

Data services

Christine Staiger, SURFsara

Agenda



1pm - 2.10pm

- Introduction (Christine)
- GUI to archive, Cyberduck and FileZilla (Narges)
- **Coffee break**

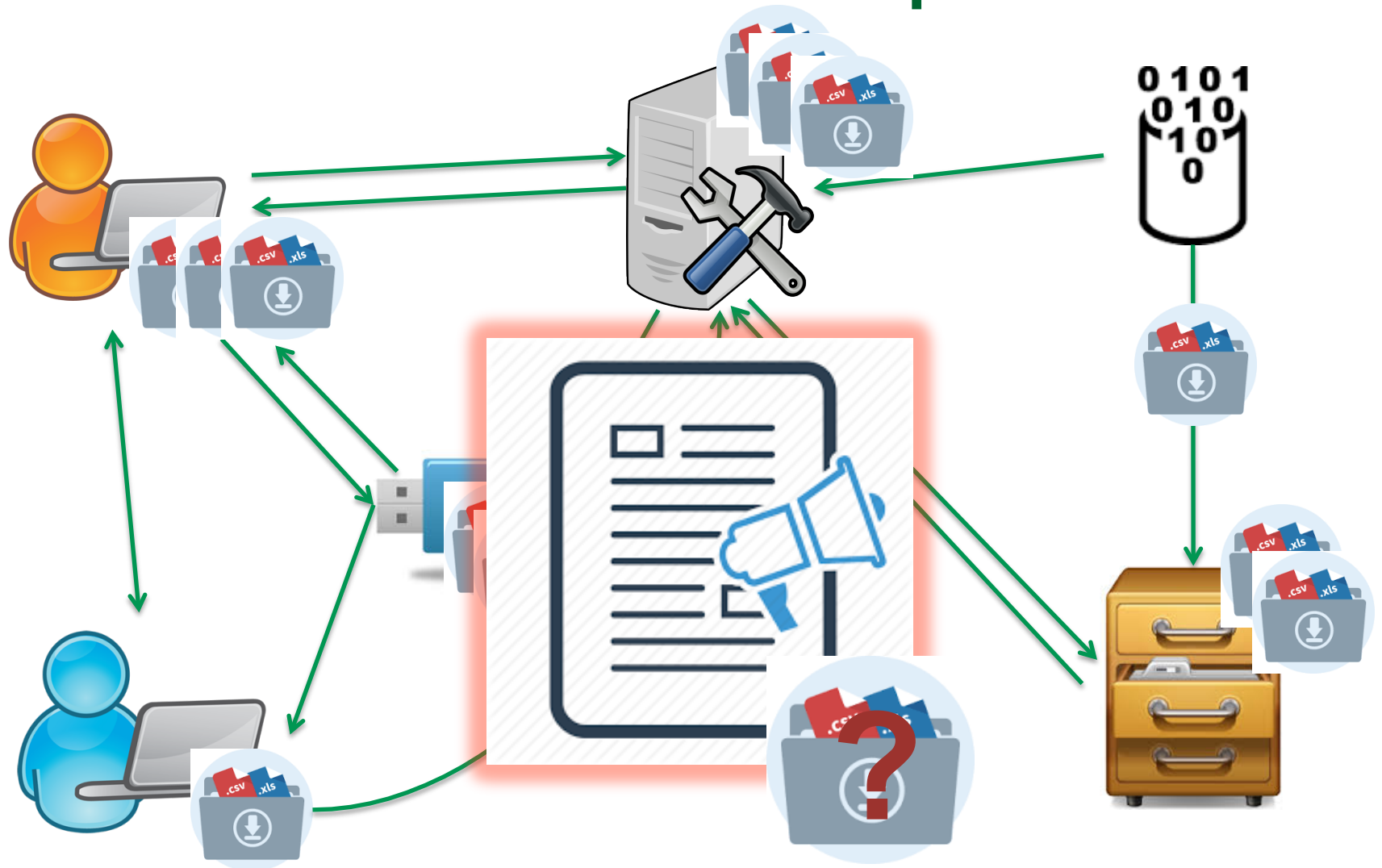
2.10pm - 4.10pm

- scp + rsync (Jeroen)
- Tarring data with tar and dmftar (Hans)
- **Coffee break**

4.10pm - 5pm

- Wrap-up (Christine)
- Final Challenge:
Putting it all together in a python workflow

Data – where is the problem?



