



Tools for Creating and Validating SIPs

Reducing the Technological Know-How of Preparing Content for Digital Preservation

Anni Järvenpää
Finnish National Digital Preservation Service
CSC – IT Center for Science Ltd



Non-profit state
organization with
special tasks



Turn over
in 2019
51 M€



Headquarters in
Espoo,
datacenter in
Kajaani



Owned by state **(70%)**
and all Finnish higher education
institutions **(30%)**



Circa
395
employees
in 2019

Our customers



Researchers,
research institutes and
organizations

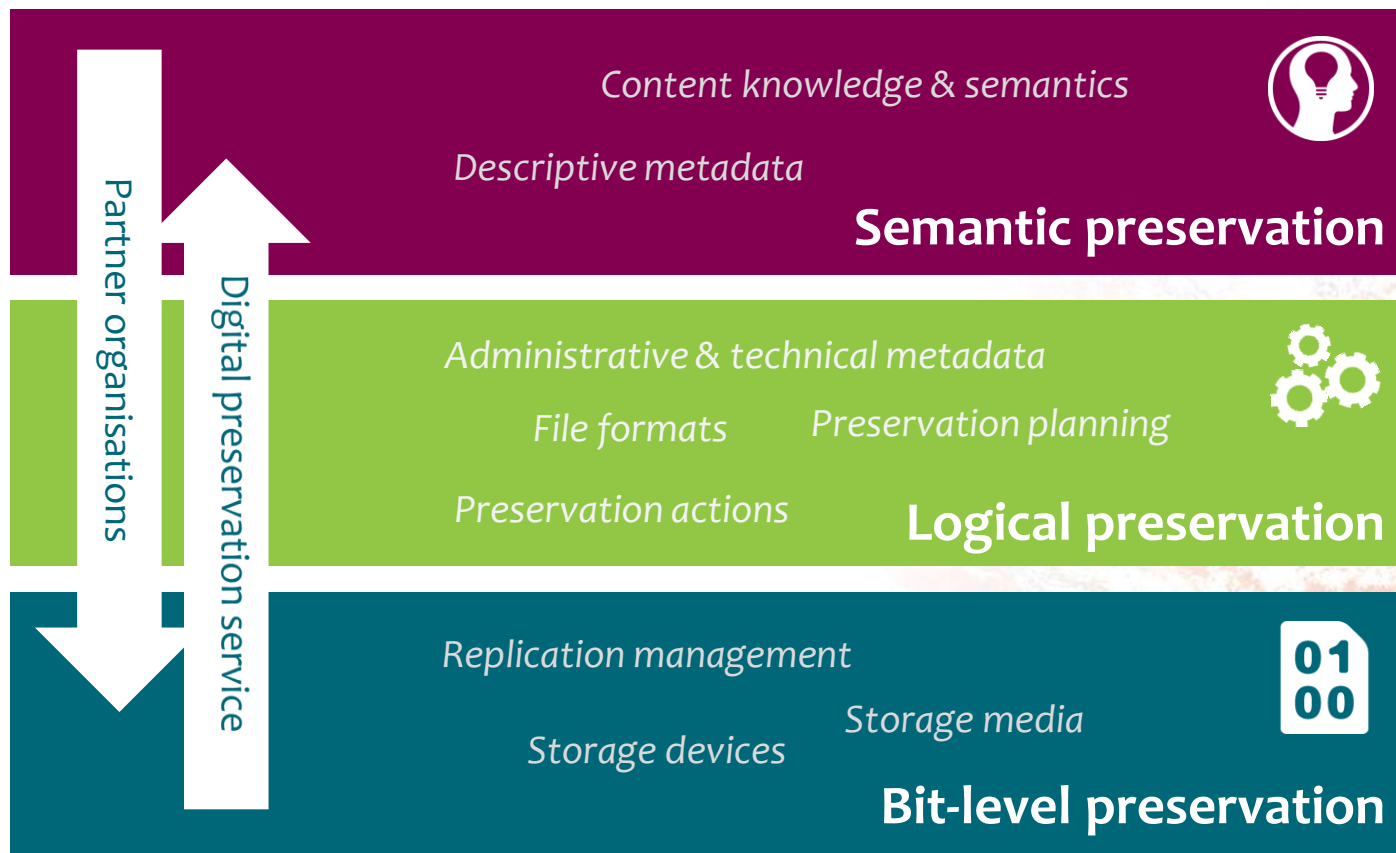


Organizations providing
education

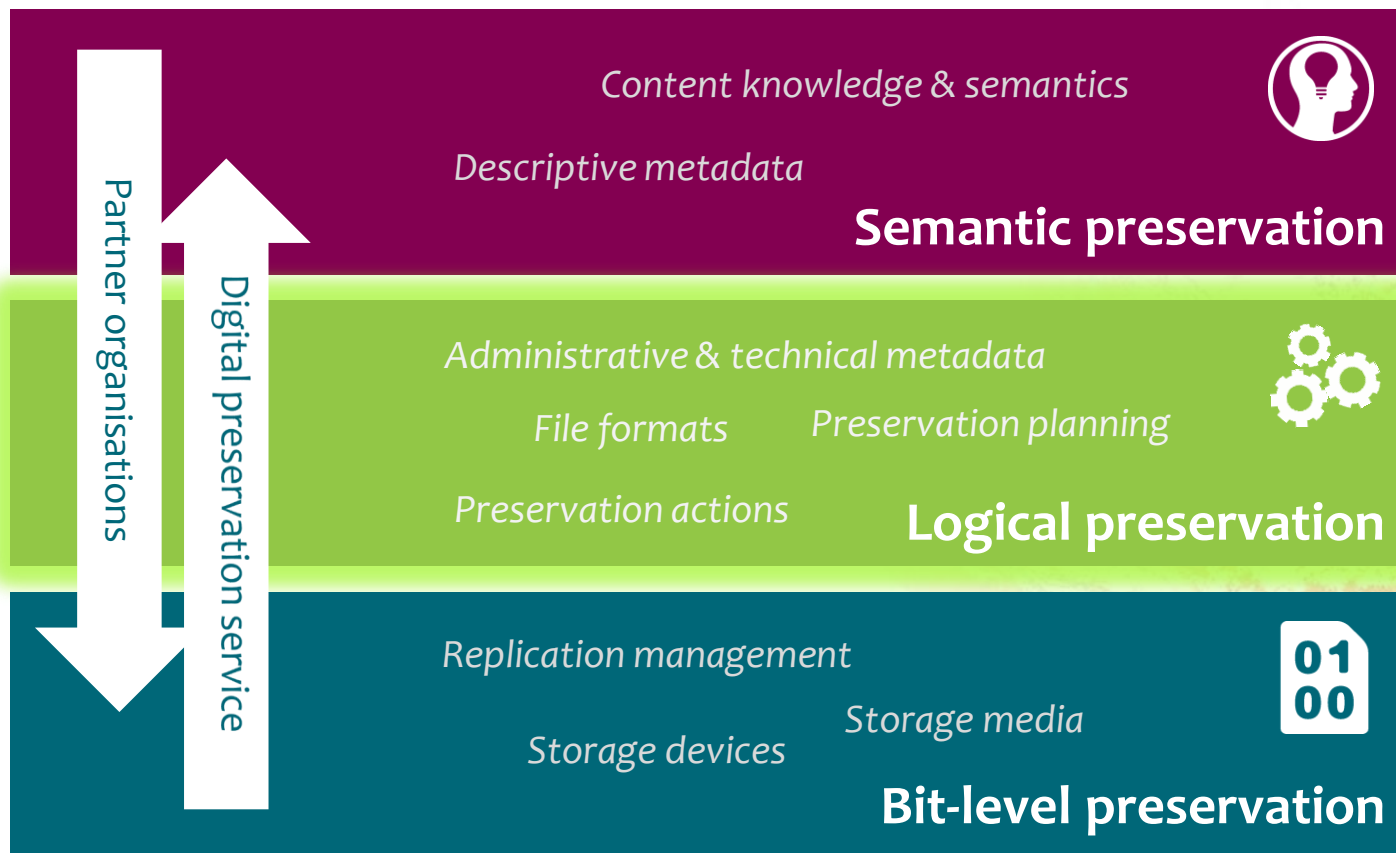


Memory organizations,
state and public
organizations

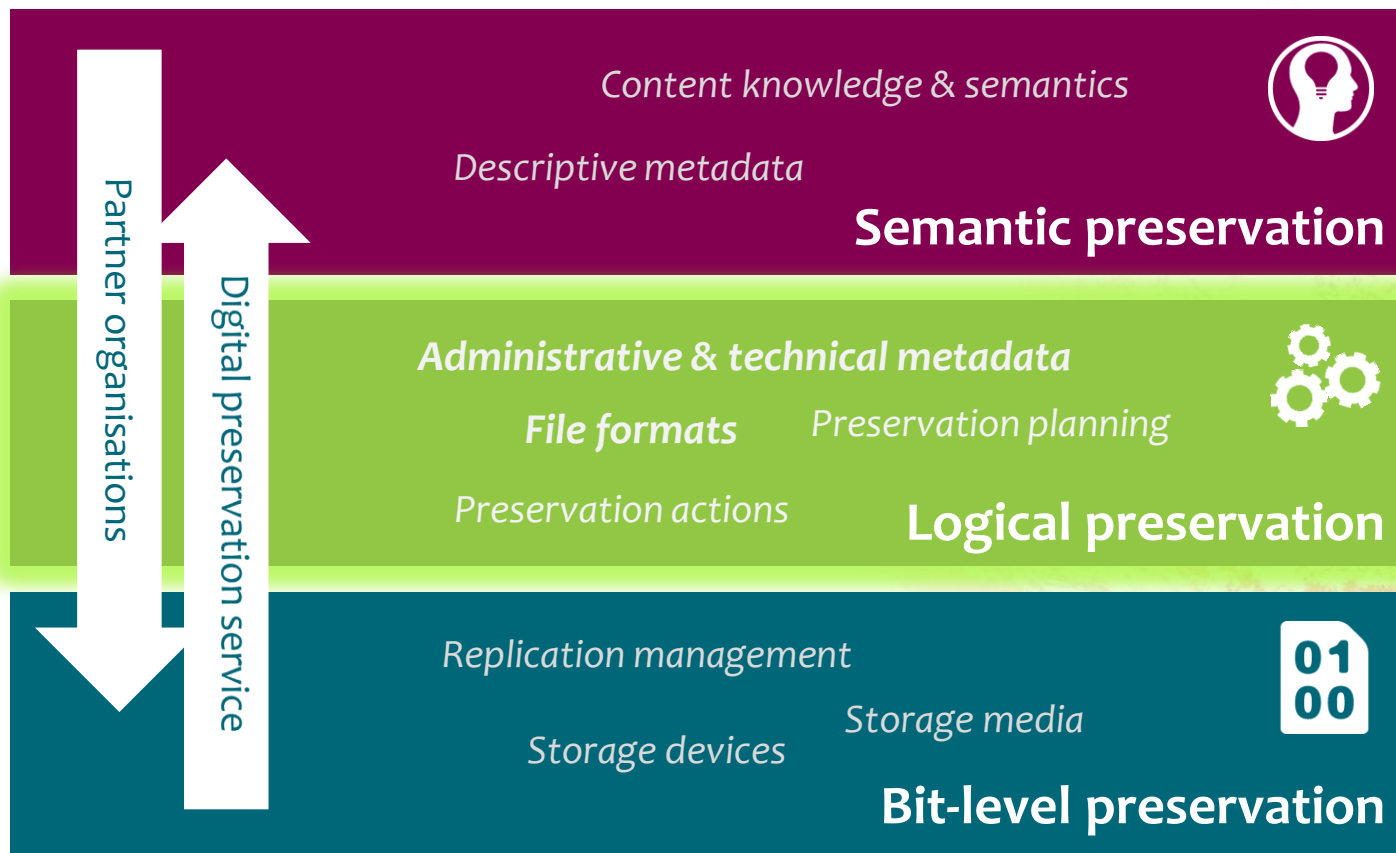
