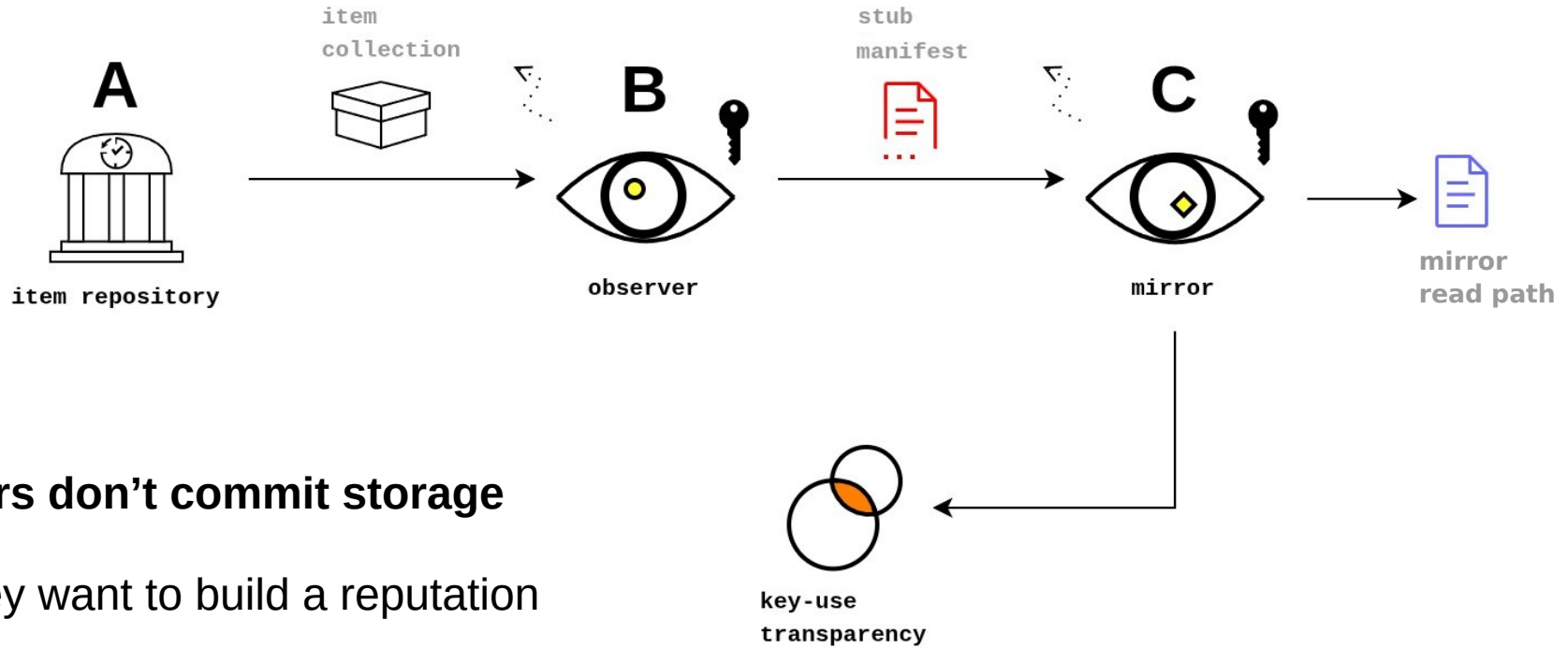
manifest.txt
+manifest.txt.sig
+manifest.txt.cosig

“I keep this manifest online”



“I keep this manifest online”

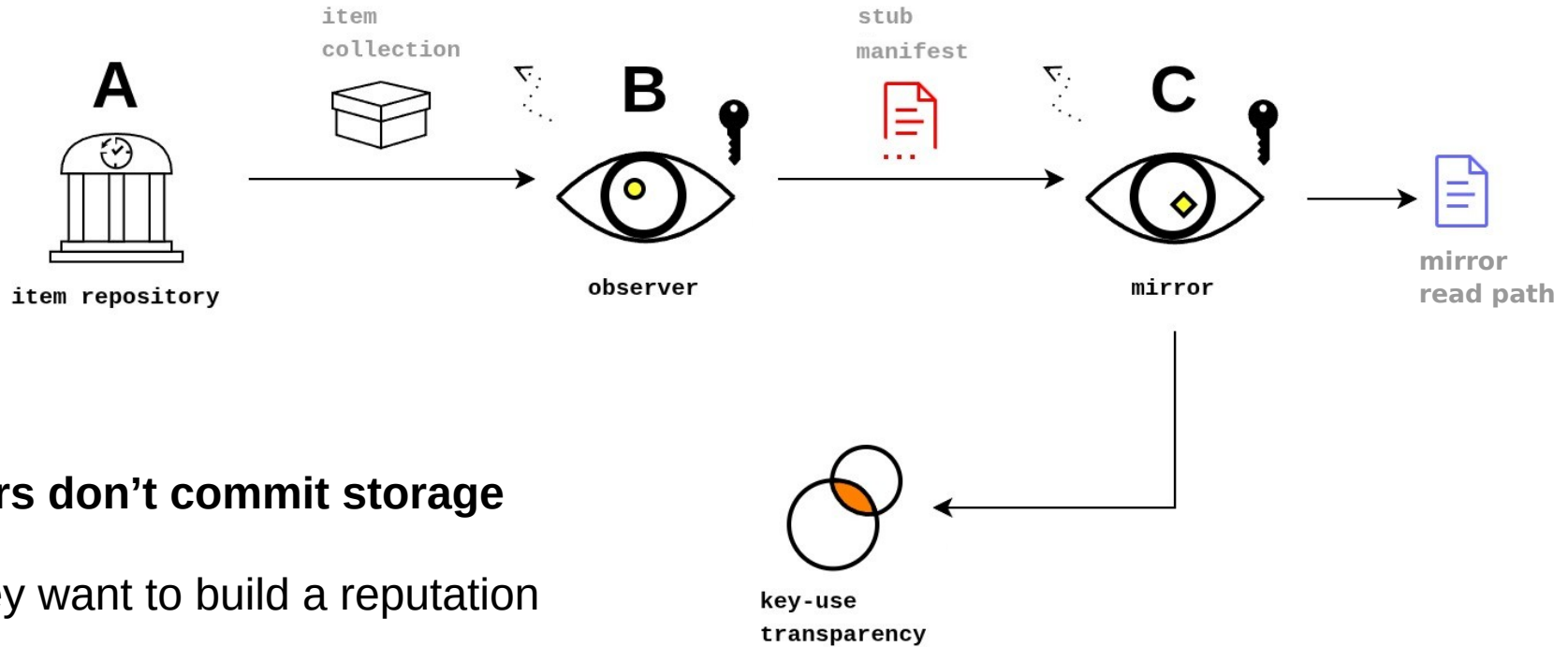
> mirror log this commitment



observers don't commit storage
...but they want to build a reputation

"I keep this manifest online"