Data services in the Netherlands – there are a lot of solutions



4TU.Centre for Research Data



Handle.Net®

Data ingest service



Data services at SURFsara

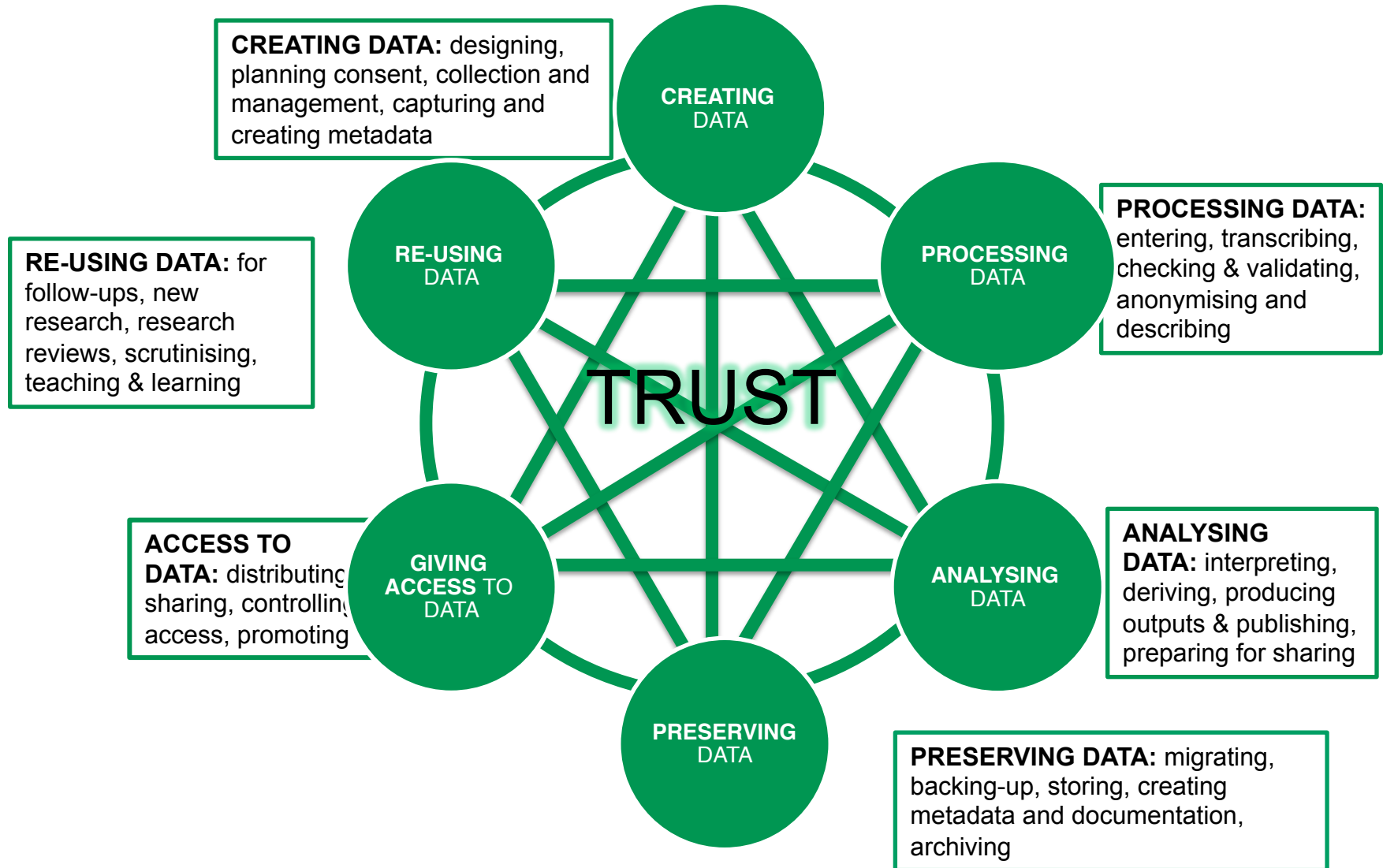


Data ingest service

Handle.Net[®]