Focus of This Talk



Focus of This Talk



Focus of This Talk



What We Needed



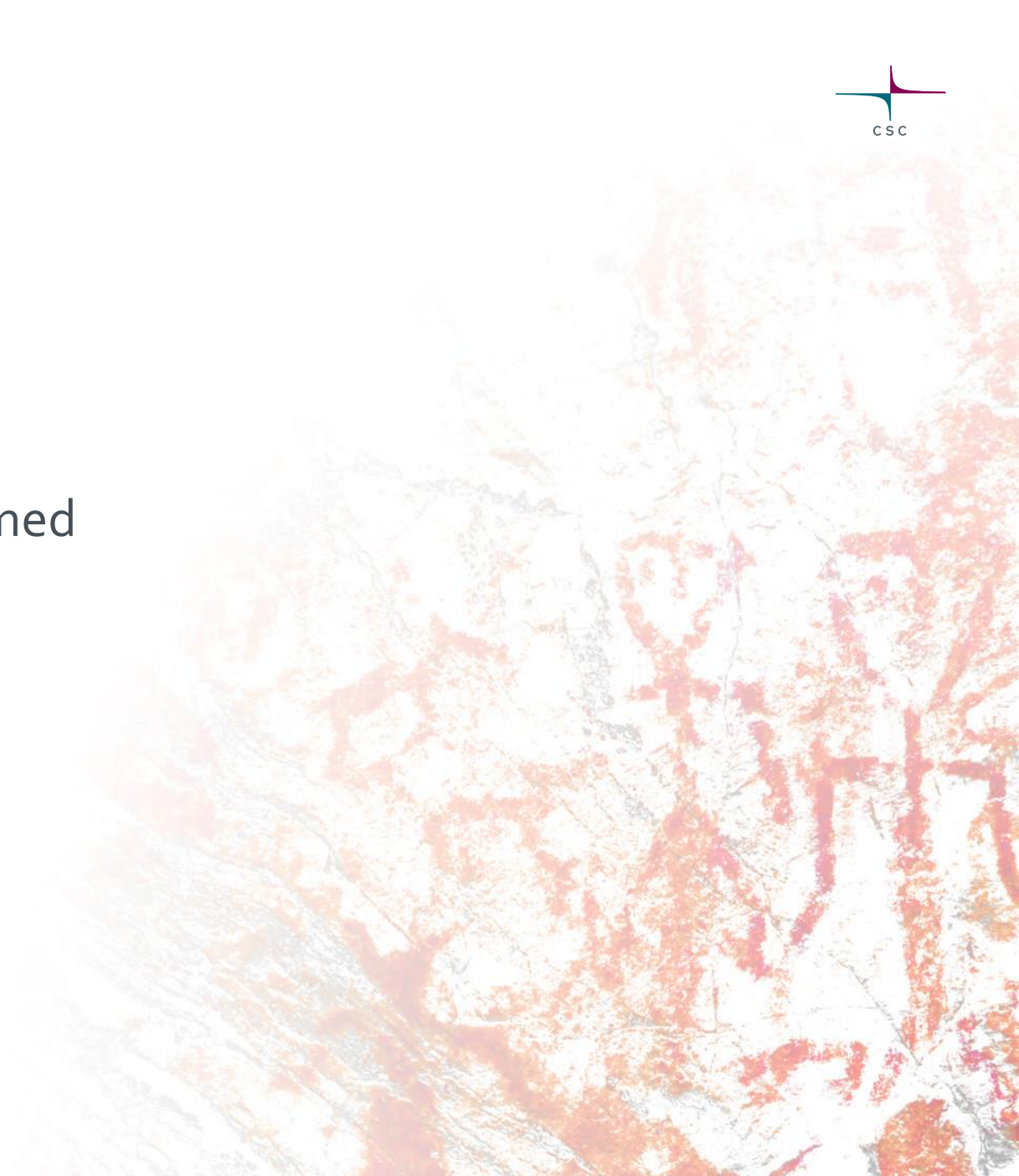
What We Needed

- A tool that can identify file formats
- A tool that can extract metadata from files



What We Needed

- A tool that can identify file formats
- A tool that can extract metadata from files
- A tool that can determine if a file is well-formed