> mirror log this commitment



observers don't commit storage

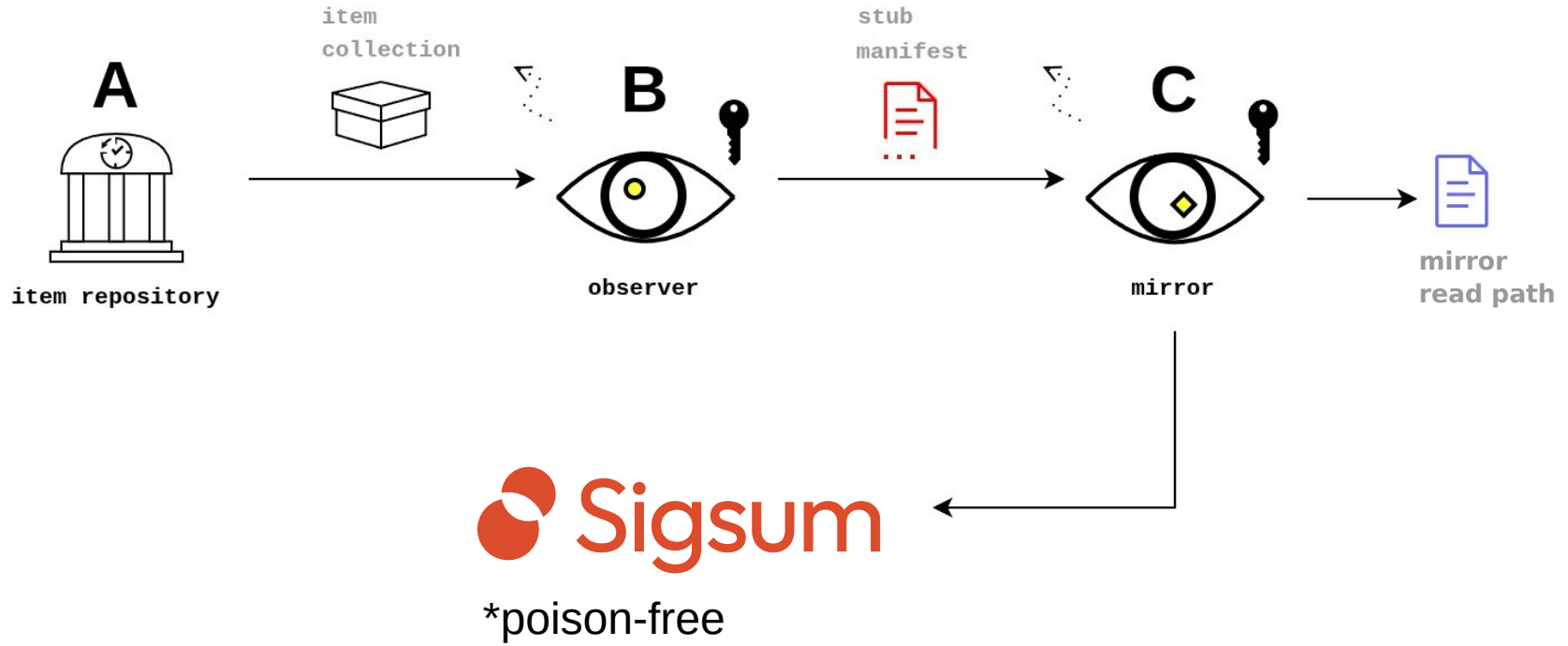
...but they want to build a reputation

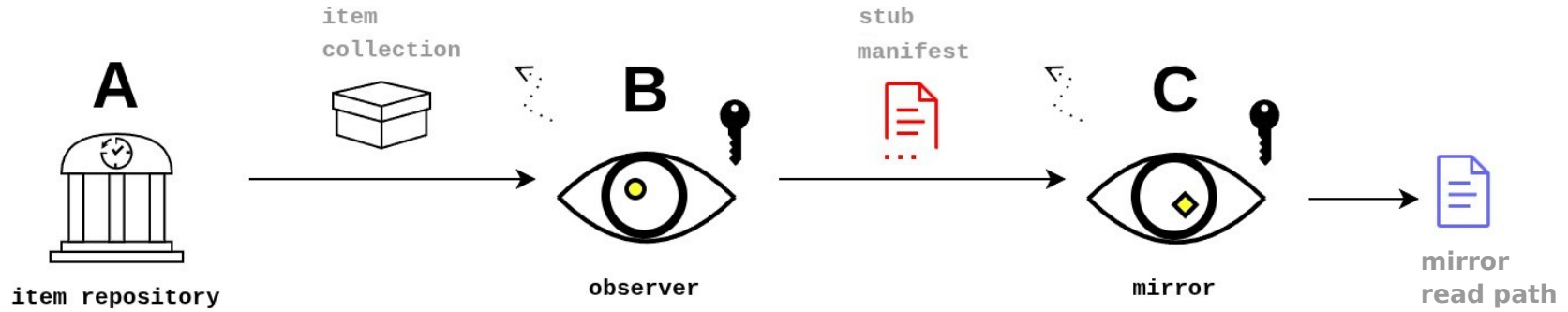
...thus they want monitoring

But who keeps the manifest published?

"I keep this manifest online"