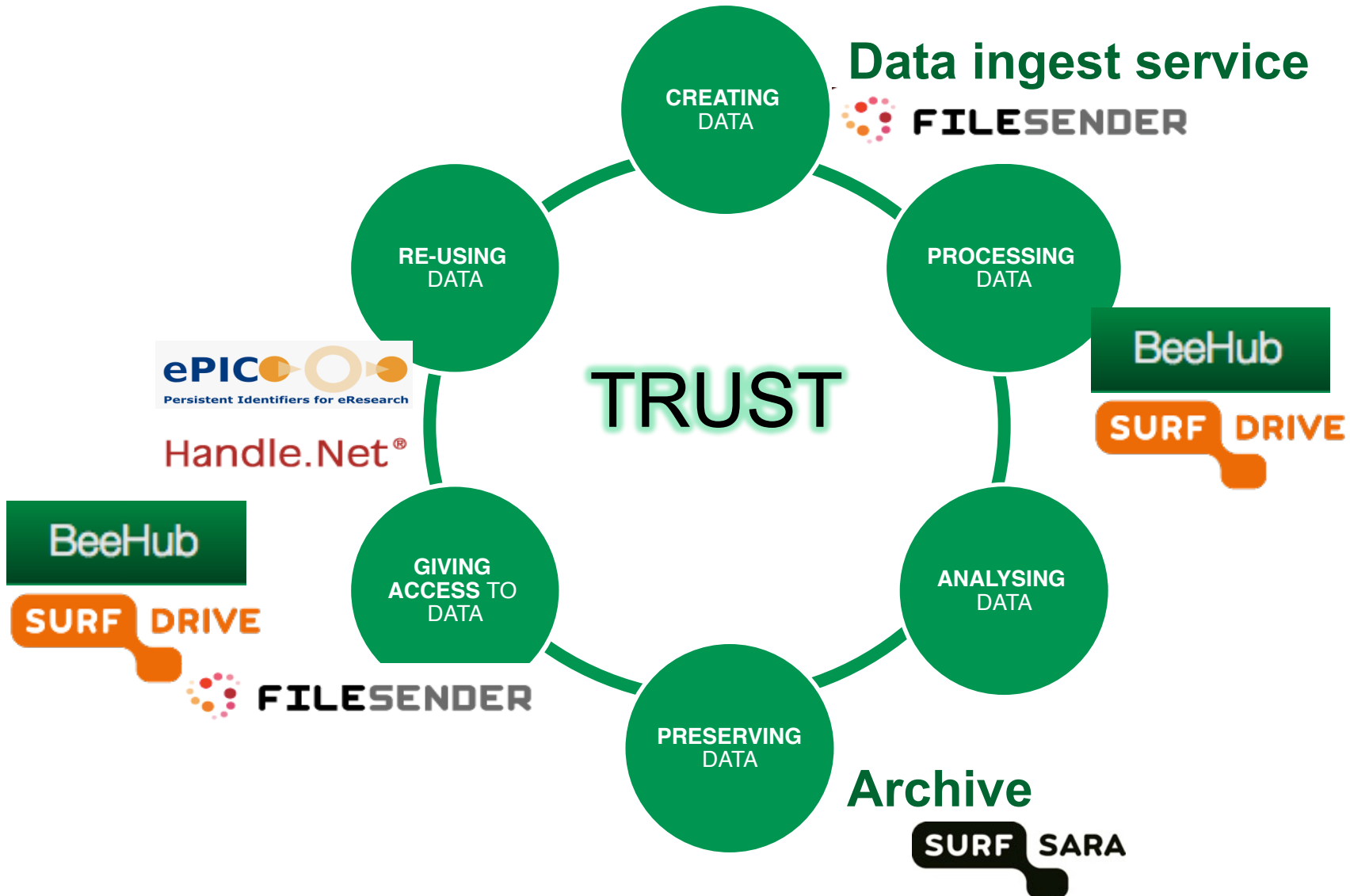
The data Life cycle



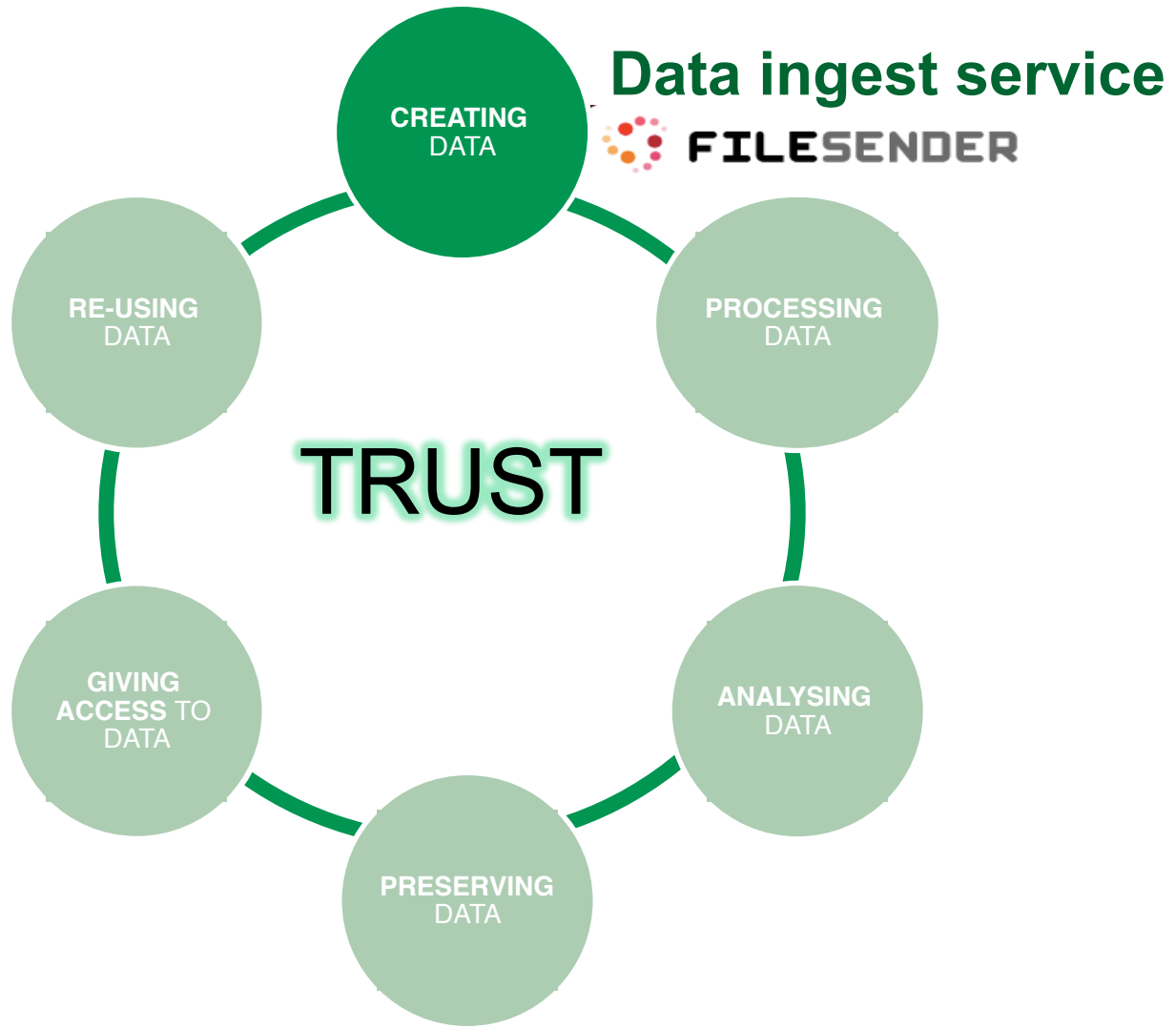
The researchers' needs

- **Store** data during research
- **Share** data during and after research
- **Archive** data
- **Synchronise** data across different locations
- **Link** publication to processed and raw data
- **Publish** data
- **Find** data and **make data findable** by others
- Data **transfers**