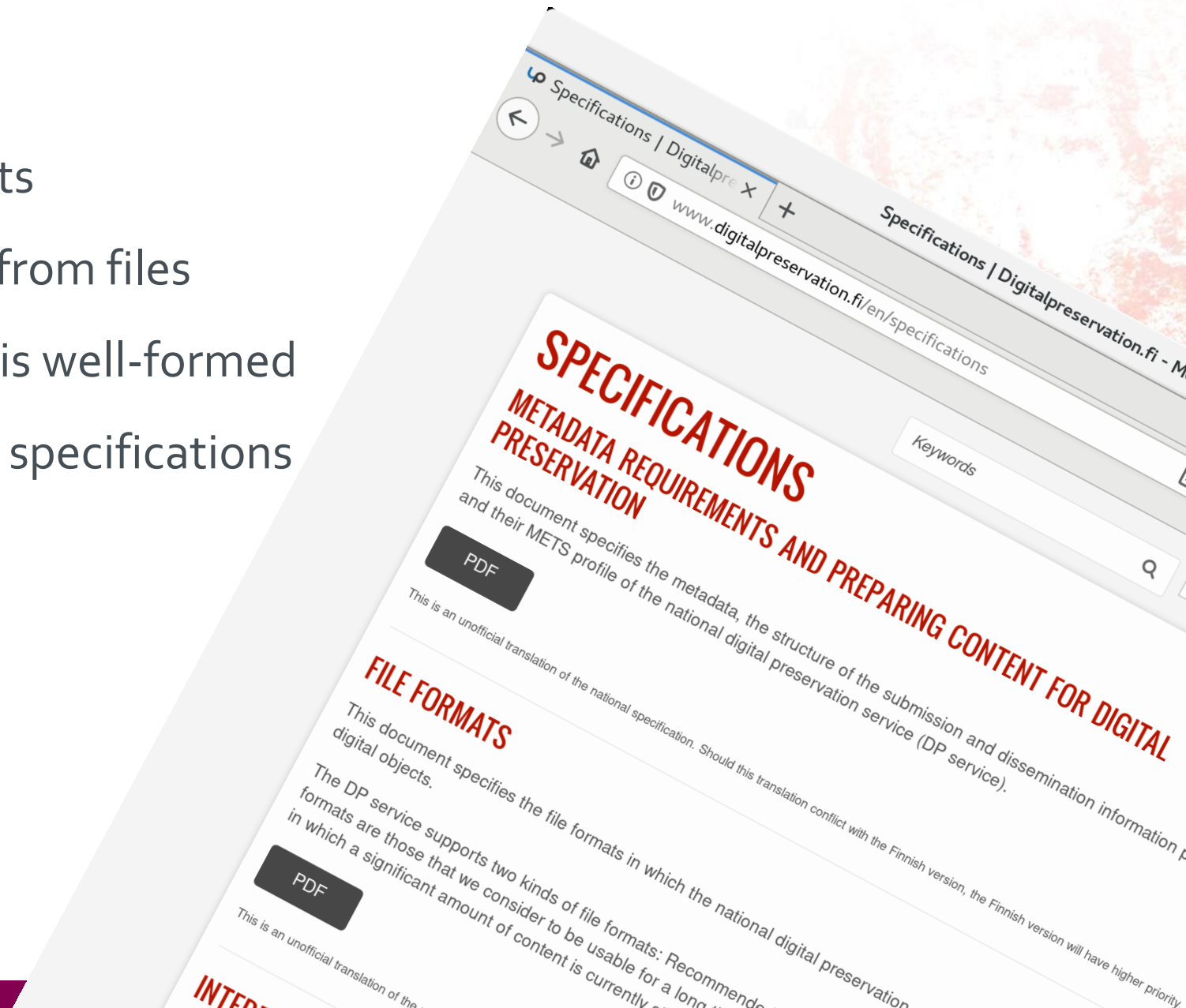


What We Needed

- A tool that can identify file formats
- A tool that can extract metadata from files
- A tool that can determine if a file is well-formed
- A tool whose output matches our specifications

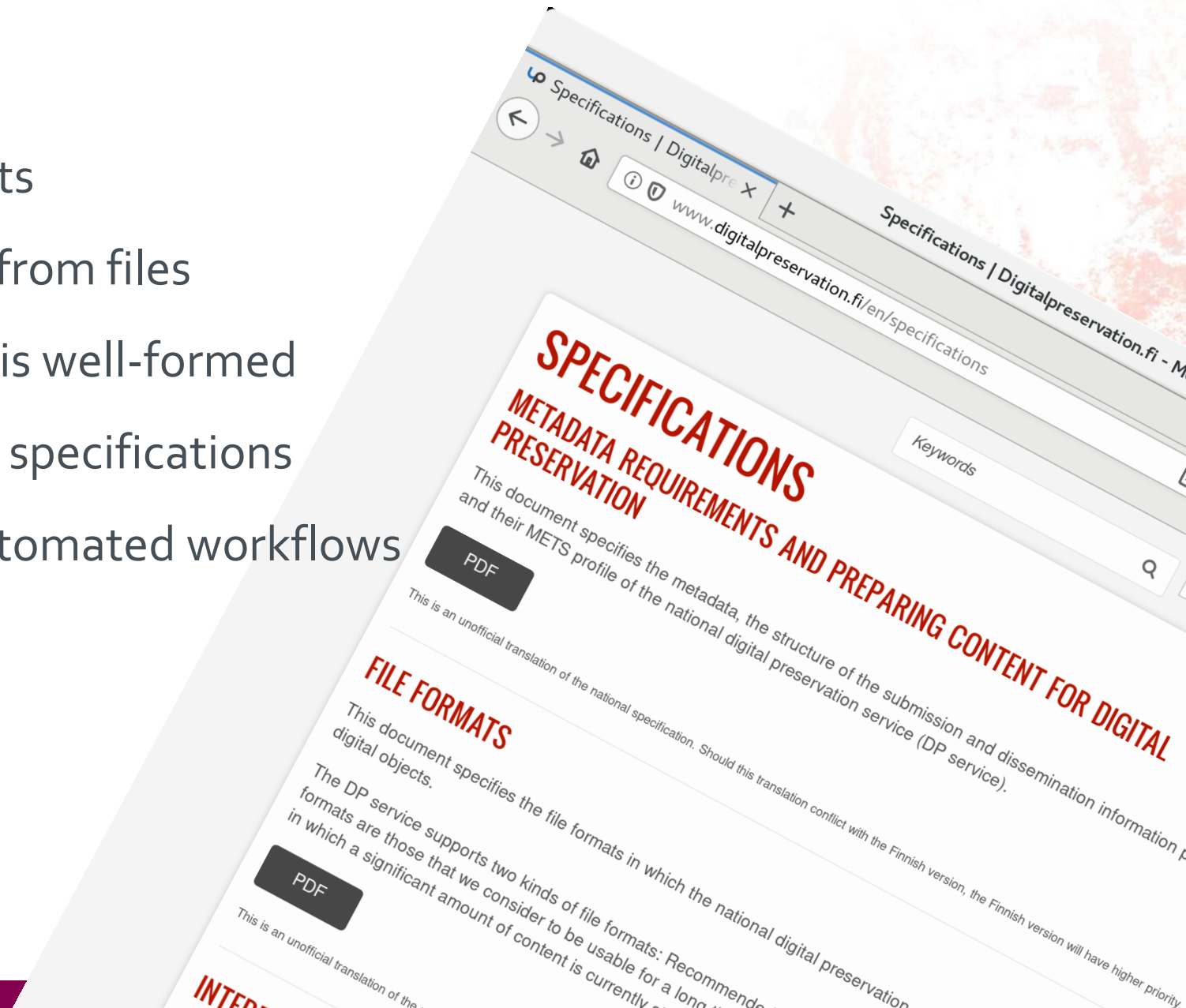
What We Needed

- A tool that can identify file formats
- A tool that can extract metadata from files
- A tool that can determine if a file is well-formed
- A tool whose output matches our specifications