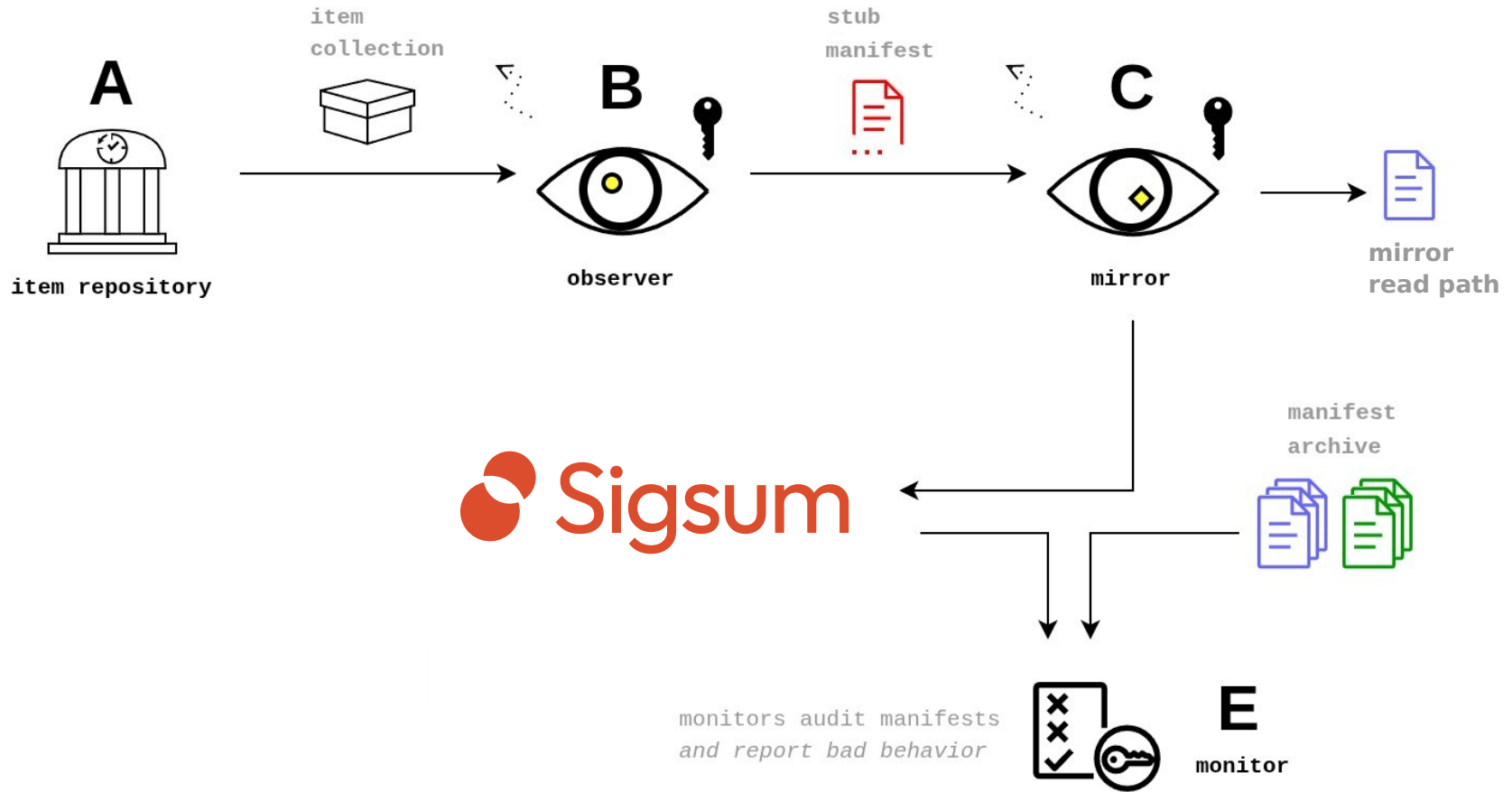
> mirror log this commitment

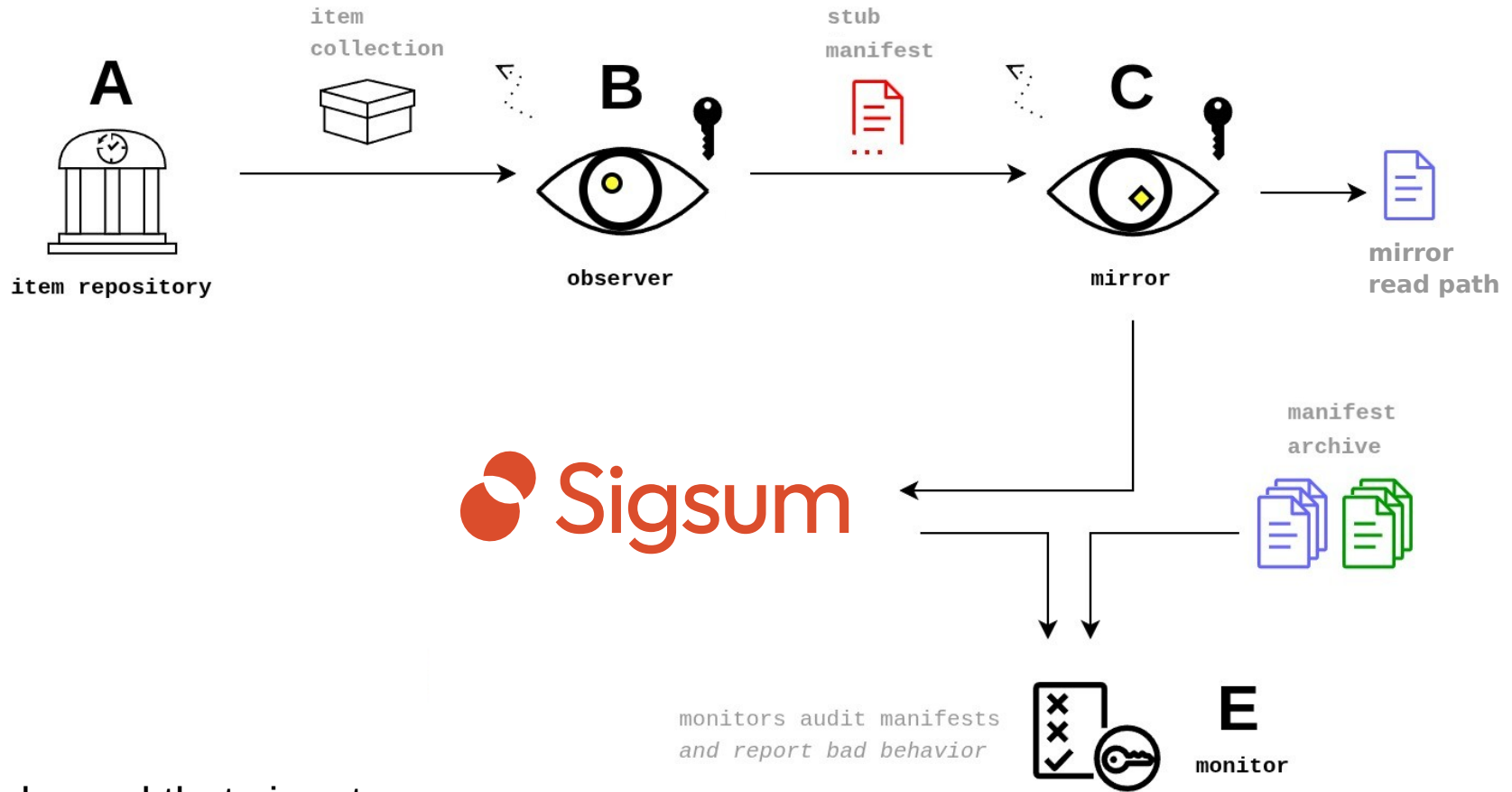




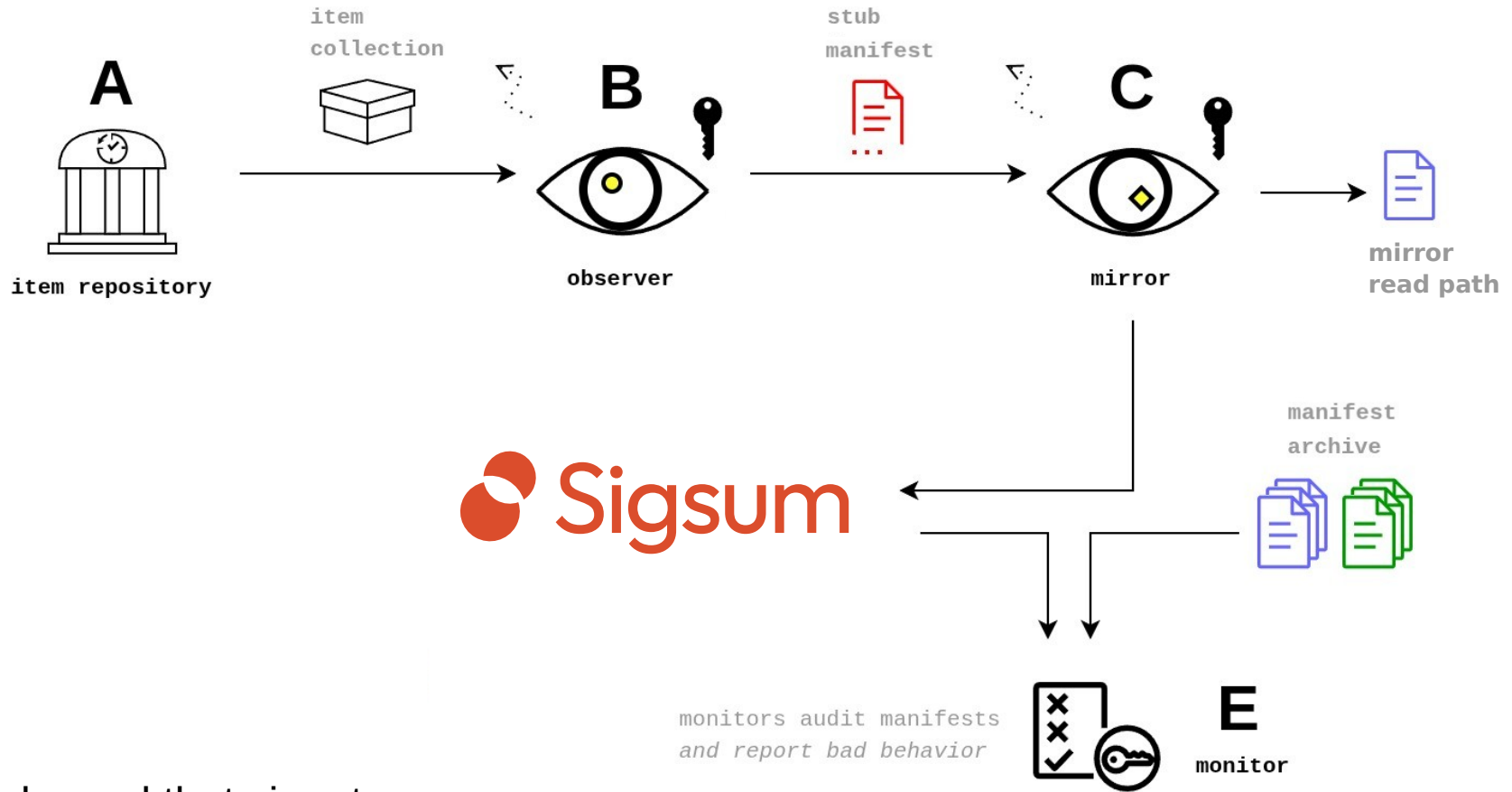
+ content-addressable storage
(mirror can get poison)

manifest hash <> manifest bytes



