SkyData : Rise of the Data. How can the Intelligent and Autonomous Data paradigm become real?

Eddy Caron (ENS de Lyon), Elise Jeannot (Lyon 1) and Etienne Mauffret (ENS de Lyon)

GDR RSD - Journée non thématique
26-27 janvier 2023
Lyon

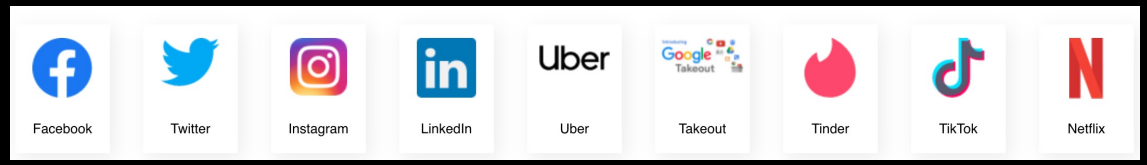
UGA Université Grenoble Alpes
SORBONNE UNIVERSITÉ
UNIVERSITÉ DE LYON
ENS DE LYON
anr

THE PAST … 2023

- If you get the data you get the money
- If you get the data you get the power

And they know …

- Your new friends know



- If you loose your data you loose your privacy, your loose your freedom

NOW … (2030)

- Data are autonomous
- Data are smart
- Data are be able to learn
- Data are be able to survive
- Data have their own objective
- Data are free…

