Industry relations in the ESFRI domain of Social & Cultural Innovation (aka SSH)

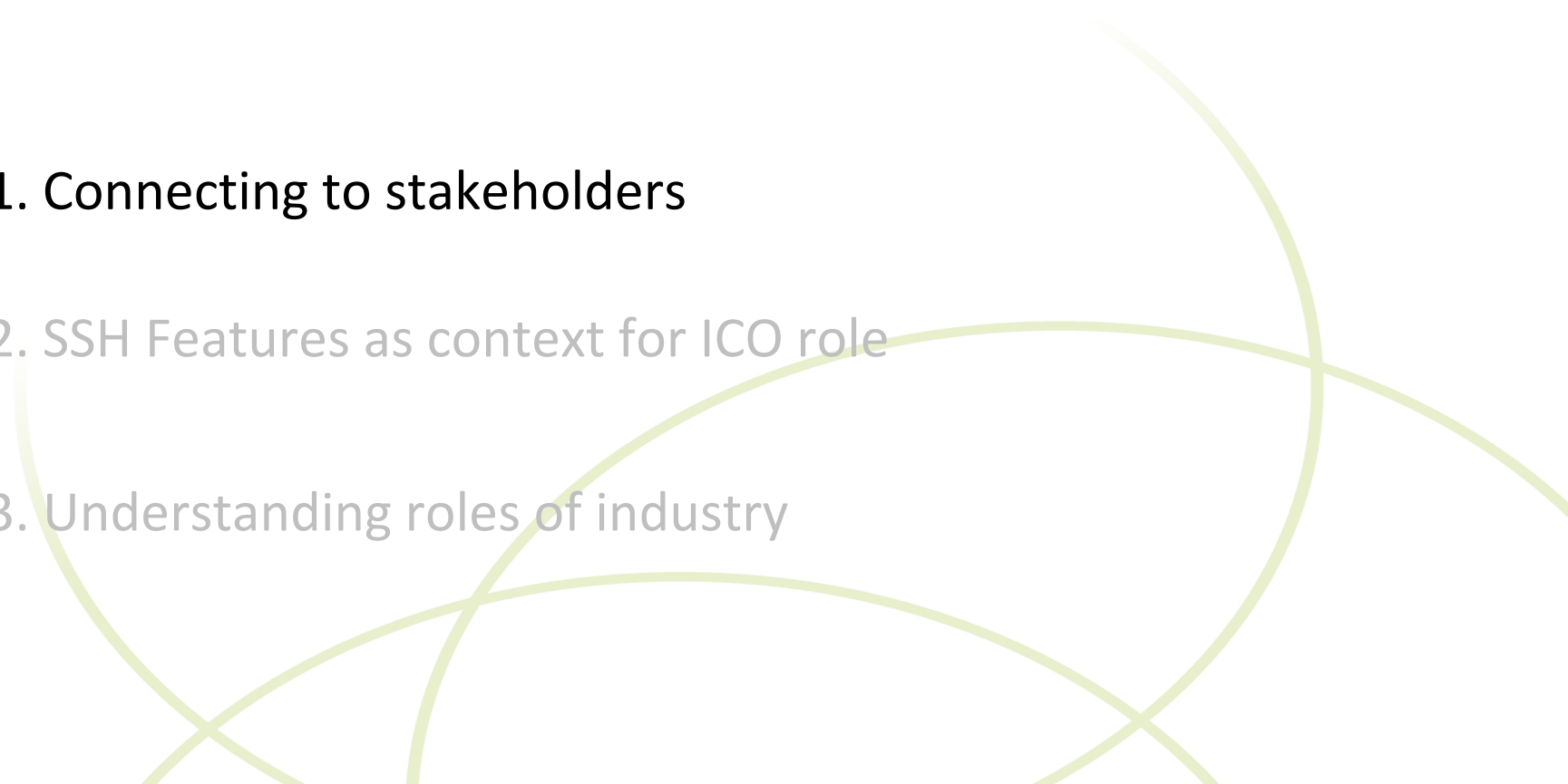## Franciska de Jong
executive director CLARIN ERIC