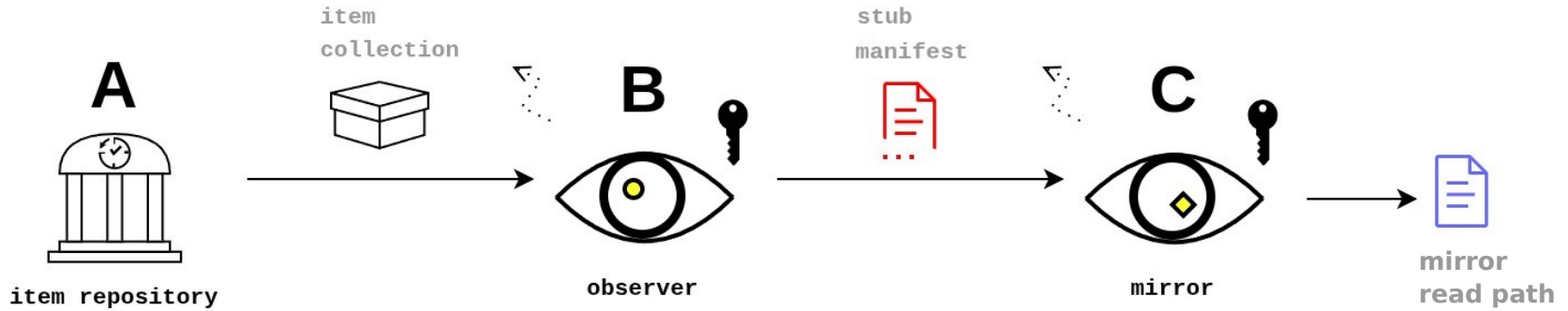


“you logged that signature...”



“you logged that signature...

...please show me the manifest behind it :)”

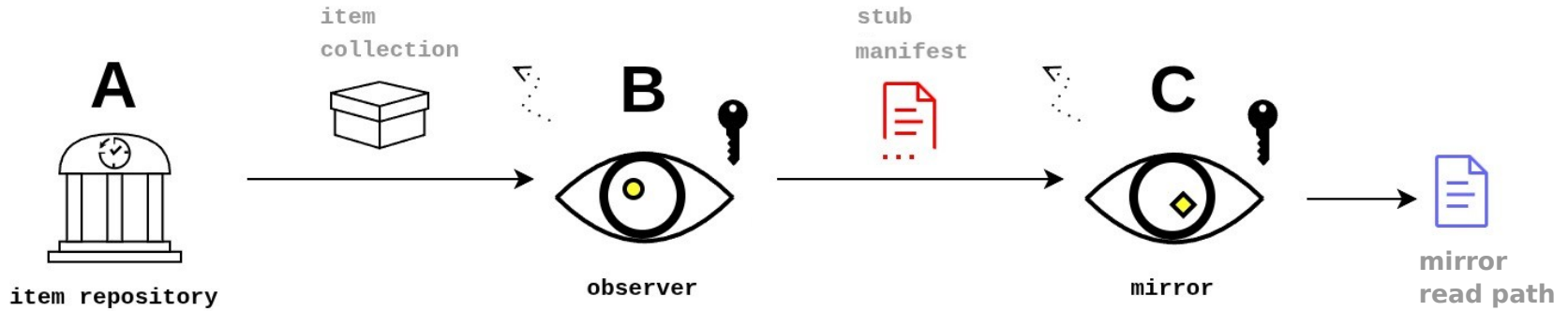


Monitor will check:
1/ mirror keeping manifests online



“you logged that signature...

