

# TOWARDS CERTIFICATION OF GFBIO DATA CENTERS

Birgit Klasen<sup>1</sup>, Peter Grobe<sup>1</sup>, Christian Ebeling<sup>2</sup>, Jitendra Gaikwad<sup>3</sup>, Maren Gleisberg<sup>4</sup>, Falko Glöckler<sup>5</sup>, Anton Güntsch<sup>4</sup>, Joachim Holstein<sup>7</sup>, Birgitta König-Ries<sup>3</sup>, Anke Penzlin<sup>6</sup>, Mareike Petersen<sup>5</sup>, Stefan Seifert<sup>8</sup>, Dina Sharafeldien<sup>3</sup>, Sabine von Mering<sup>5</sup>, Tanja Weibulat<sup>8,9</sup>, Dagmar Triebel<sup>8</sup>



## CHOICE OF CERTIFICATION AUTHORITY

- Four potential data repository certifying options were evaluated: [nestor](#), [ISO16363](#), [DINI-Certificate 2016](#) and the [CoreTrustSeal \(CTS\)](#).
- Based on an analysis of certification requirements and community acceptance, **CoreTrustSeal certification was selected** as the most suitable candidate.
- To get CTS certification an adequate documentation at the data centers is sufficient and a physical visit from the CTS Authority is not necessary.
- Prior knowledge and experience of PANGAEA (GFBio partner) about CTS certification was also helpful in addressing the CTS Requirements.
- Due to the different infrastructures of the specific data centers, each goes through the certification process separately.

## DEMAND ANALYSIS AND INVENTORY

- The **17 Core Trustworthy Data Repositories Requirements** have to be fulfilled.
- All processes regarding data handling and related issues need to be described in detail.
- Publicly accessible documents** must be available as supporting evidence.
- Documentation and portrayal of the data centers was done in the years before as part of GFBio work and contributed to some of the Requirements.

## COLLABORATIVE WORKFLOW

- Although we apply for certification separately, we compiled the documents together.
- Each data center answered 2-3 of the 17 CTS Requirements and shared outcomes among data centers as templates.
- We **jointly discussed the Requirements** and work status and addressed concerns during regular virtual meetings.
- All data centers shared their CTS Requirement texts.
- Work progress (Fig.1) was **tracked in an overview table** within the GFBio Wiki (Fig.2).