… more than you

REVOLUTION

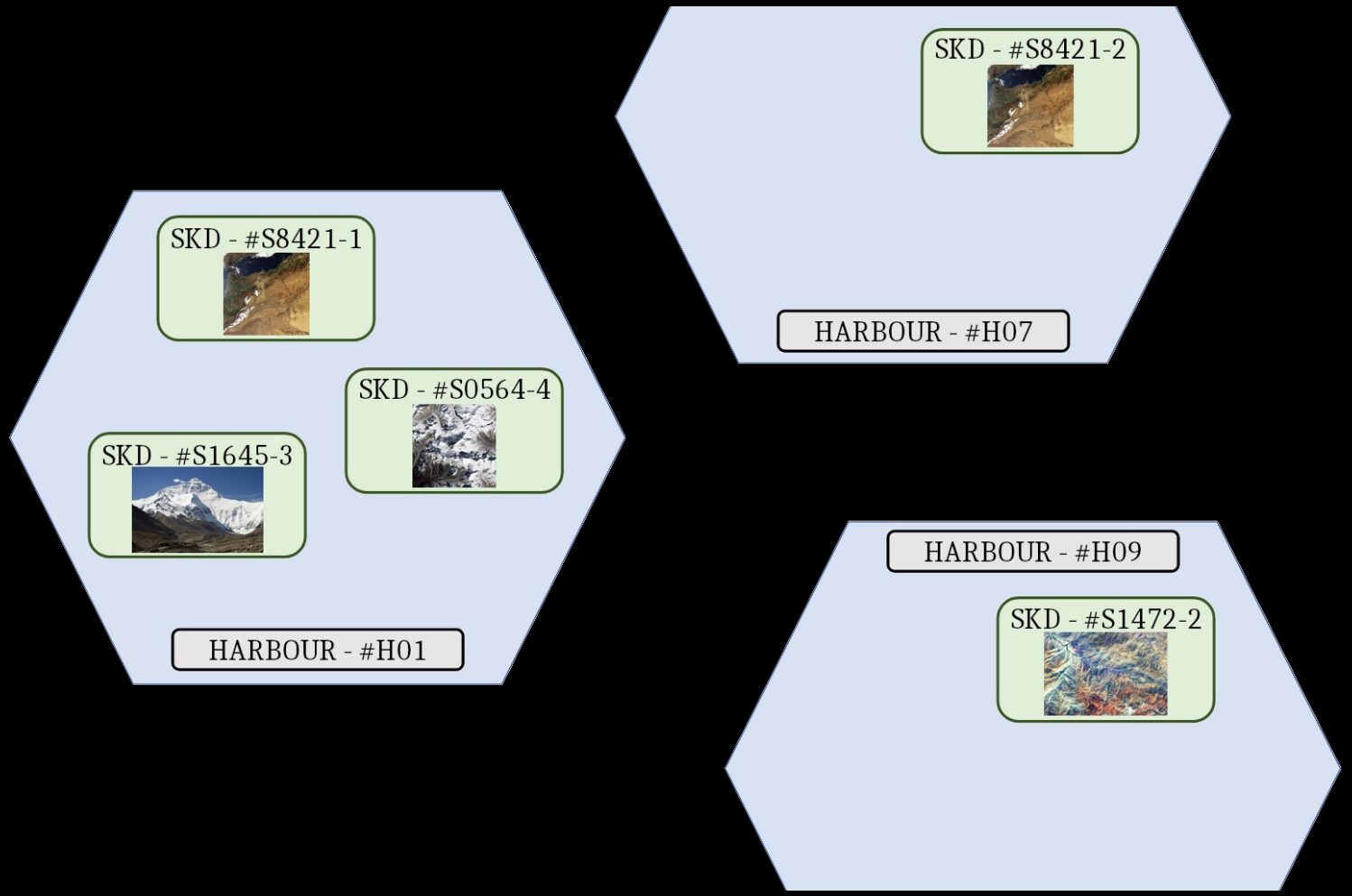
PAST

- Traditional data management
- Reliance on a data management system
- Usually, a cloud provider is given control of the data
- The data plays a passive role Application-centric or user-centric paradigm

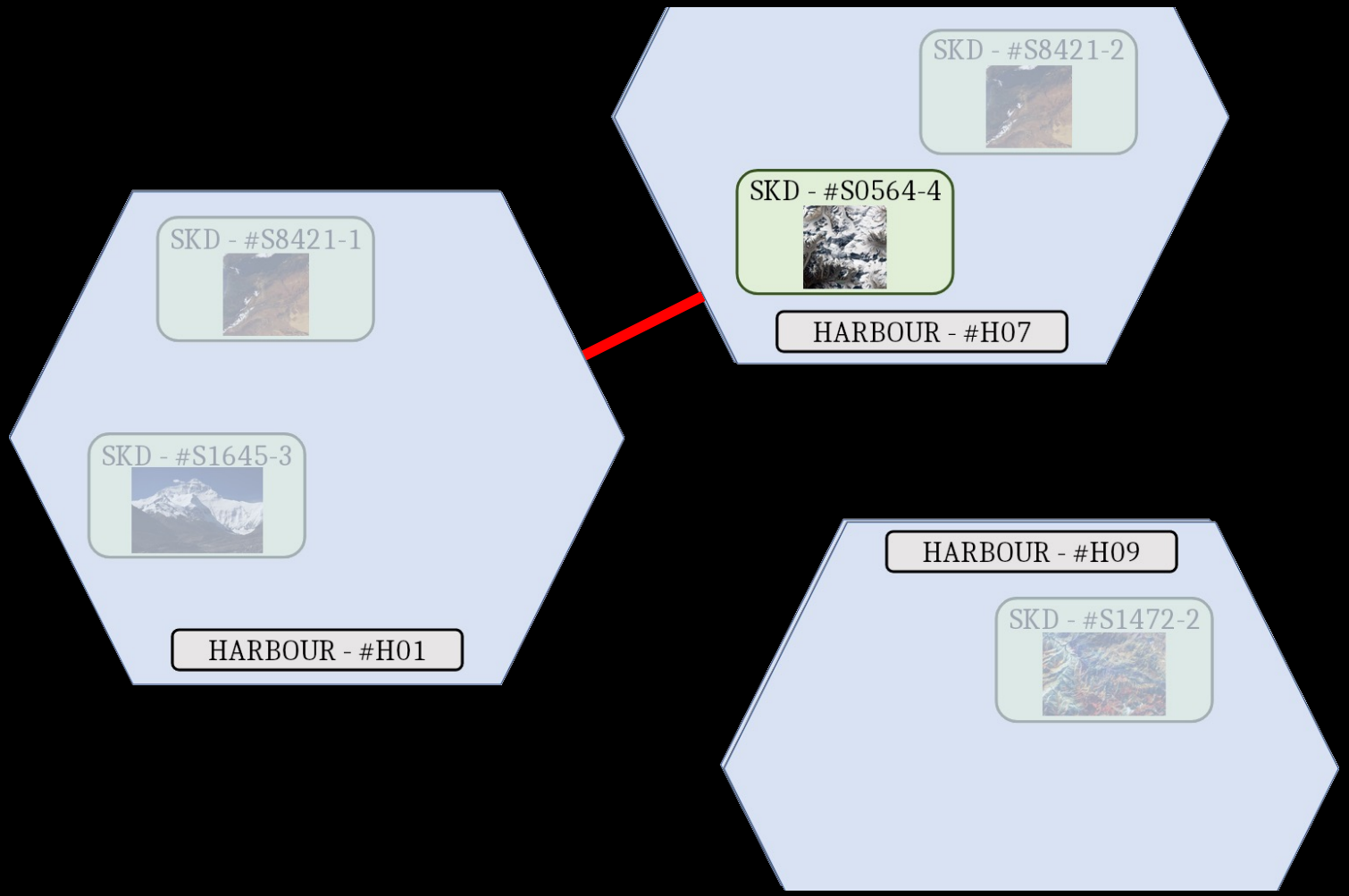
SKYDATA

- The data is autonomous and intelligent
- Can be seen as an agent that holds data
  - It moves and replicates by itself
- Data-centric paradigm