...please show me the manifest behind it :)”



Monitor will check:

- 1/ mirror keeping manifests online
- 2/ story on the manifest content

example: detect identity misuse

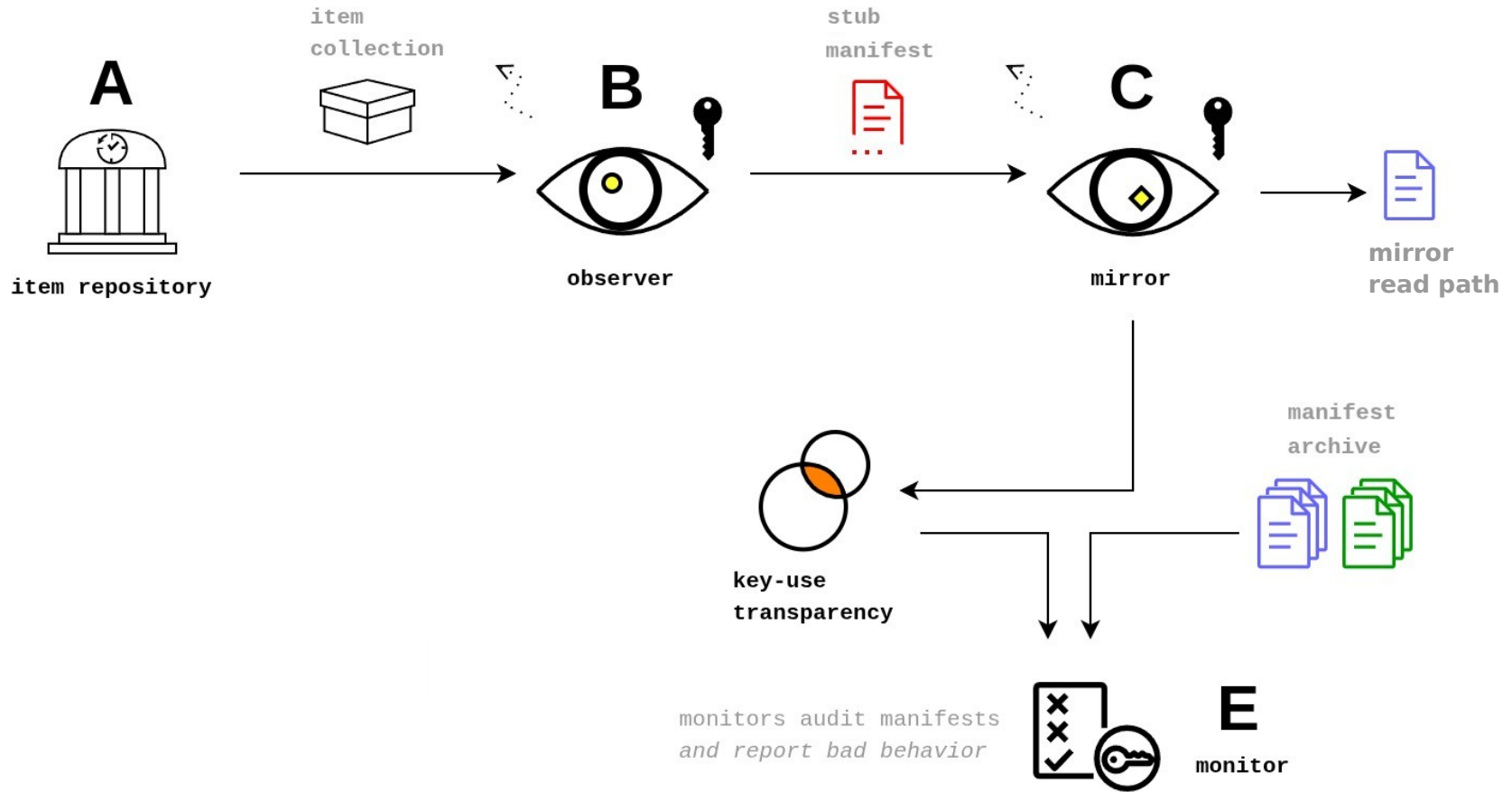


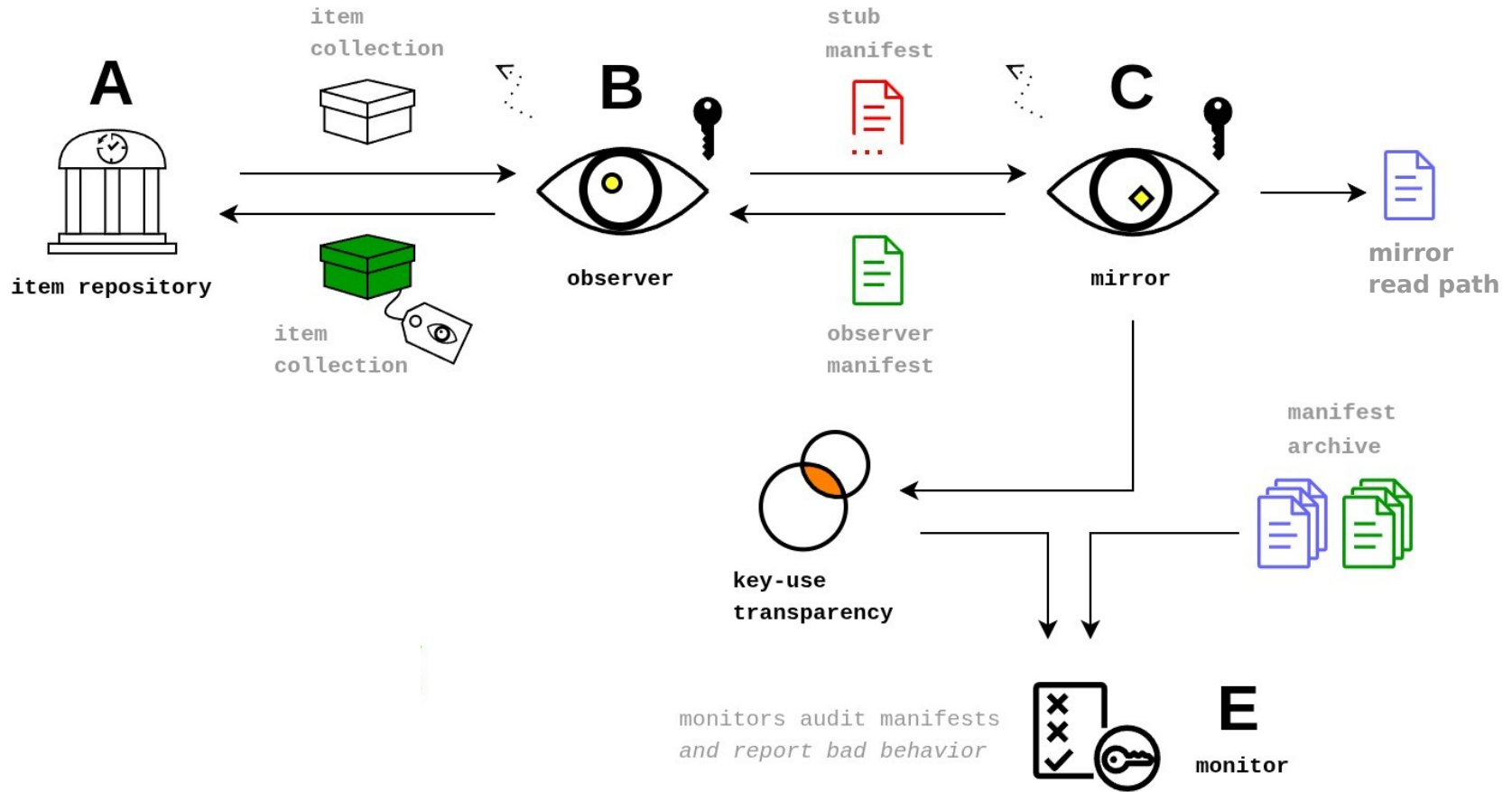
monitors audit manifests
and report bad behavior