- Data **provenance**: what happened with the data
- ...

The data Life cycle

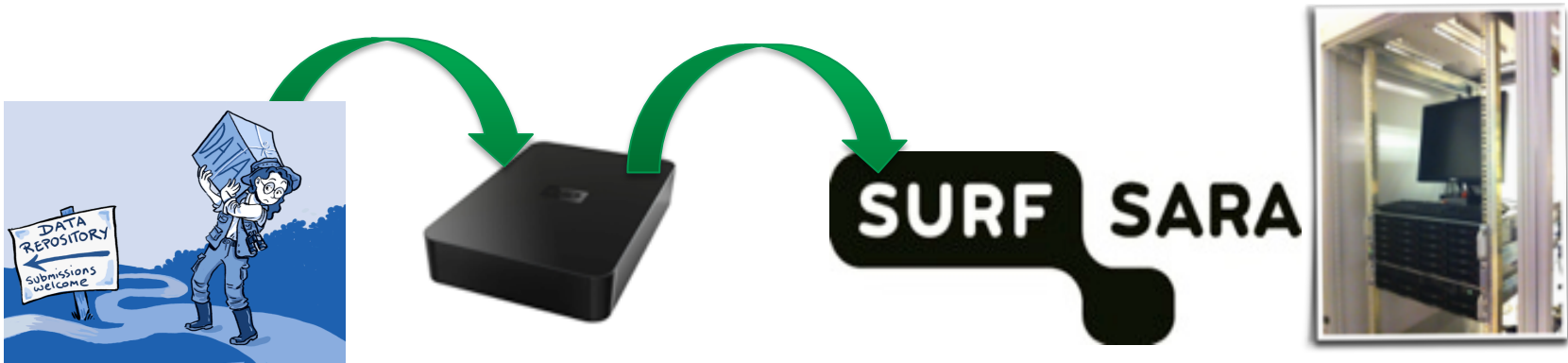


The data Life cycle



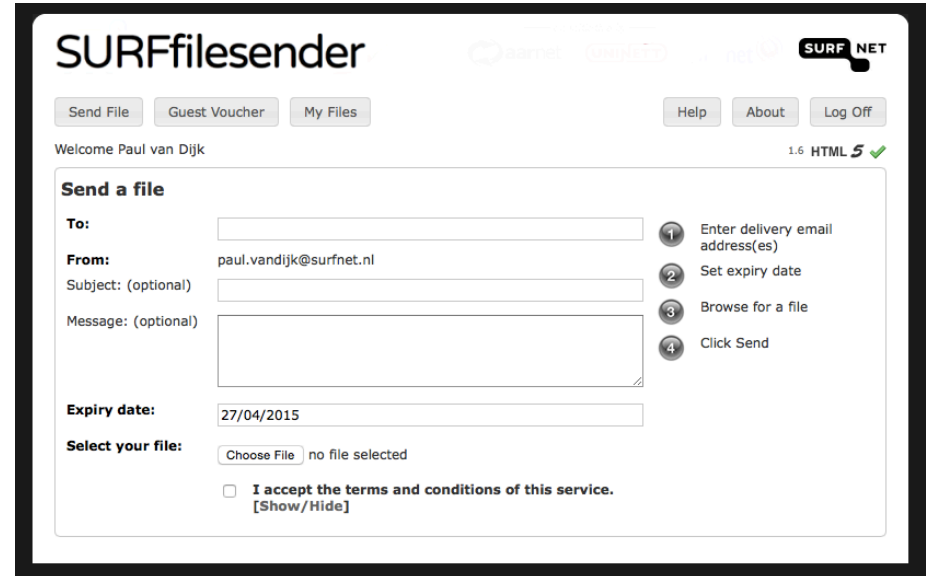
Data ingest service

- Data often resides on external storage media, USB sticks, external hard drives
- Slow or no internet connection
- Easy way to upload large data from disk to SURFsara facilities
- Upload data from 45 disks in parallel