ENRIITC Webinar#1
27 May 2021

The various aspects of the ICO-role in SSH RIs

1. Connecting to stakeholders

2. SSH Features as context for ICO role

3. Understanding roles of industry

The various aspects of the ICO-role in SSH RIs

1. Connecting to stakeholders

2. SSH Features as context for ICO role

3. Understanding roles of industry

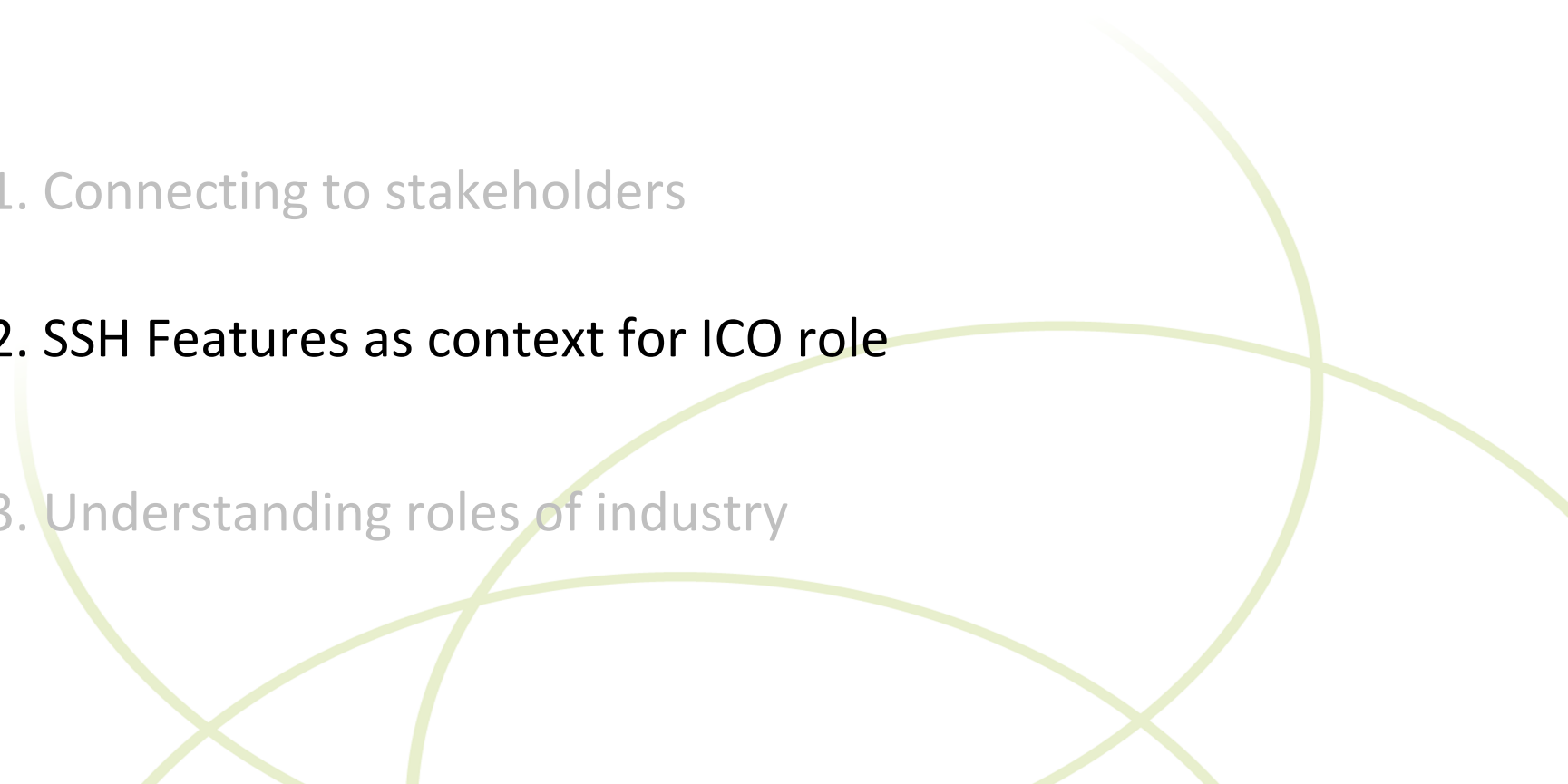# CLARIN data types and communities of use in SSH and beyond