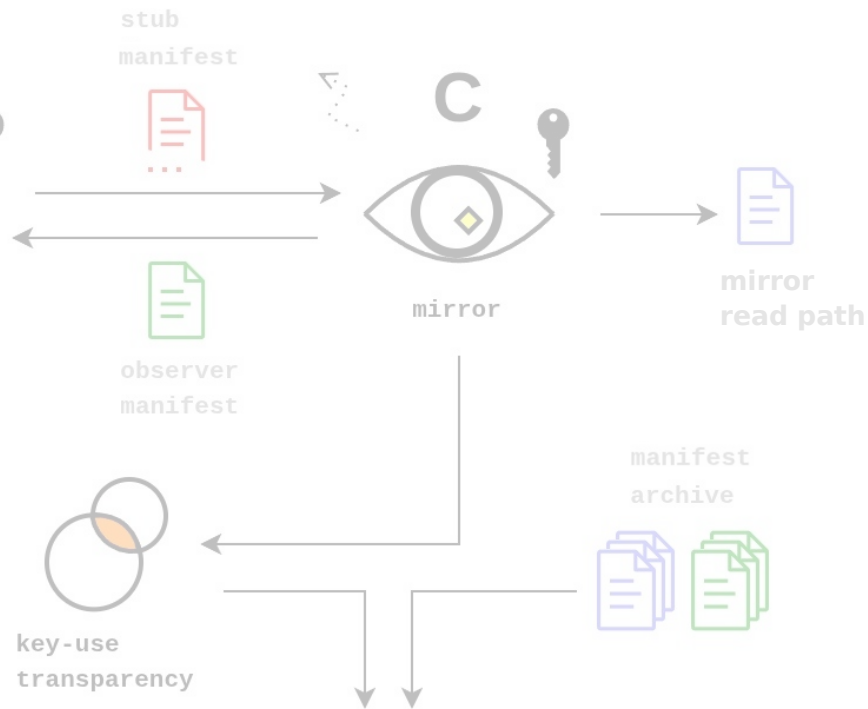
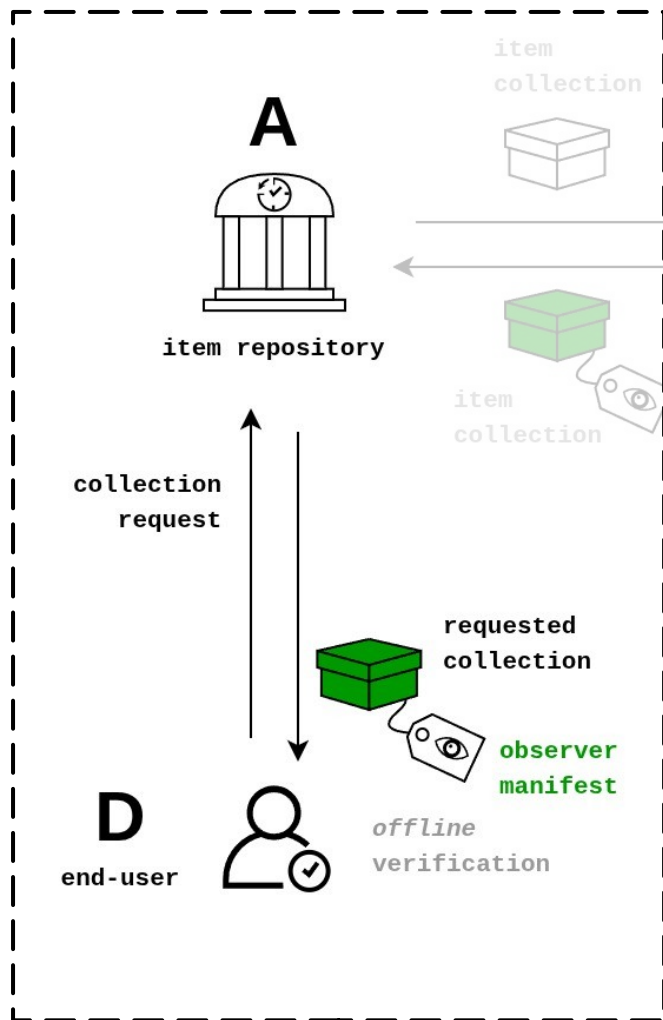


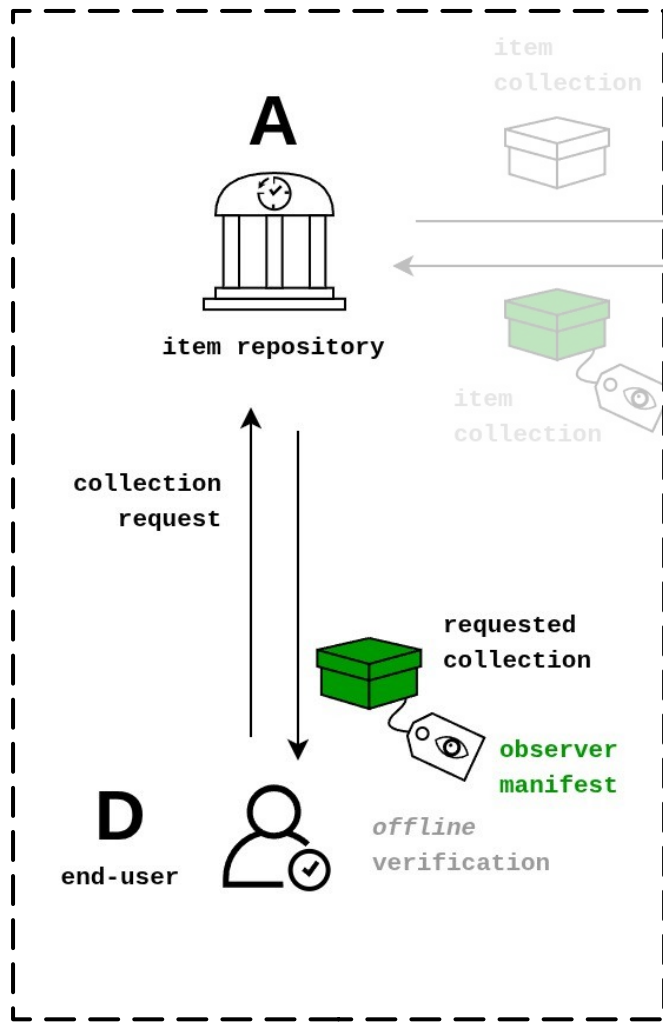
“you logged that signature...