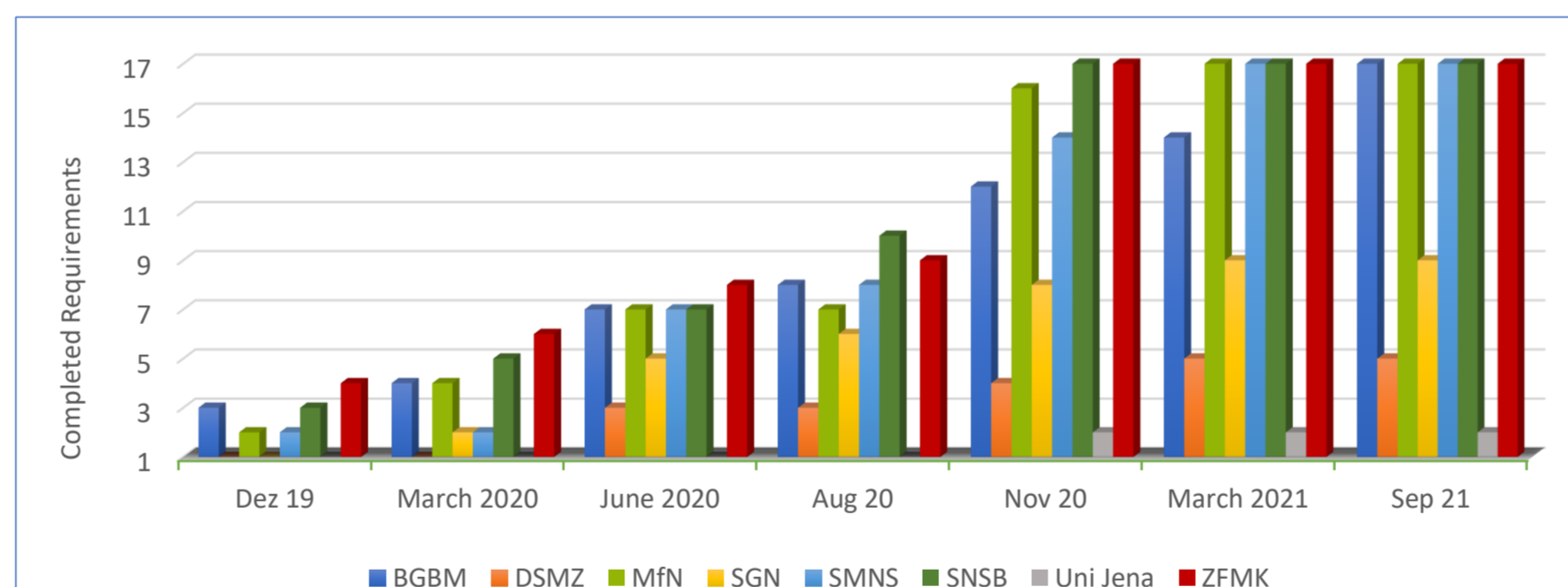


Fig. 1: Progress in fulfilling CTS Requirements

## MANDATORY CORE DOCUMENTS

- The certification process revealed that core documents such as **Data Policy**, **Privacy Policy**, **Terms of Use** and a **Digital Preservation Plan** are mandatory.
- These documents must be openly accessible on the website of the organization and require special care in their wording due to its potential legally binding content.

No	Title	BGBM	DSMZ	MfN	SGN	SMNS	SNSB	Uni Jena	ZFMK	Repository description: accepted data	Important:
R0	Context of data repository @ BGBM	draft	draft	draft	draft	draft	draft	draft	draft	Repository description: accepted data. If repository part of a network, add diagram. Designated Community: check with the requirement documentation on p. 8. the interaction with the community seems ok. Outsourced partners: List partners (computing centers, etc.) and provide certificates, SLAs, ... Level of curation: demonstrate that assurance of long-term accessibility of data as the needs of the Designated Community. "Cooperating OAI archives" versus "federated OAI archives" etc. (note D.T. 2020-11-22): OAI compliant archive reference needed under 6.1.	
R1	Mission and Scope of data repository @ SGN	draft	draft	draft	draft	draft	draft	draft	draft	include "data preservation" in the mission of the repository!	
R2	Licenses @ Jena, DSMZ	draft	draft	draft	draft	draft	draft	draft	draft	Licensing of own Software not necessary	
R3	Continuity of access @ ZFMK	draft	draft	draft	draft	draft	draft	draft	draft	include risk management	
R4	Confidentiality/Ethics @ SNSB	draft	draft	draft	draft	draft	draft	draft	draft	No separate list section. Include links in the text	
R5	Organizational infrastructure @ ZFMK	draft	draft	draft	draft	draft	draft	draft	draft	No noticeable problems	
R6	Expert guidance @ ZFMK	draft	draft	draft	draft	draft	draft	draft	draft	No noticeable problems	
R7	Data integrity and authenticity @ BGBM	draft	draft	draft	draft	draft	draft	draft	draft	Data policy? (date of the data, data ownership change, etc.): See: <a href="https://wiki.pangaea.de/wiki/Data_policy">https://wiki.pangaea.de/wiki/Data_policy</a> and Data policy	
R8	Appraisal @ Jena	draft	draft	draft	draft	draft	draft	draft	draft	Preferred formats? Is "Data Submission Formats" @ on Datacenter site sufficient? Or do we need something like <a href="https://www.terminology.com">https://www.terminology.com</a> ?	
R9	Documented storage procedures @ SMNS	draft	draft	draft	draft	draft	draft	draft	draft	What about redundant copies between data centers?	
R10	Preservation plan @ ZFMK	draft	draft	draft	draft	draft	draft	draft	draft	transfer of custody and responsibility handover? => Data policy?	
R11	Data quality @ MfN	draft	draft	draft	draft	draft	draft	draft	draft	ability of the Designated Community to comment on, and/or rate data and metadata? => Annotate? citations to related works or links to citation indices are provided	
R12	Workflows @ MfN	draft	draft	draft	draft	draft	draft	draft	draft	include details of how any issues are resolved	
R13	Data discovery and identification @ BGBM	draft	draft	draft	draft	draft	draft	draft	draft	Searchable metadata catalogue (we currently have provenance data, what about taxonomic coverage, geographic area / taxonomic terms)?	
R14	Data reuse @ SMNS	draft	draft	draft	draft	draft	draft	draft	draft	Listing in registry of resources?	
R15	Technical infrastructure @ SNSB	draft	draft	draft	draft	draft	draft	draft	draft	Evolving formats	
R16	Security @ SNSB	draft	draft	draft	draft	draft	draft	draft	draft		

Fig. 2: Overview table with Responsibilities for Requirements, Progress of work and Important issues (status: September 2021)

## SUBMISSION AND REVIEW

- Museum Koenig Bonn and Naturkundemuseum Stuttgart submitted their applications in April and July 2021 respectively.
- Both **applications are currently under review**.
- Most of the other data centers have finalized their texts of the CTS Requirements, but final approval by the governing bodies of the organizations is pending.
- Internal restructuring measures and the optimization of internal processes delayed more submissions.

**The entire certification process helped each data center to optimize processes and implement measures.**