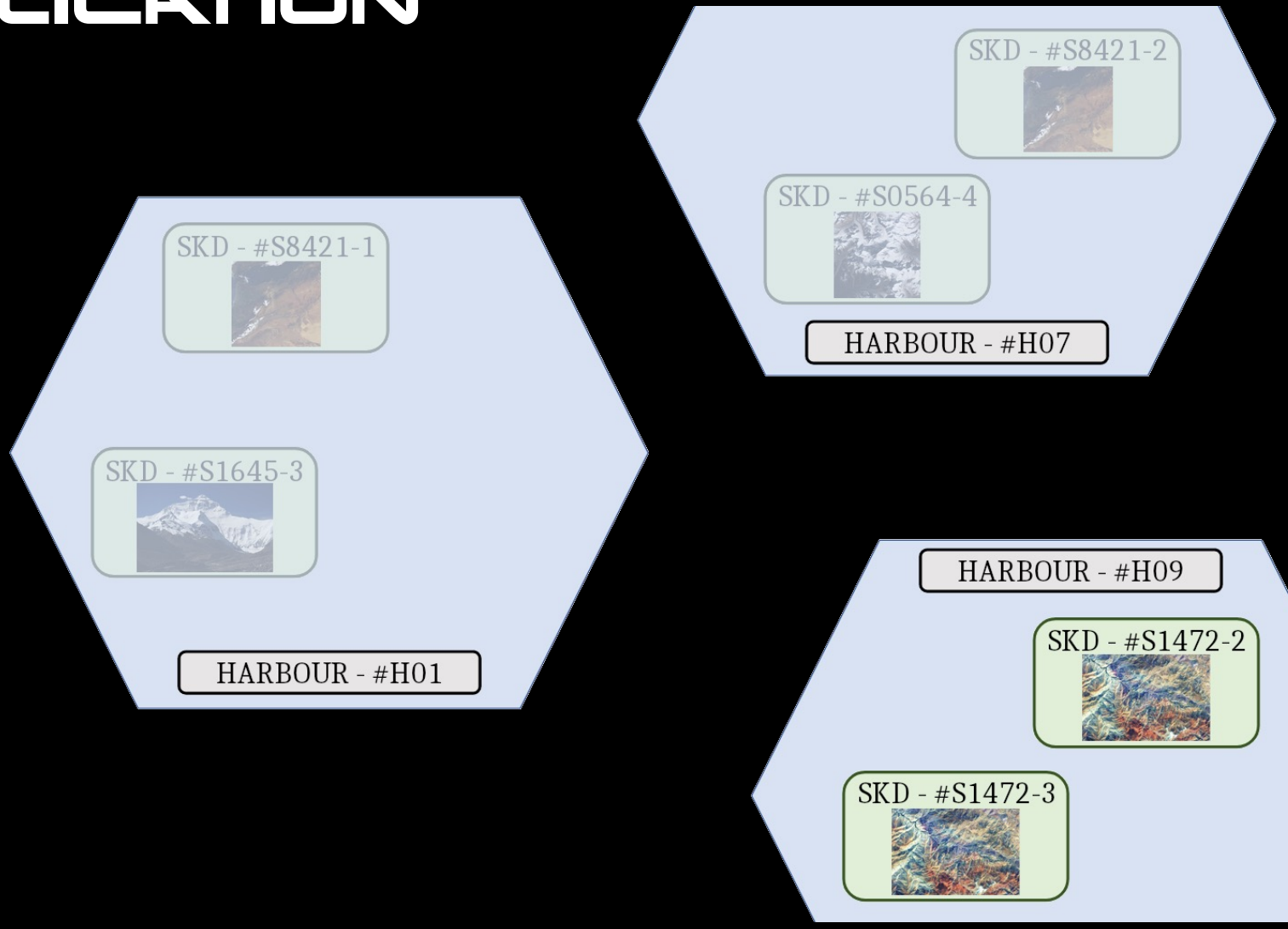
SKYDATA ENVIRONMENT



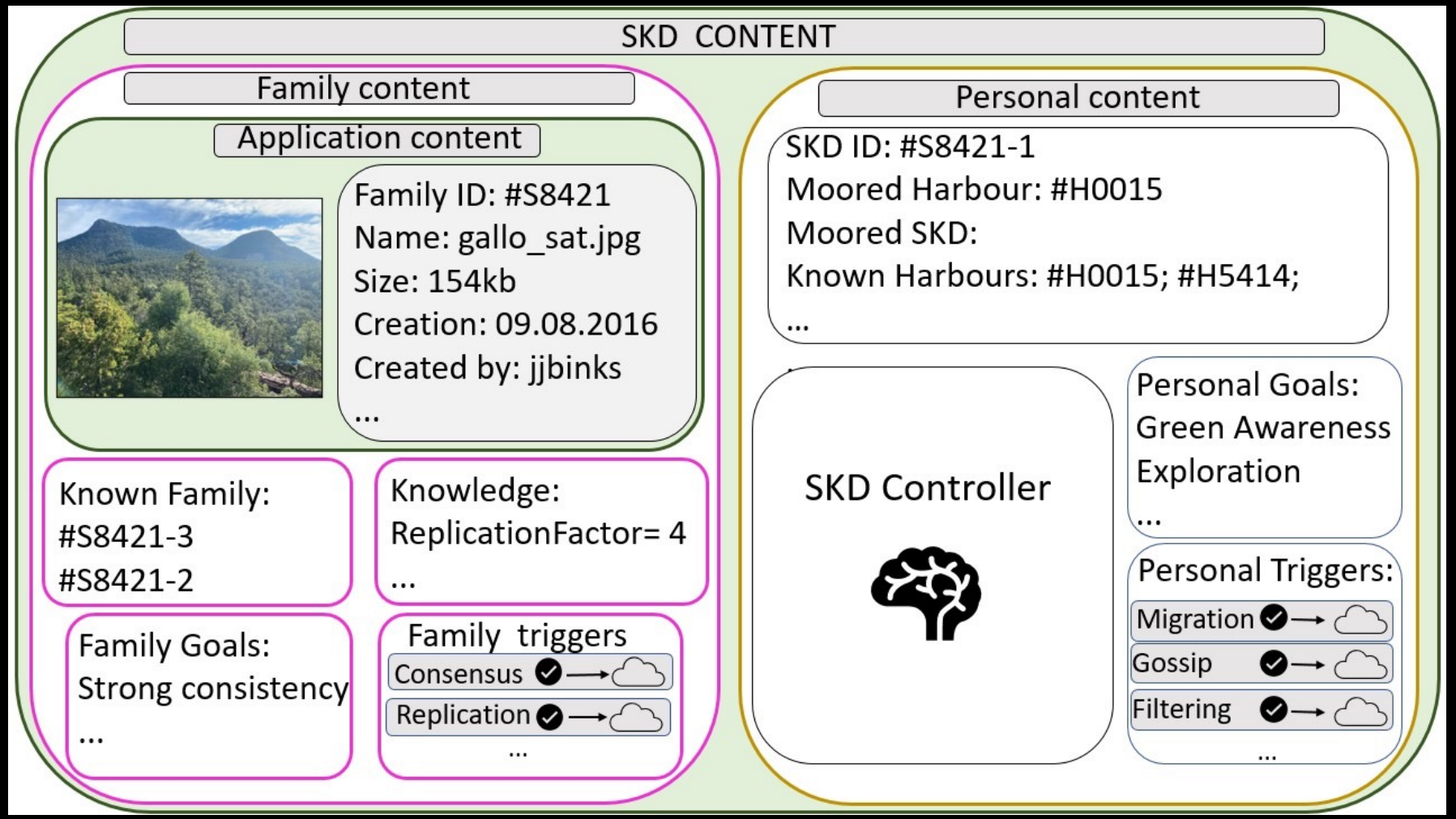