...please show me the manifest behind it :)”



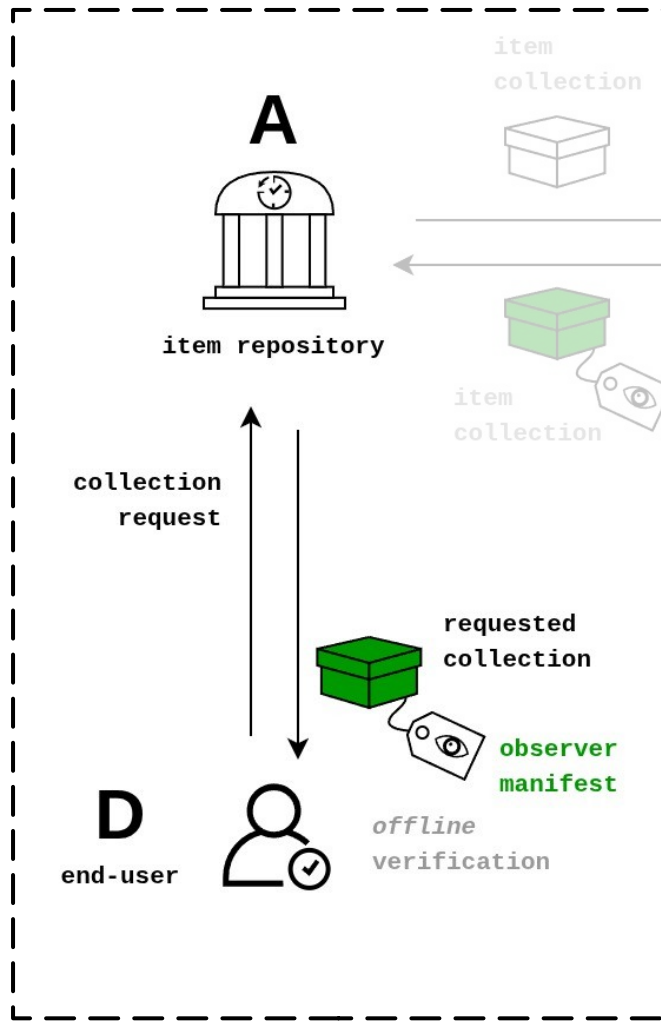






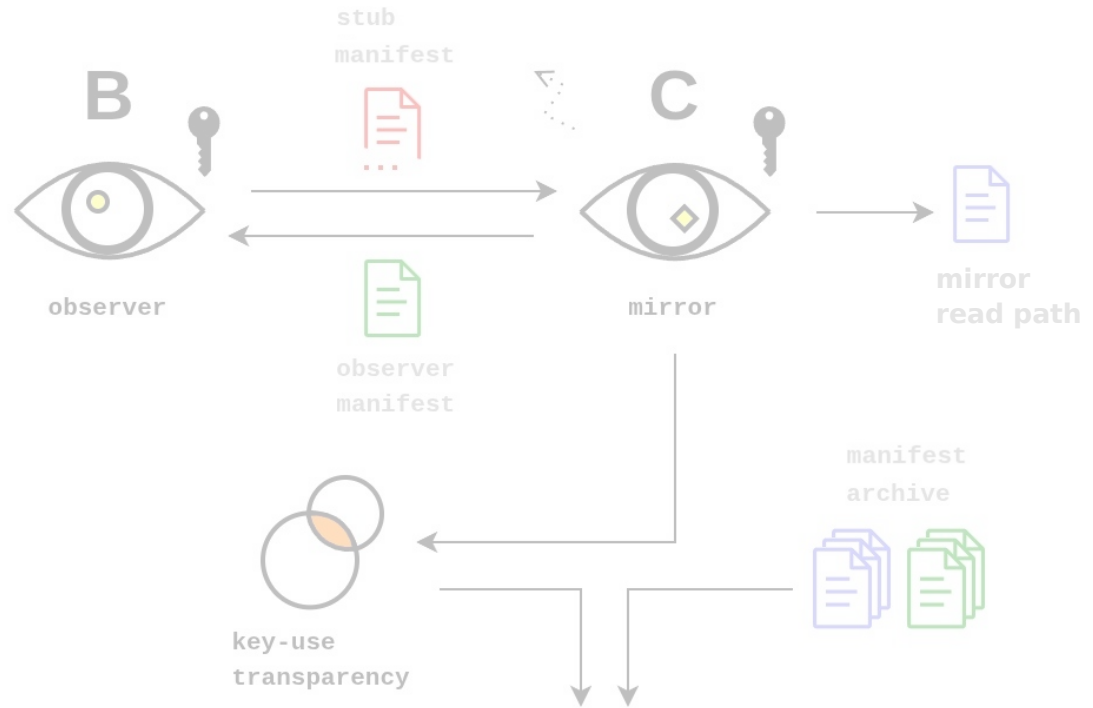
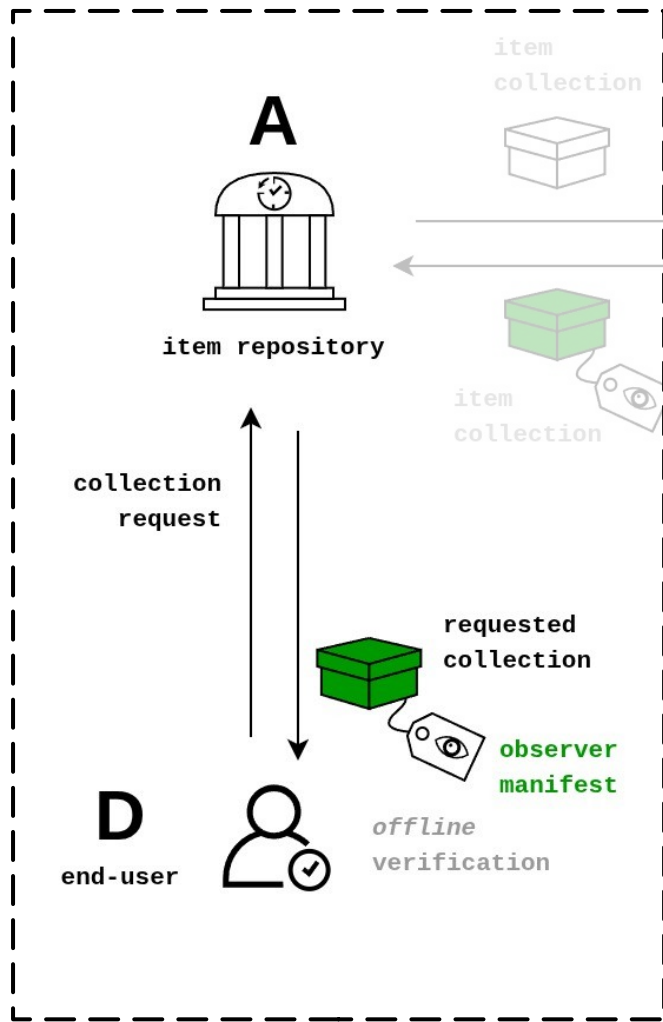
end-user learns:

1/ observer X signed manifest and saw items Y



end-user learns:

- 1/ observer X signed manifest and saw items Y
- 2/ manifest born public and was available for everyone

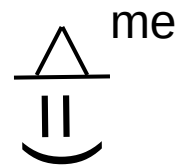


end-user learns:

- 1/ observer X signed manifest and saw items Y
- 2/ manifest born public and was available for everyone