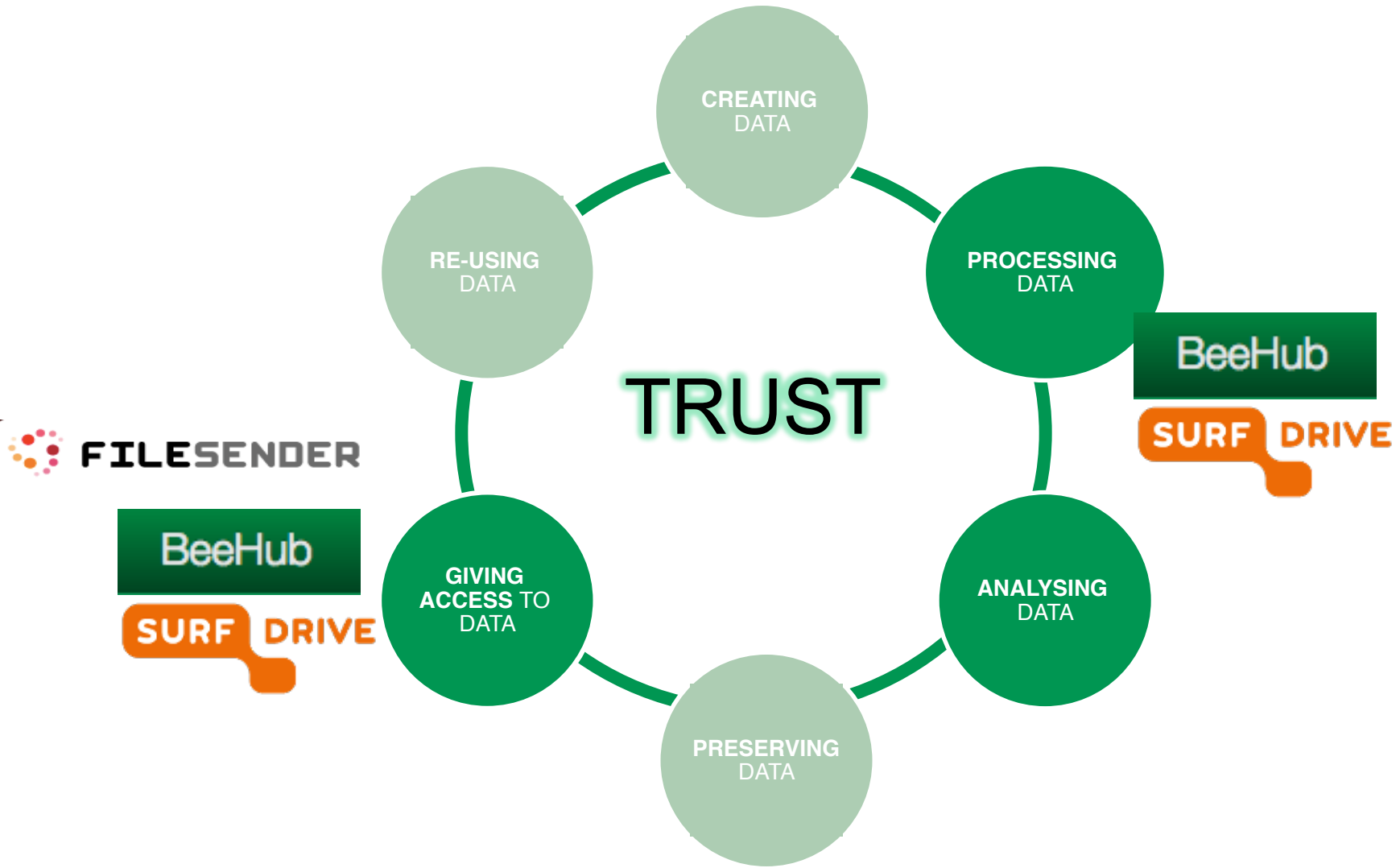
FILESENDER

- Trusted community service
- Transferring BIG files from person to person
- File Transport service not File Storage (!)
- Simple interface
- Up to 200GB per file
- Option to apply end-to-end encryption (250MB browser limit)
- Vouchers for guest usage



The screenshot displays the SURFfilesender web interface. At the top, the title 'SURFfilesender' is visible alongside navigation links for 'Send File', 'Guest Voucher', and 'My Files'. A welcome message 'Welcome Paul van Dijk' is shown, along with a '1.6 HTML5' status indicator. The main section is titled 'Send a file' and contains a form with the following fields: 'To:' (empty), 'From:' (paul.vandijk@surfnet.nl), 'Subject: (optional)' (empty), and 'Message: (optional)' (empty text area). To the right of the form, a numbered list of steps is provided: 1. Enter delivery email address(es), 2. Set expiry date, 3. Browse for a file, and 4. Click Send. Below the message field, the 'Expiry date:' is set to '27/04/2015'. The 'Select your file:' section shows a 'Choose File' button and the text 'no file selected'. At the bottom, there is a checkbox for 'I accept the terms and conditions of this service.' with a '[Show/Hide]' link.

