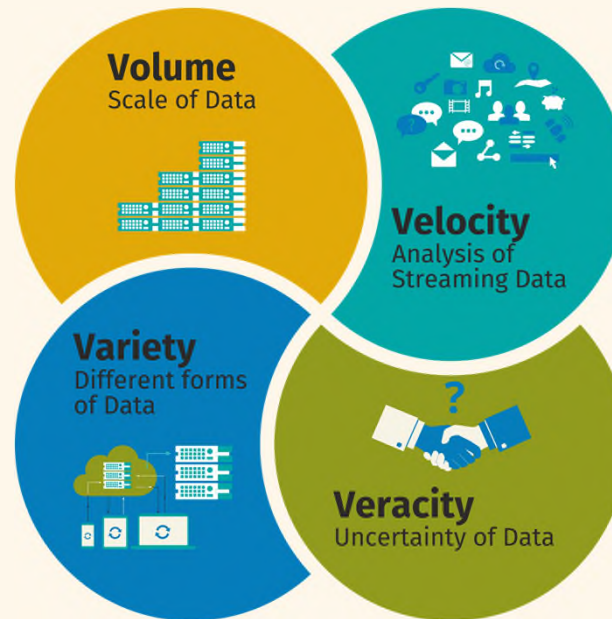**RESILIENCE**

Towards Big Religious Data –

RESILIENCE Research Infrastructure for Data on Religion in the Digital Age

Marco BÜCHLER, Sarah RIEGERT, **Federico ALPI**, Francesca CADDEDU

What is Big Data?

Why Big Data in the context of Religious Studies?

$$f = \frac{k}{r^\alpha} \tag{1}$$

$$R_1^n = \sum_{f=1}^{n} \frac{1}{f \cdot (f + 1)} \tag{2}$$

90-10 rule for textual data: 90% of the words in a corpus occur ten times or less

Table 1: Quantitative description of the long tail of rare concepts

Frequency f	Rank ratio	R_1^n
1	0.5000	0.5000
2	0.1667	0.6667
3	0.0833	0.7500
4	0.0500	0.8000
5	0.0333	0.8333
6	0.0238	0.8571
7	0.0179	0.8750
8	0.0139	0.8889
9	0.0111	0.9000
10	0.0091	0.9091
15	0.0042	0.9375
20	0.0024	0.9524
25	0.0015	0.9623
50	0.0004	0.9804
75	0.0002	0.9868
100	0.0001	0.9901

Types of data in Religious Studies

- Audio recordings (e.g. interviews, radio, podcasts)
- Bibliographies
- Pictures, Photographs
- Databases and/or tables
- Editions
- Questionnaires
- GIS data (measurement data)
- Manuscript books / other handwritten material (e.g., notes)
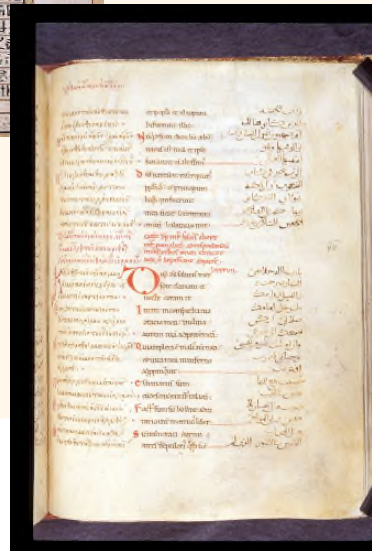
- Interview transcripts
- Maps
- Simulations, models
- Software (development), program code
- Text corpora and texts
- Video recordings
- Website, blogs, Twitter entries, etc.

Challenges

- **destruction of rare manuscripts or icons**: even if digitization made enormous progress, access to objects and knowledge (esp. about religion) will remain an issue for the decades ahead;

- **incomplete digitisation**;

- **property/market-driven restrictions**;

- **lack of specialized tools** to find, analyse, annotate and compare the digital data effectively;

- **fragmentation**: data are often developed and placed on different platforms and there is a lack of infrastructures allowing for one single access point to data on religion and to easy-to-use toolkits

Big Humanities Data – Scholarly DESIRE for Big Data

- **Diversity of languages**: humanities have to deal with texts in several languages, often rare, ancient or even undeciphered.

- **Evolution of historical records**. The semantic of concepts changes over time. To understand a concept's meaning, later research must understand a word in its epochal usage and meaning.