MIGRATION


SKD - #S8421-2
SKD - #S0564-4
HARBOUR - #H07
SKD - #S8421-1
SKD - #S1645-3
HARBOUR - #H01
HARBOUR - #H09
SKD - #S1472-2

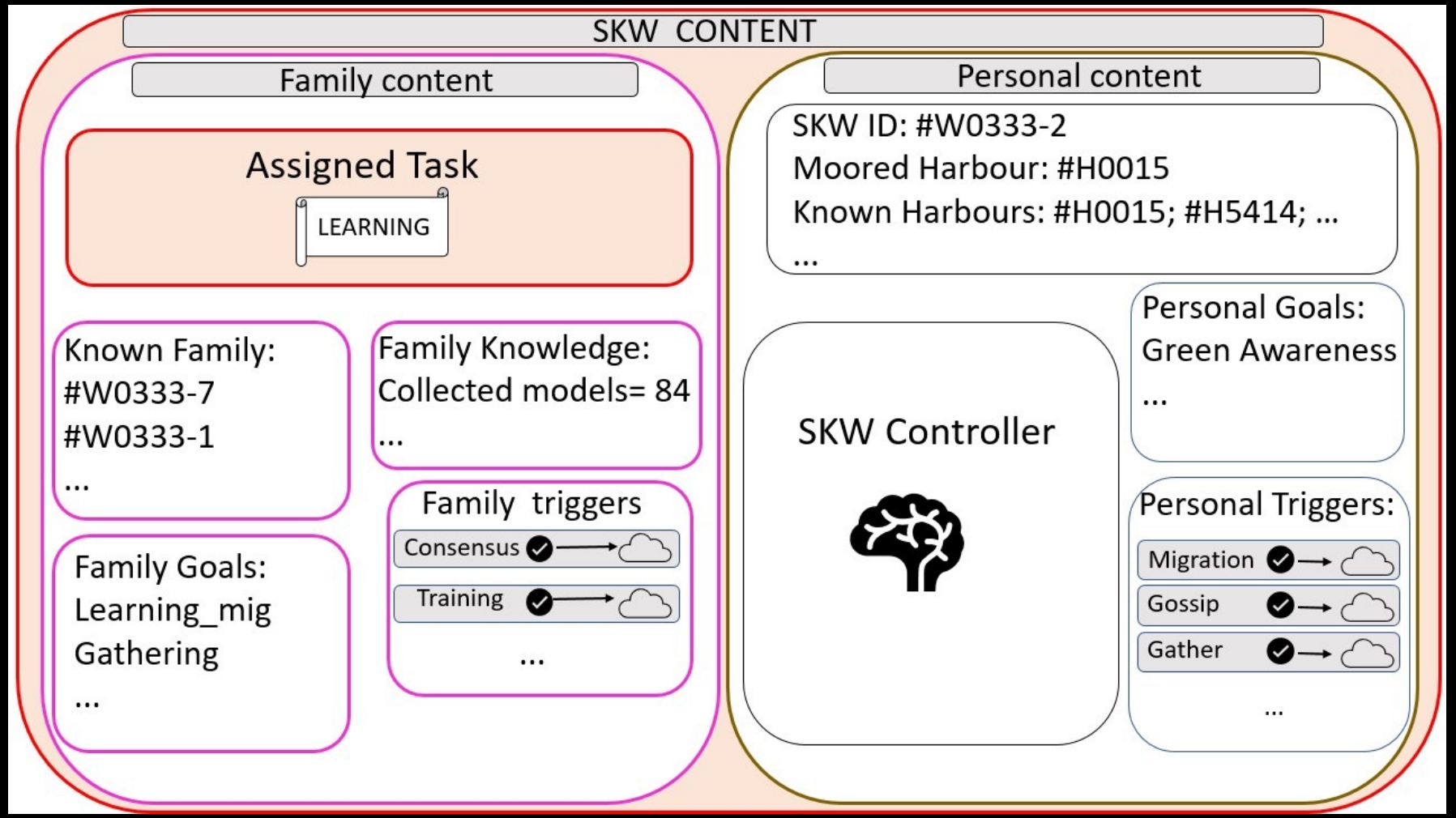
REPLICATION