The data Life cycle



SURFdrive

- Trusted community cloud for personal storage
- Sharing smaller data files
- Collaboration between SURFsara, SURFnet and Dutch universities
- Specifications and service determined by end-users (universities)
- 100 GB storage capacity per user
- Based on ownCloud, synchronises with local storage













Beehub

BeeHub Files (new client) Files (old client) Groups Docs 1 notifications ▾ Machiel Jansen ▾ SURF SARA

BeeHub root » home / mgjansen

Contents ACL

Up Home Upload New Copy Move Delete

<input type="checkbox"/>	Name	Size	Type	Modified	Owner
<input type="checkbox"/>	 ebooks-2011/			16-2-2012 10:42	Machiel Jansen
<input type="checkbox"/>	 .DS_Store	6 KB	application/x-empty; charset=binary	26-3-2013 15:07	Machiel Jansen
<input type="checkbox"/>	 ._DS_Store	4 KB	application/x-empty; charset=binary	4-3-2013 15:30	Machiel Jansen
<input type="checkbox"/>	 DA6ACEFF-D736-48B9-BE66-7AB94CEAFC0E.jpg	1.71 MB	image/jpeg	13-1-2012 11:42	Machiel Jansen
<input type="checkbox"/>	 ._ILR2.0.pptx	4 KB	application/x-empty; charset=binary	29-3-2013 11:25	Machiel Jansen
<input type="checkbox"/>	 rbn2.0-distributie2.zip	9.06 MB	application/zip	9-7-2013 11:12	Machiel Jansen
<input type="checkbox"/>	 Cornetto_2.0.zip	177.97 MB	application/zip	9-7-2013 11:13	Machiel Jansen
<input type="checkbox"/>	 beoordelingen.zip	3.98 MB	application/zip	10-10-2013 17:18	Machiel Jansen
<input type="checkbox"/>	 resourceplanning.zip	27.39 MB	application/zip	10-10-2013 17:25	Machiel Jansen

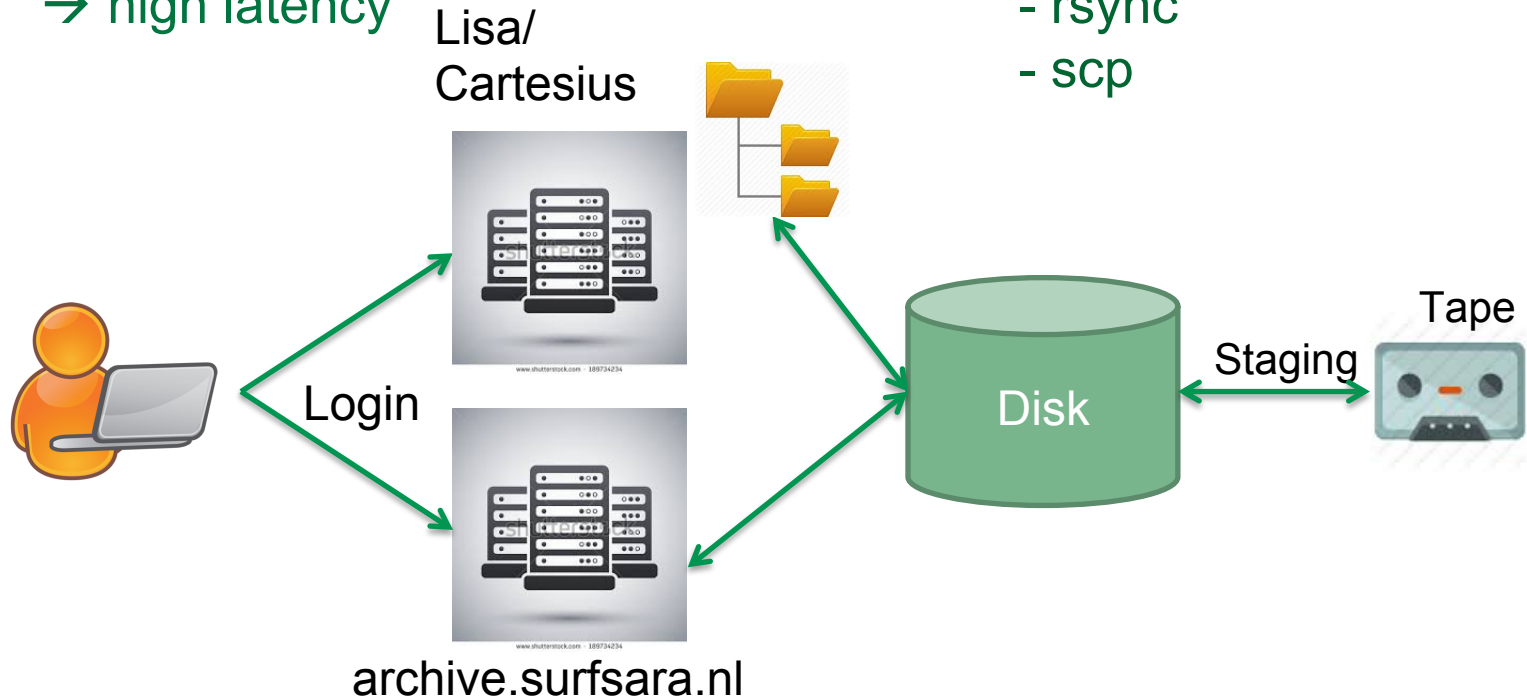
- Sharing large data
- Mountable via webdav (no synchronisation)
- Capacity: 80 TB
- Temporary storage
- 100GB/user
- You can apply for more via SURF e-infrastructure grant