- **Stemma**. Historical data of more than 2000 years have nearly no chance to survive by its originals. Scribes copied text, made corrections, additions, and sometimes including errors due to e.g. reading mistakes in the text. The tracking of historical documents with its dependencies leads to stemmata.

- **Intertextuality** of data. Not only for texts but for all types of historical data, lines of transmission are important indicators not only for textual dependencies but also for ancient travel routes that delivered that information.

- **Reliability** of the source. Even if texts, paintings, stained glasses, or e.g. gable stones transmitted information from past to today, it is often unclear, how precise and reliable a certain information survived or only reflects a subjective perspective.

- **Elaborated interpretation**: As historical data can not necessarily be read with today's meanings, a critical interpretation and explanation of the data including further knowledge about a specific epoch or a dedicated geographical region is necessary to properly reflect data's content.

RESILIENCE

Consequence of Big Humanities Data: Co-existence of information overload vs. Information poverty

Big Data lead due to a huge amount of data to an information overload

Humanities Data its nature cause an information poverty due to lots of lost data

Big Religious Data

- Big Religious Data is the consequence of
  - Big Data paradigm
  - DESIRE principle
  - The co-existence of information overload and information poverty
  - Extra dimension of dedicated and specific normative texts such as the Bible or Quran

How to ensure Big Religious Data - RESILIENCE Research Infrastructure

- A unique, interdisciplinary and invigorating research infrastructure for all **Religious Studies**,

- building a high-performance **platform**,

- supplying evolving **tools** and **big data** to scholars from all the scientific disciplines crossing religions in their diachronical and synchronical variety.

RESILIENCE in numbers

- 34 years duration
- 13 partners from
- 8 European member states and...
- ... 3 associated countries
- 1 access point

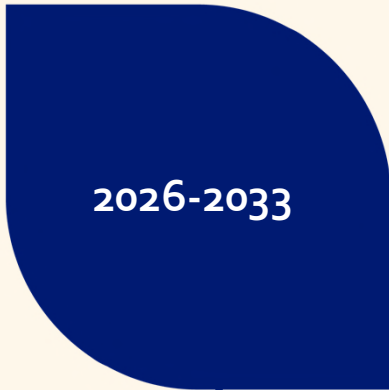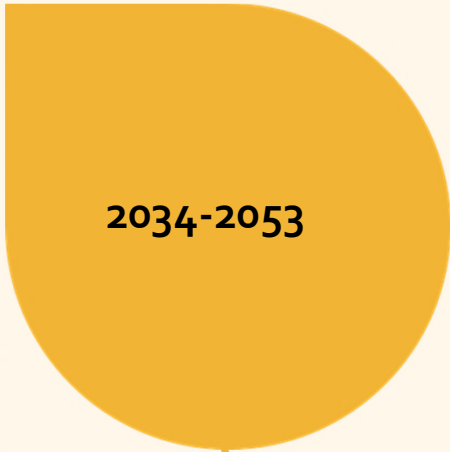**318,4 mio euro**

80% in-kind contributions

RESILIENCE

TERMINATION
Decomissioning

2054-2055

PREPARATION
Organisational entities set-up
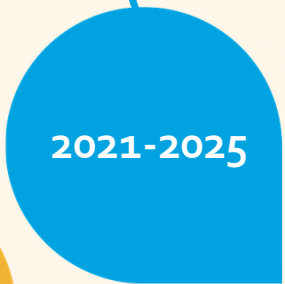
2021-2025
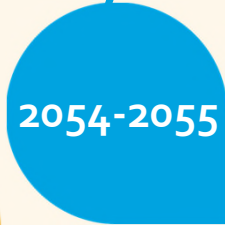
DESIGN
Support from the EC until September 2021

2014-2020

2026-2033

2034-2053

OPERATION
Starting 2034: 100% services available
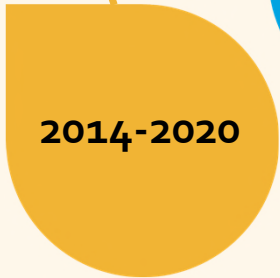
IMPLEMENTATION
2030: 33% of the services is ready.
The remaining 67% is fully operational before the end of 2033.

What's our timeline?

Our services

- Research enhancing services
- Research enabling services
- Research data management
- Data center…
- …and more

RESILIENCE

Let's stay in touch

- Join our focus groups
- Subscribe to our newsletter
- Follow us on social media
- Tell us your feedback

www.resilience-ri.eu

Serving Research

Building Knowledge

This project has received funding from The European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871127