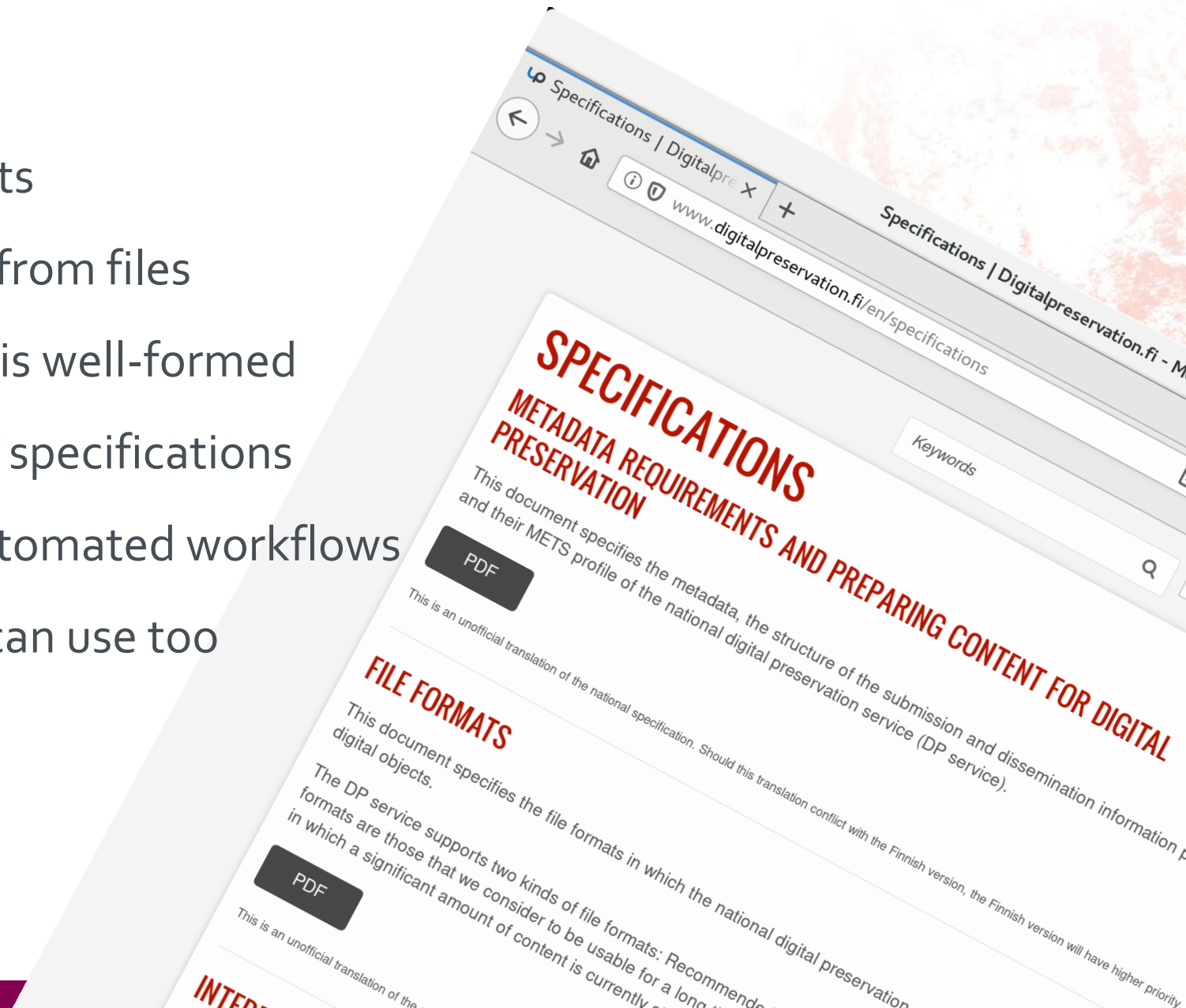
What We Needed

- A tool that can identify file formats
- A tool that can extract metadata from files
- A tool that can determine if a file is well-formed
- A tool whose output matches our specifications
- A tool that can be used in fully automated workflows



What We Needed

- A tool that can identify file formats
- A tool that can extract metadata from files
- A tool that can determine if a file is well-formed
- A tool whose output matches our specifications
- A tool that can be used in fully automated workflows
- A tool our partner organizations can use too



What We Created

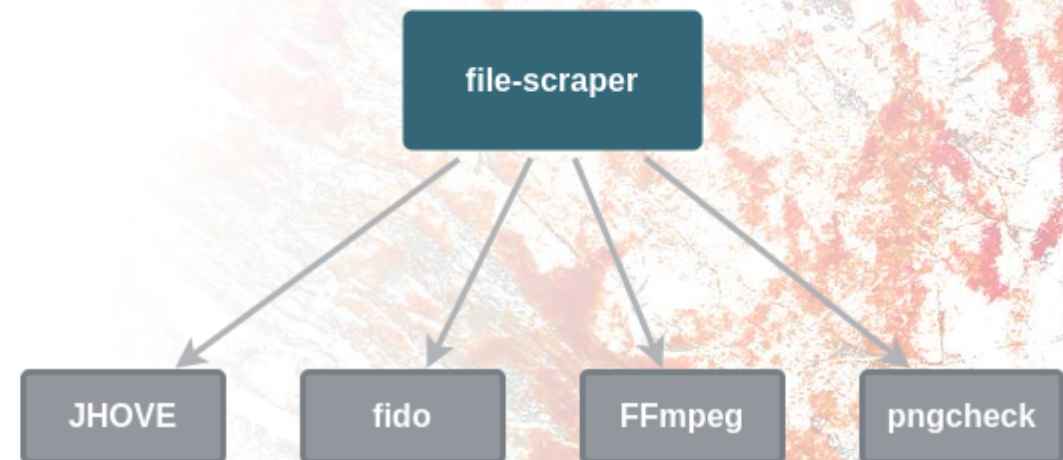
- Determine file format
- Extract metadata
- Check well-formedness
- Follow the spec
- Easy to use for machines
- Easy to use for people