The data Life cycle



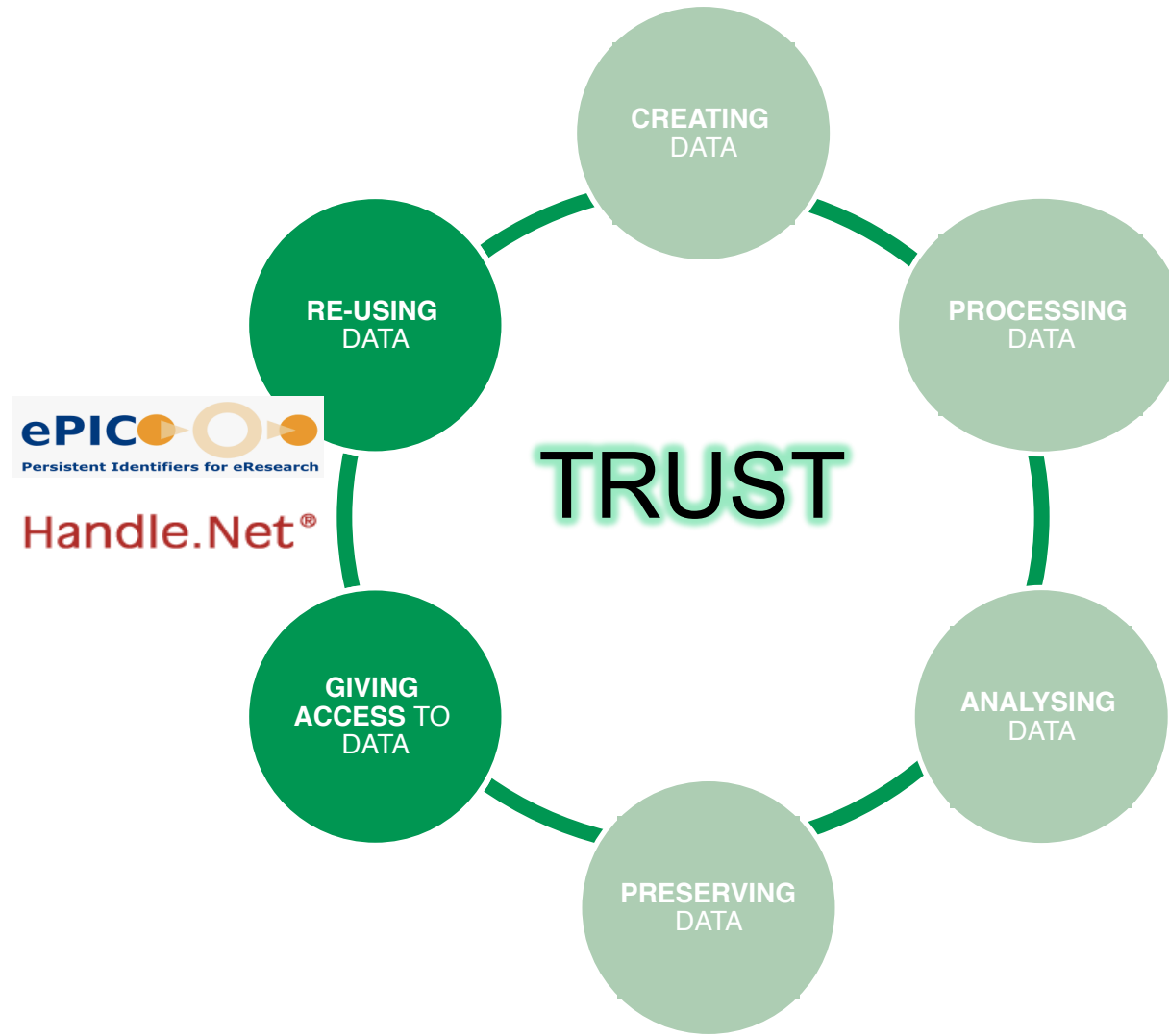
The Archive – Long-term storage

- Long-term storage of big data
- Storage medium: Tape
→ high latency
- Powerful transfer protocols:
 - gridFTP
 - rsync
 - scp