- Newspaper archives
- Parliamentary records
- Literary texts
- Historical letters
- Broadcast archives
- Oral History data
- Social Media data
- L-2 Learner Resources
- Survey data
- Patient recordings
- Excavation reports

- Digital Humanities
- Linguistics and Philology
- Data Science /AI
- Translation and Lexicography
- Literary Studies
- History
- Political and Social Sciences
- Media Studies
- Culture, Folklore, Anthropology
- Speech therapy
- General Public

Connecting to Stakeholders

- Research communities:
  - tool developers and data providers
  - data-driven research on SSH topics

- Industry:
  - tech industry: search, translation, matching, text mining/AI
  - SMEs: dedicated search engines, translation services, workflows based on speech recognition, matching tools for recruitment, etc.

- Public sector: Galleries, Libraries, Archives, Museums (GLAM)

- Funders: countries, EC

The various aspects of the ICO-role in SSH RIs

1. Connecting to stakeholders

2. SSH Features as context for ICO role

3. Understanding roles of industry

CLARIN ERIC – a distributed research infrastructure for language resources

After 10 years: a consortium of

- 21 members

- 3 observers

- 1 linked party

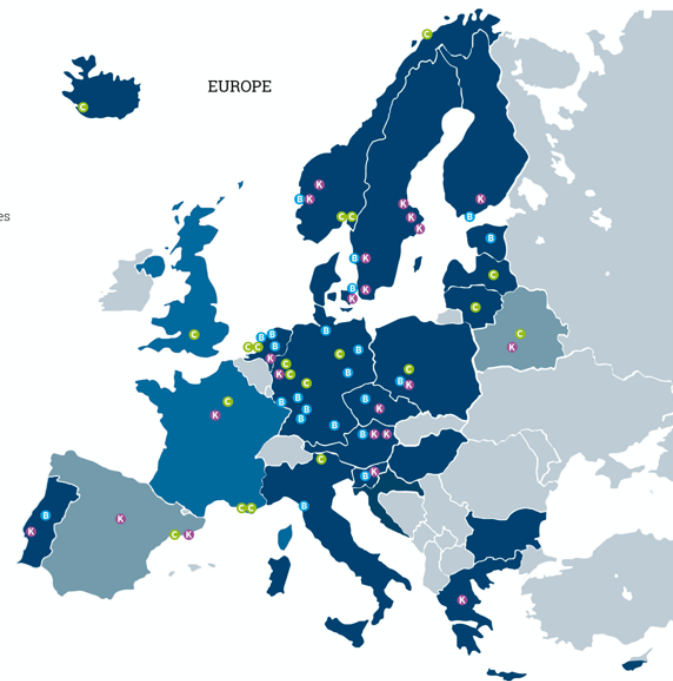## A distributed network of >60 centres

25 CTS certified data centres;

strong focus on FAIRness & interoperability

- *federated login:*