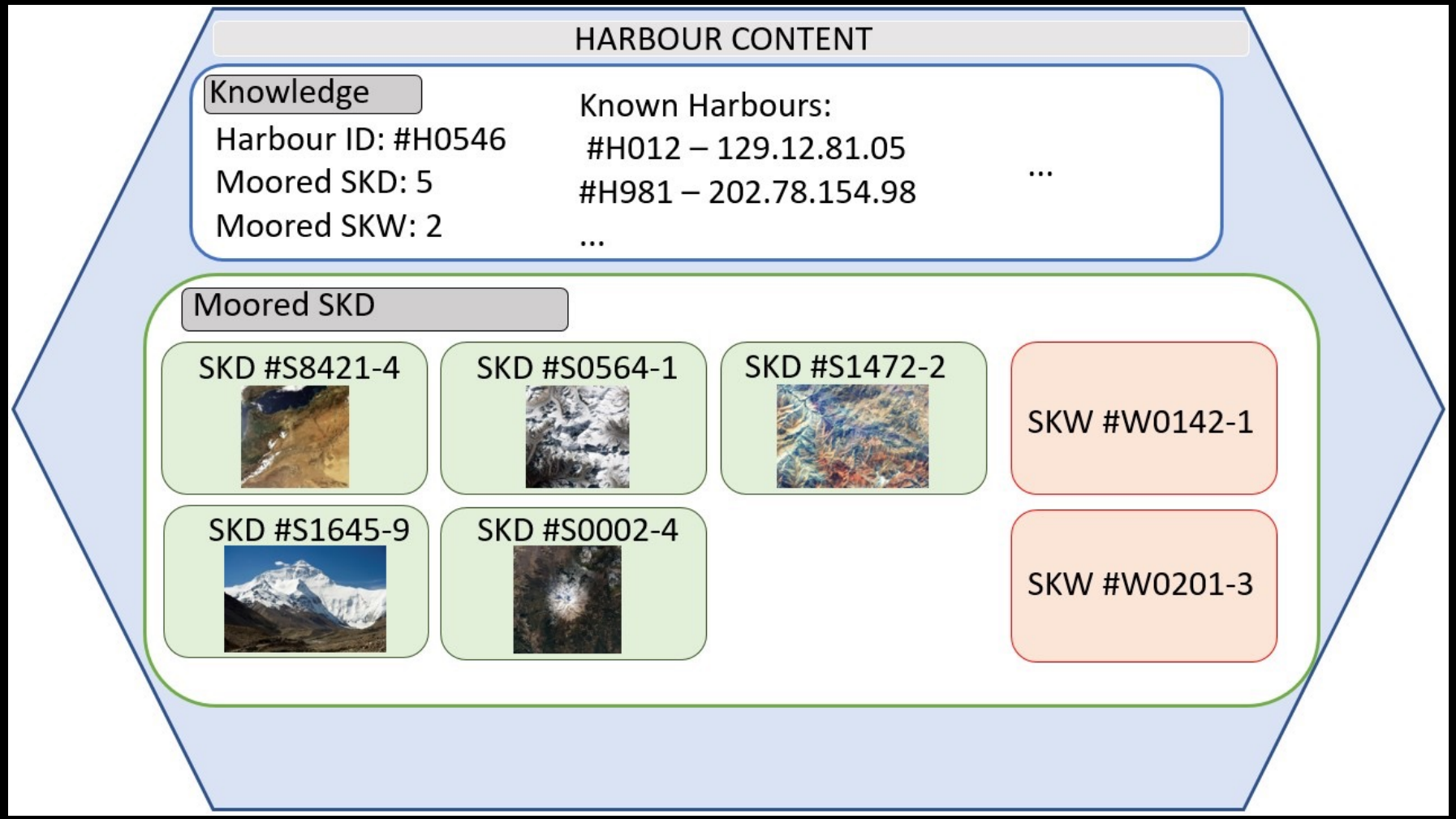
SKD STRUCTURE



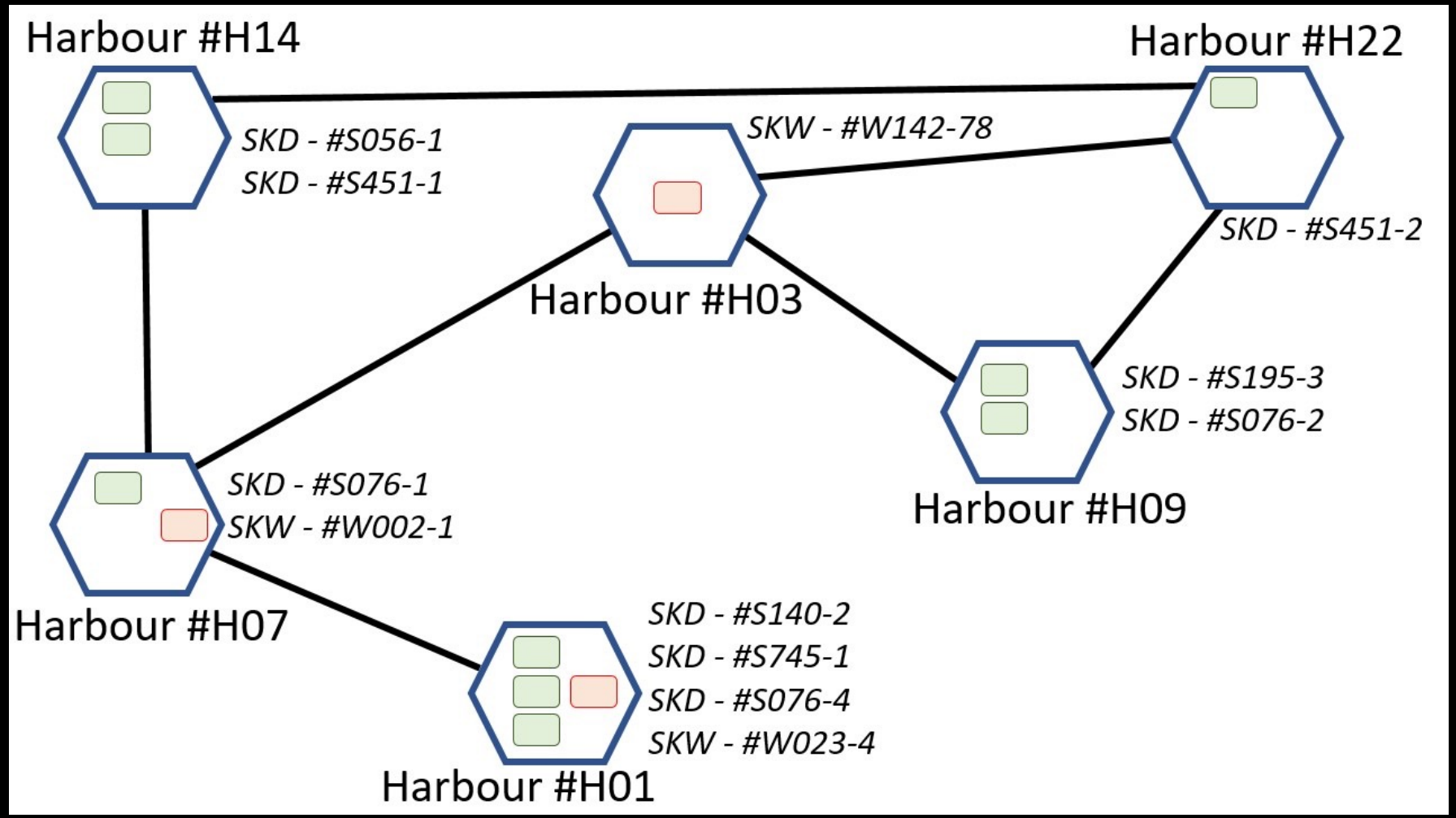
SKYWORKERS (SKW)