- Easy access from HPC services lisa and cartesius via NFS mounts → use archive as yet another directory

The data Life cycle

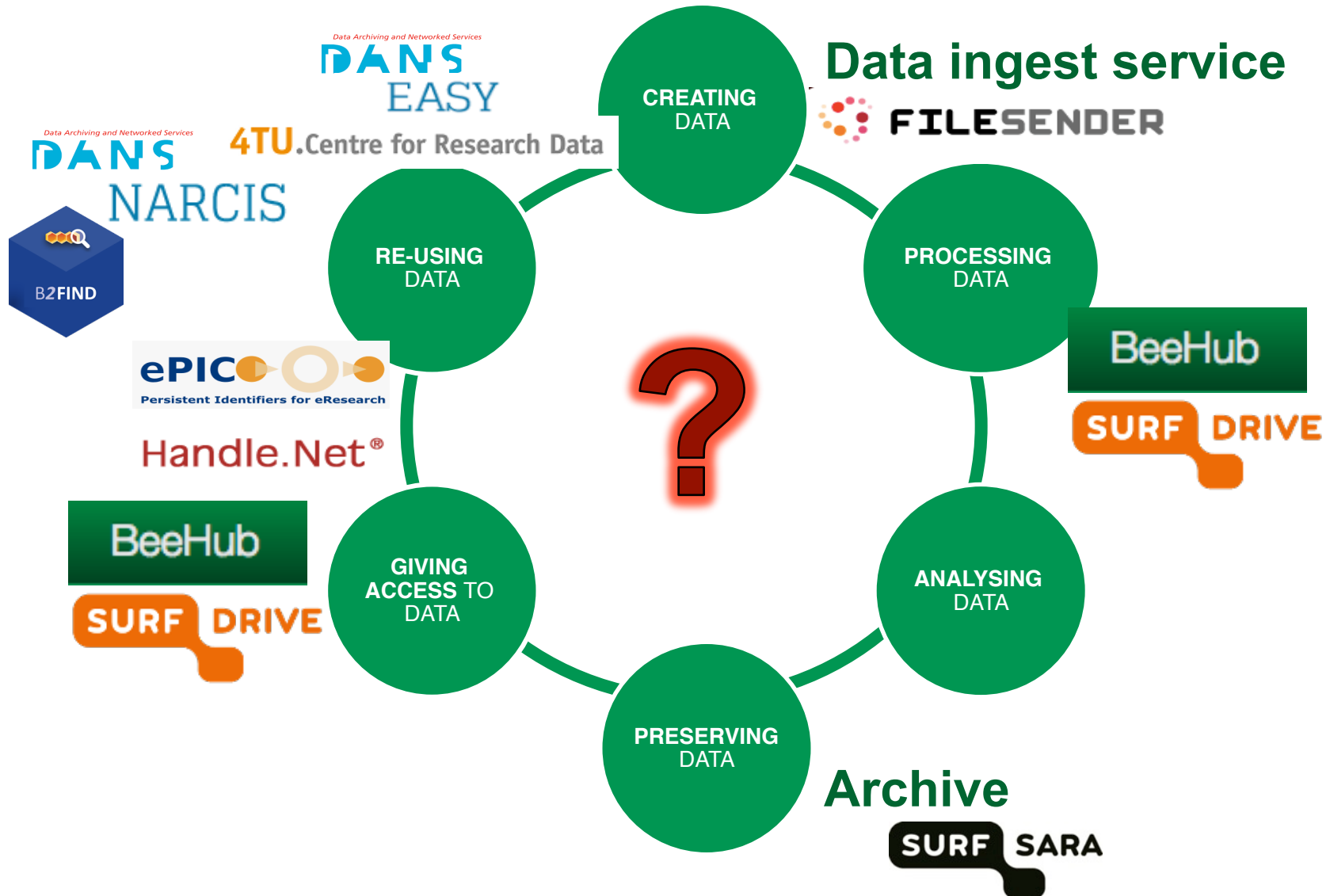


PIDs

- PIDs (Persistent Identifiers) are
 - Pointers to resources like files, folders, webpages, real world objects
 - Globally unique
 - Resolvable via http
- Example resolvers: <https://dx.doi.org/> and <http://hdl.handle.net/>



The data Life cycle



CONTACT and more information

Information on SURFsara services:

<https://userinfo.surfsara.nl/>

Helpdesk:

info@surfsara.nl - General information

helpdesk@surfsara.nl - Information and
guidance on

infrastructure