What We Created

- Determine file format
- Extract metadata
- Check well-formedness
- Follow the spec
- Easy to use for machines
- Easy to use for people

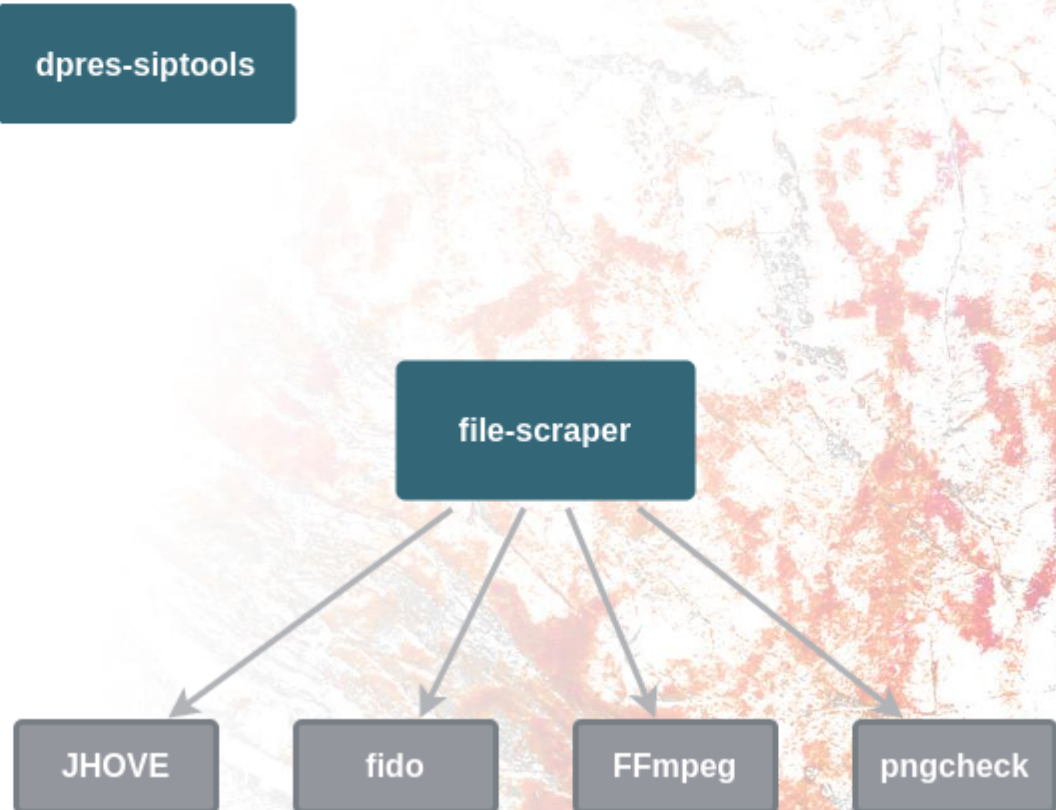
What We Created

- Determine file format
- Extract metadata
- Check well-formedness
- Follow the spec
- Easy to use for machines
- Easy to use for people