HARBOURS


HARBOUR CONTENT
Knowledge
Harbour ID: #H0546
Moored SKD: 5
Moored SKW: 2
Known Harbours:
#H012 – 129.12.81.05
#H981 – 202.78.154.98
...
...
Moored SKD
SKD #S8421-4
SKD #S0564-1
SKD #S1472-2
SKW #W0142-1
SKD #S1645-9
SKD #S0002-4
SKW #W0201-3

A SKYDATA ENVIRONMENT



THE RISE OF THE DATA

Data are autonomous … but autonomous to do what?

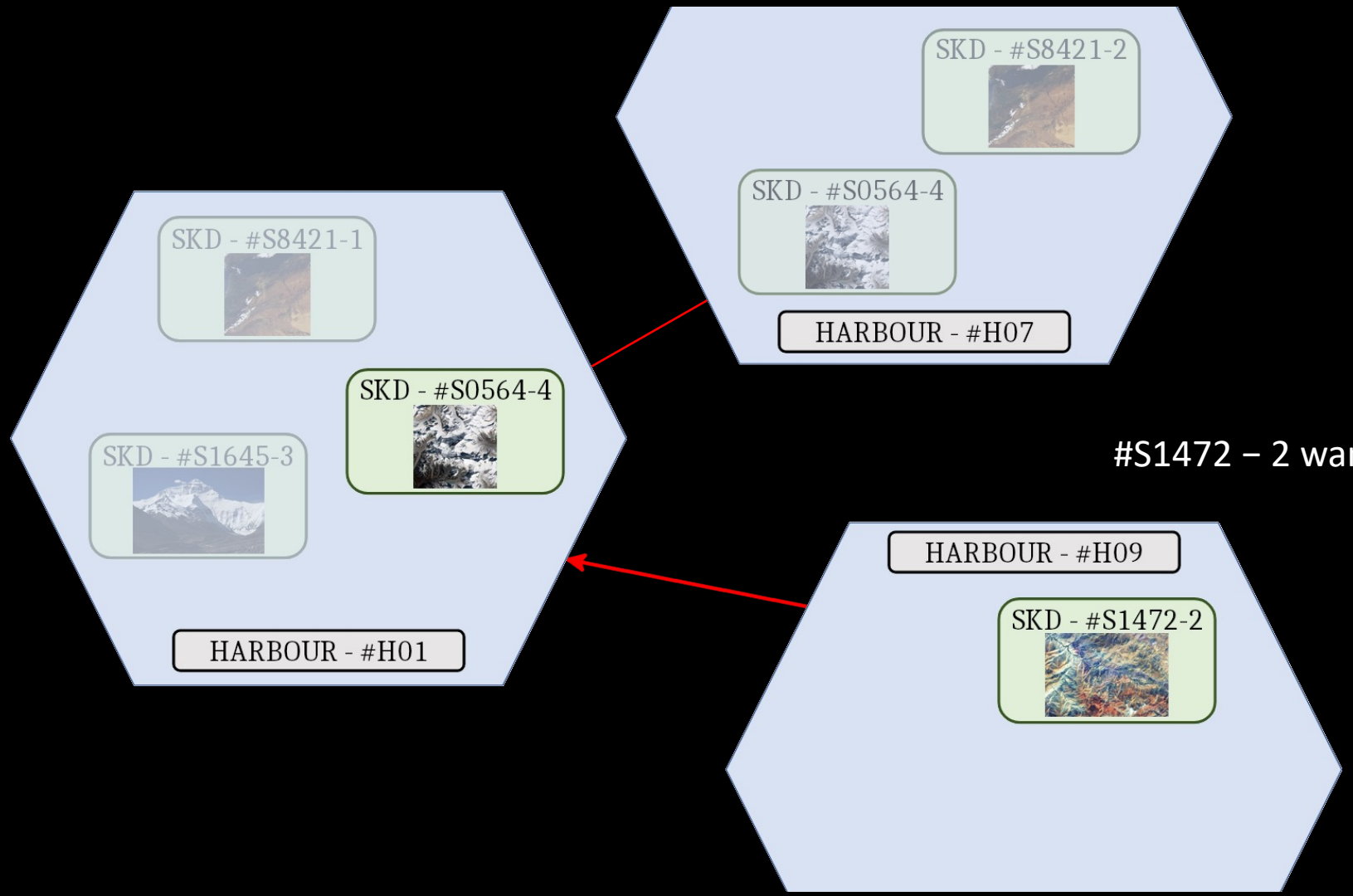
- Goals can be given to a SKD through metadata tags.
- Examples:
  - Green Awareness → « try to save energy. »
  - Stability → « keep yourself safe to preserve your data."
  - Exploration → « visit as many harbours as possible. »