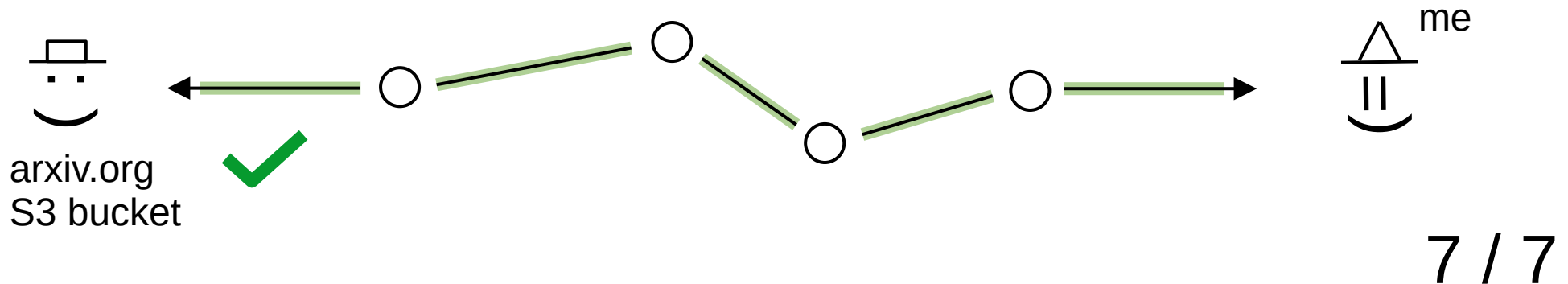
+quorum of witnesses signing time as reference!

not finished & not deployed (yet)



not finished & not deployed (yet)

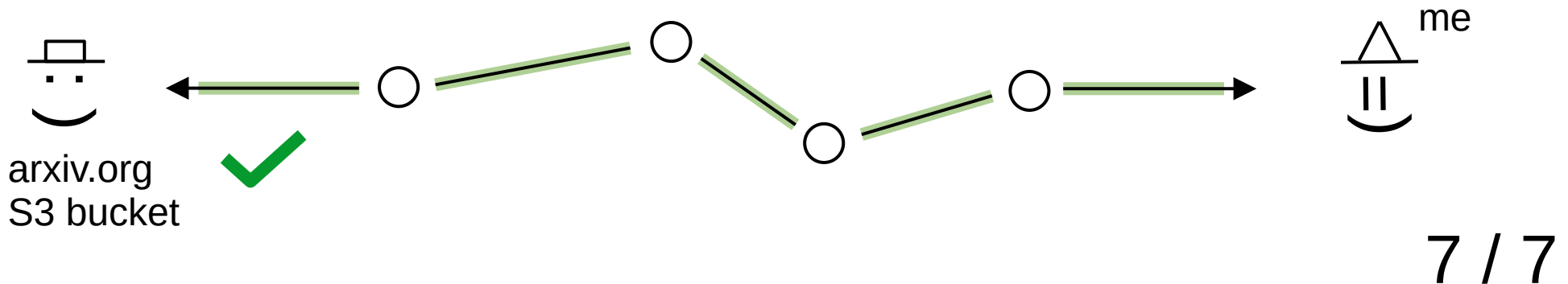
See test.archive.rip/arxiv for ~3M papers in ~150 manifests ;



not finished & not deployed (yet)

See test.archive.rip/arxiv for ~3M papers in ~150 manifests ;

Logs stay lightweight & predictable costs for mirrors ;

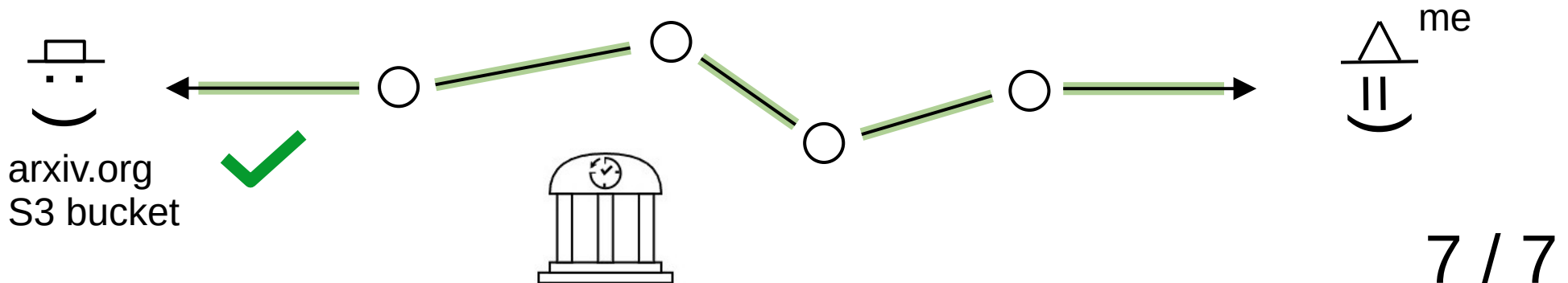


not finished & not deployed (yet)

See test.archive.rip/arxiv for ~3M papers in ~150 manifests ;

Logs stay lightweight & predictable costs for mirrors ;

Should scale to *billions* of items ;
(a whole national archive)

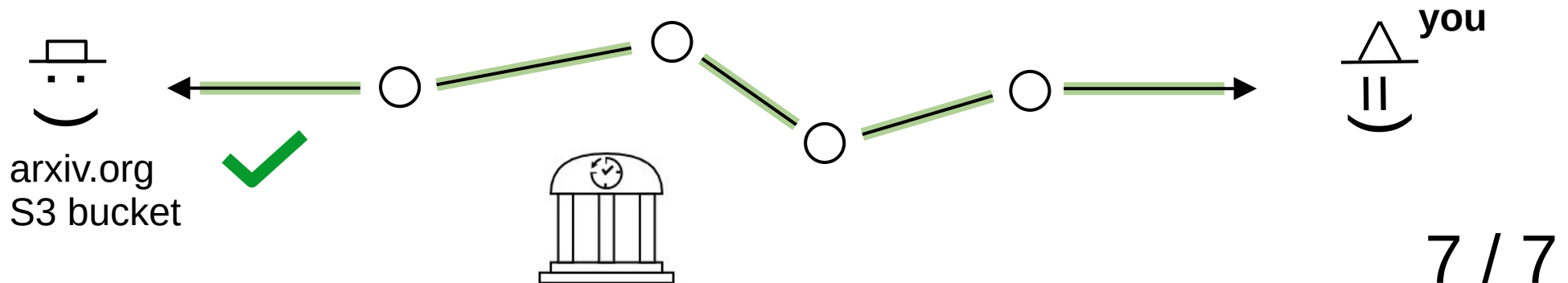


not finished & not deployed (yet)

See test.archive.rip/arxiv for ~1.5M papers in ~150 manifests ;

Logs stay lightweight & predictable costs for mirrors ;

Should scale to *billions* of items ; at home, on your NAS :)
(a whole national archive)



not finished & not deployed (yet)

See test.archive.rip/arxiv for ~3M papers in ~150 manifests ;

Logs stay lightweight & predictable costs for mirrors ;

Should scale to *billions* of items ; at home, on your NAS :)
(a whole national archive)