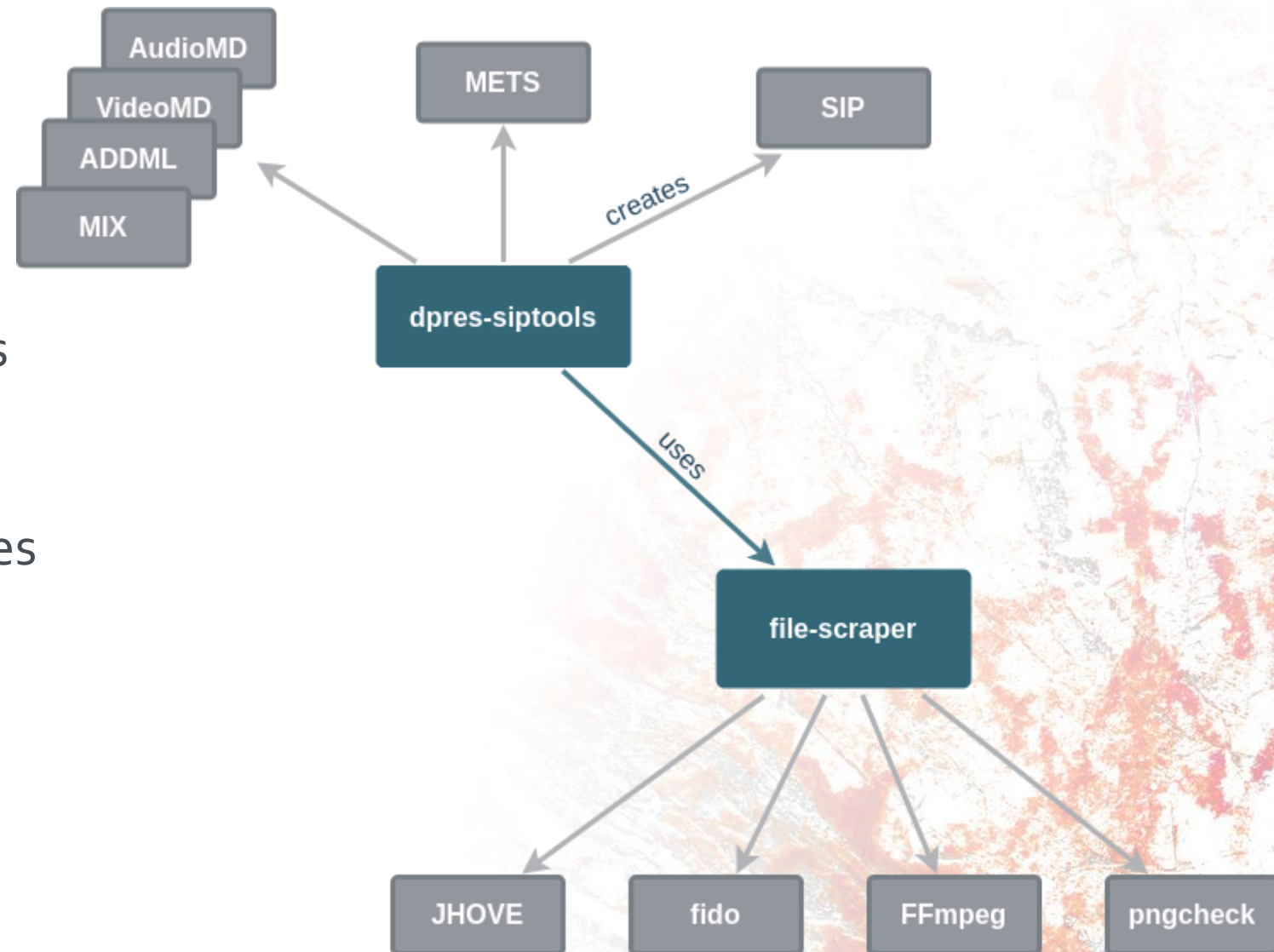
What We Created

- Determine file format
- Extract metadata
- Check well-formedness
- Follow the spec
- Easy to use for machines
- Easy to use for people



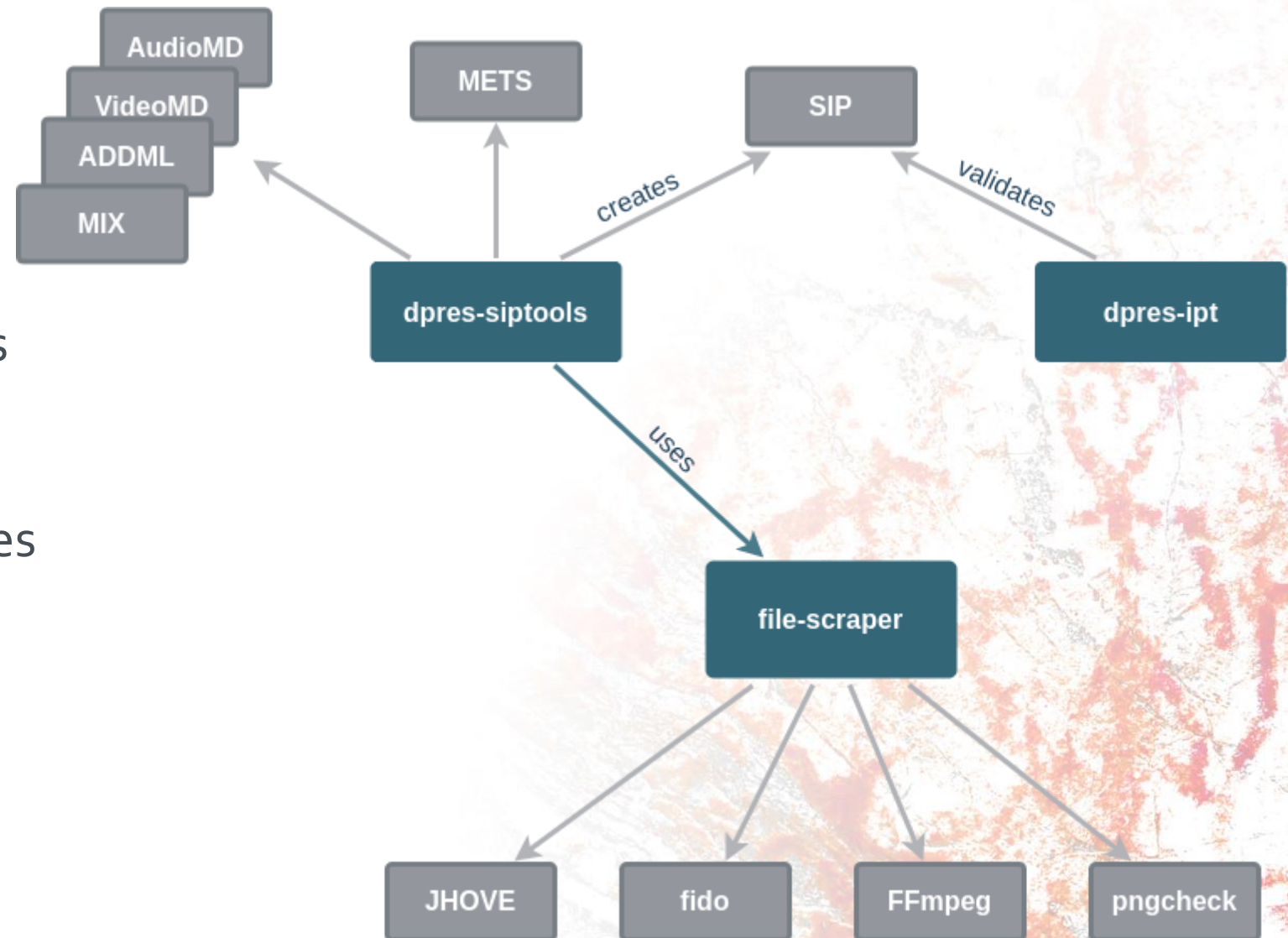
What We Created

- Determine file format
- Extract metadata
- Check well-formedness
- Follow the spec
- Easy to use for machines
- Easy to use for people