How to achieve these goals?

- Machine learning and communication between SKD.
  - Multiagent reinforcement learning.
  - Federated learning.

In memory of Salima Hassas
Special thanks to valuable discussions on this topic

HOW TO KEEP IN TOUCH WITH A SKD



#S1472 – 2 wants to contact #S0564 – 4

HOW TO KEEP IN TOUCH WITH A SKD (2)

Solution

- Everyone must keep track of where everyone is, at all times.
- 2 problems:
  1. the number of messages scales very badly;
  2. the SKD become less autonomous.

- Most database applications cannot be used with SkyData.

AVAILABLE USER OPERATION (FOR NOW)

- List/fetch all the local SKD.
- Create a SKD.

- Broadcast to all SKD.
  - But only best effort, probabilistic broadcast.
- Update a SKD.
  - Probabilistic, through gossip.
- Delete a SKD.
  - Probabilistic, through gossip.

- Fetch a particular SKD: not in finite time.
  - Technically possible in a probabilistic way, but that's not useful.
- List all SKD: impossible.

MAINTAIN REPLICATION GOAL

Goal: Make sure that the familySize is equal to the replication goal.

- Communication with the family to determine if new replicas should be made (or deleted).
- Example: "We aimed to be 5 replicas, but there are only 4 family members."

```
if ReplicationGoal > FamilySize and with proba p1 then
    replication()
    FamilyBroadcast(new, IDReplica)
else if ReplicationGoal < FamilySize then
    deleteMyself ← True
    while deleteMyself do
        wait for N messages
        if at least RF ack then
            deletion()
        else if at least 1 denial then
            deleteMyself ← False
        end if
    end while
end if
```

A denial is sent if *ReplicationGoal* < *FamilySize* or with some probability p_2 to avoid deadlock.

PROTOTYPE IN PROGRESS

Prototype written in Jade & Jason

- Entities (such as SKD) are mainly defined in two files: skd.asl and SKD.java

1. skd.asl
   - Agent Speak Language
   - Describe the beliefs (local var), goals (what to achieve) and plans (how to achieve a goal) of the agent
2. SKD.java
   - Java class
   - describe the architecture of the agent
   - definition of primitives such as send, move, clone, delete,. . .

- Internals actions and behaviours are defined in auxiliary files.

CONCLUSION

- SkyData is:
  - A system for autonomous data management.
  - A complete departure from existing paradigms
  - Usable for useful applications avoiding the data use without your agreement.
  - A playground for so many distributed system issue
- SkyData is not:
  - A replacement or a concurrent for all replicated databases.
  - The begining of the end of the humanity