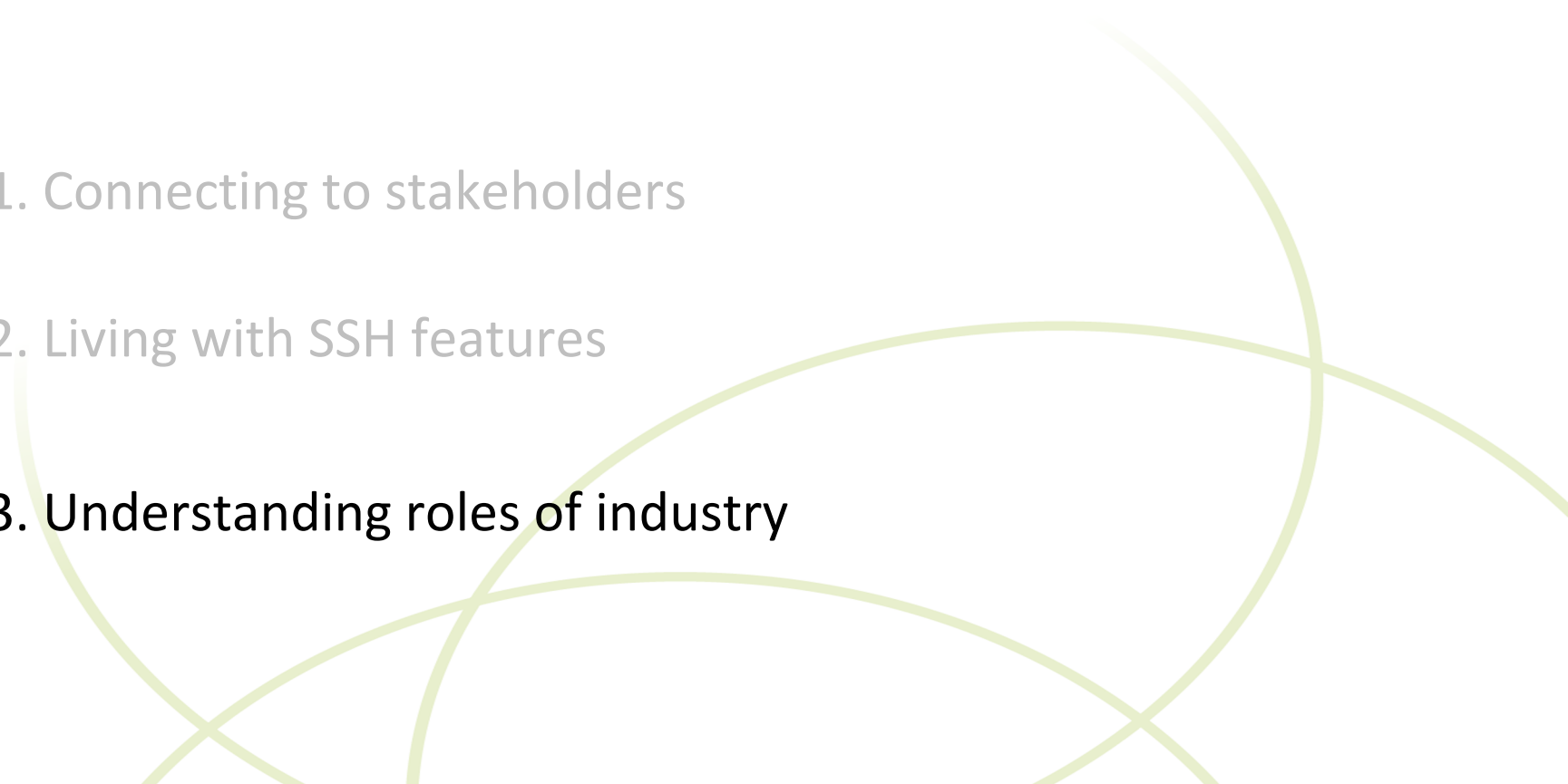- *central metadata harvesting for easy discovery:*

- *chained services:*

CLARIN

- ERIC members
- Observers
- Countries with participating centres
- **B** Centre Providing Data
- **C** Centre Providing Metadata
- **K** Knowledge Centre

EUROPE

USA

SOUTH AFRICA

SSH Features as context for ICO role

- Networks of distributed data repositories, certified

- RI as broker, ICO role distributed over the RI network, mostly in part-time and combined with other responsibilities

- Focus on Open Access to data and tools

- Services developed in open source

- SSH knowledge transfer partly through big but fragmented sectors such as communication and media

- No ILOs with focus on SSH

- Complementary network of academic group aimed collaboration with industry, with focus on Machine Translation

The various aspects of the ICO-role in SSH RIs

1. Connecting to stakeholders

2. Living with SSH features

3. Understanding roles of industry

Understanding the Roles of Industry

- Industry as user:
    - advice on how to use data to generate models for MT, ASR, AI
    - advice on how to handle heterogenous data
    - advice on data curation, terminology harmonisation
    - …

- Industry as supplier:
    - campus licenses for commercial software: e.g. LIWC, SketchEngine, spaCy, …
    - Provider of certification services, permanent data identifiers
    - …

- Industry as co-creator