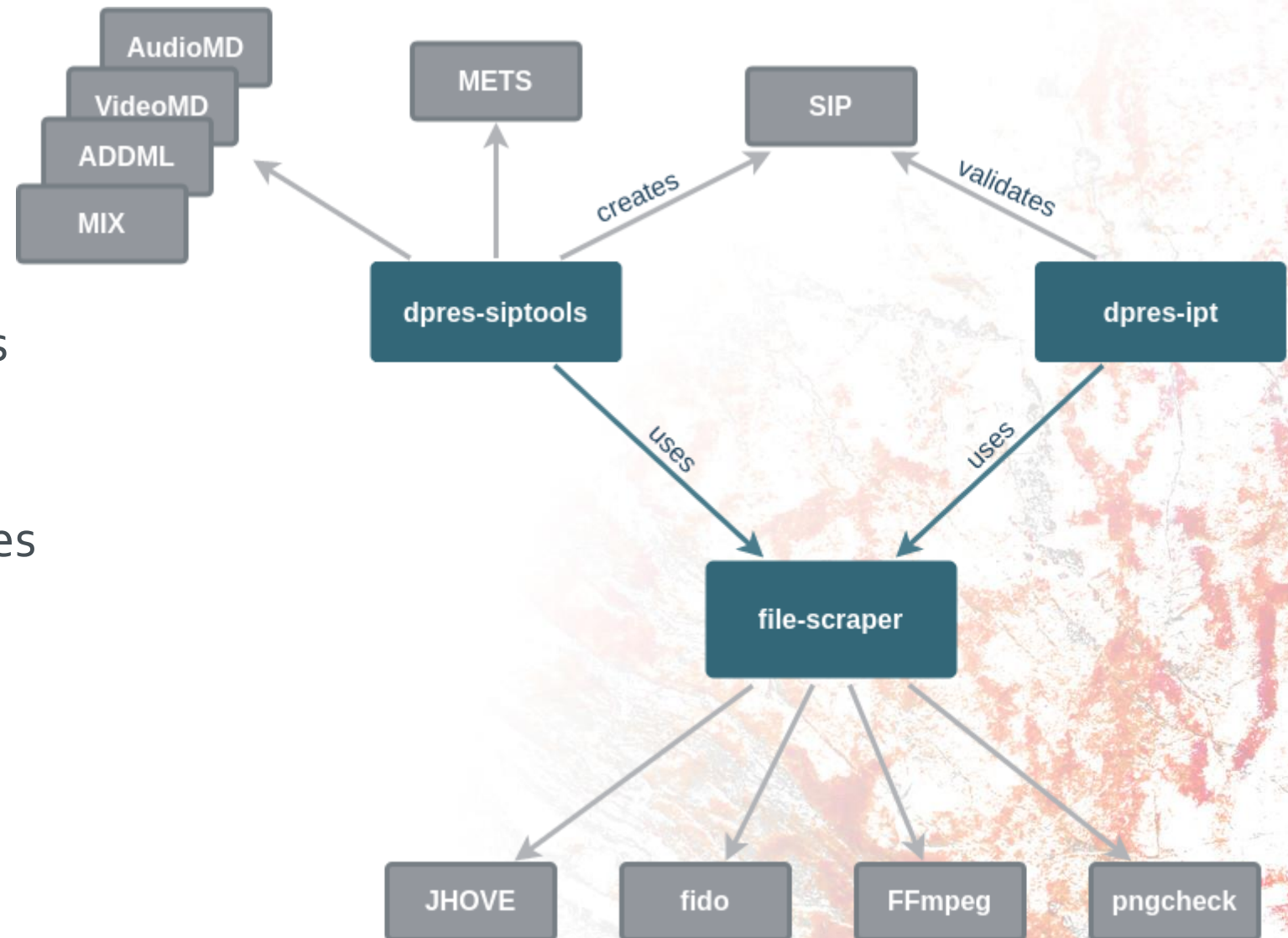
What We Created

- Determine file format
- Extract metadata
- Check well-formedness
- Follow the spec
- Easy to use for machines
- Easy to use for people



What We Created

- Determine file format
- Extract metadata
- Check well-formedness
- Follow the spec
- Easy to use for machines
- Easy to use for people



Is It Any Good?



Is It Any Good?

+ Works with a wide range of file formats



Is It Any Good?

- + Works with a wide range of file formats
- Only supports the file formats from our specs

Is It Any Good?

- + Works with a wide range of file formats
- + Extensible: adding support for new formats can be very easy
- Only supports the file formats from our specs

Is It Any Good?

- + Works with a wide range of file formats
 - + Extensible: adding support for new formats can be very easy
 - + It's FOSS!
-
- Only supports the file formats from our specs

Is It Any Good?

- + Works with a wide range of file formats
- + Extensible: adding support for new formats can be very easy
- + It's FOSS!
- + Easy to use by hand
- + Easy to use in other Python projects
- Only supports the file formats from our specs

Is It Any Good?

- + Works with a wide range of file formats
- + Extensible: adding support for new formats can be very easy
- + It's FOSS!
- + Easy to use by hand
- + Easy to use in other Python projects
- Only supports the file formats from our specs
- Full installation is somewhat heavy

Is It Any Good?

- + Works with a wide range of file formats
- + Extensible: adding support for new formats can be very easy
- + It's FOSS!
- + Easy to use by hand
- + Easy to use in other Python projects
- Only supports the file formats from our specs
- Full installation is somewhat heavy
- Sometimes you just need semantic information

Hot tip: OPF Webinar about these tools on September 10th, details available later!



Anni Järvenpää

Software Developer
Finnish National Digital Preservation Service
CSC – IT Center for Science Ltd

Contact the team at pas-support@csc.fi
or me at anni.jarvenpaa@csc.fi



digitalpreservation.fi



twitter.com/dpres_fi



github.com/Digital-Preservation-Finland