Response to the CA/Browser Forum discussion on domainComponent usage

The validation sub-committee of the CA/Browser Forum, during the *Summer 2022 CA/B Forum F2F*, discussed a specific proposal to revise the Baseline Requirements (BR) that, when adopted, would result in serious adverse consequences for the research and education community using certificates for its research infrastructures. Specifically, the proposal

> "Except where explicitly specified, each `Name` MUST NOT contain more than one instance of a given `AttributeTypeAndValue` across all `RelativeDistinguishedName`s." (https://github.com/sleevi/cabforum-docs/pull/36/files#diff-e0ac1bd190515a4f2ec09139d395ef6a8c7e9e5b612957c1f5a2dea80c6a6cfeR3030)

is concerning, since it conflicts with the use of the domainComponent attribute type when used in an RFC-compliant way. It is this 'dc' attribute sequence on which the joint trust mode is based that allows both human researchers to use browsers to access academic research resources, and at the same time enables automated agents to act on those same services for managed data transfers and other research workflow automation tasks.

This was recognised in comments, where "GRID" was highlighted as an active use case that would be impacted, such as https://github.com/sleevi/cabforum-docs/pull/36#discussion_r859176437. The further discussion in that comment does not quite convey the importance of the use case – which is understandable since the research and education user community is not part of these discussions.

With this response, we want to highlight the specific issues and put these into the context of a secure certificate ecosystem, including academic research, which of course places high value on security, appropriate validation, and adherence to the applicable RFCs.

We only recently became aware of this discussion, so our apologies for possible chiming in later than usual. The Grid community is not directly represented in these discussions, usually relying on our partners in the Forum (over time we work and have worked with many, such as DigiCert and Sectigo, earlier also with GlobalSign and the former Comodo), so the "GRID" may appear remote and distant. I will try to close that gap a bit, since what is "GRID" here, is in fact a shorthand for the global research and education authentication community that uses PKIX for web-based research infrastructures. This includes e.g. CERN's Large Hadron Collider, many US and European supercomputing centres, life science and health research, astronomy, &c - and their globally distributed web (storage) services accessed both via browsers as well as via automated transfer services. In addition, next to its server PKI deployment that is relevant here it includes a global end-user client PKI than allows researchers to authenticate.

Over almost two decades, the research and education collaborations, such as TERENA (now GÉANT) for all European research and education networks, and InCommon in the US, have worked with their partners here to create a secure and tightly controlled ecosystem providing name uniqueness, adherence to standards (and providing some security and authentication innovations thereof along the way). We, as the global non-profit research community, are very supportive of a secure and trusted ecosystem for WebPKI, also by ensuring that browser-based access and service-agents can work together to secure the ecosystem in a comprehensive way.

The use of domainComponents in the subjectName is a core elements since it provides uniqueness of subject naming through (verified domain name owner-authorized) namespace slicing. This allows the browser users to validate the server, but at the same time is used e.g. for service-agent based authentication. Since these will by necessity use the very same end-points, it is important that the server certificates used have both WebPKI browser trust for human users as well as the uniqueness properties necessary for agent-based access.

The specific global uniqueness of the subject DN and having these certificates trusted for WebPKI is critical for securing connection to services that are access by both browsers as well as automated agents. Such as storage services: end-users can retrieves files via their browser, but automated large-scale research transfer services connect to the endpoints to do data placement (e.g. for CERN's worldwide Large Hadron Collider Computing Grid) and must authenticate the endpoints based on their certificate subject name. To identify the endpoints, unique naming is necessary, which is what – for TCS and InCommon - the subject DN domainComponents supplies.

But, by design and semantics, domainComponent should appear multiple times. RFC 2247 is quite clear: "The value of this attribute is a string holding *one component* of a domain name." So, given that a domain name usually consists of multiple components, also domainComponent should have multiple occurrences. With the domain name system being hierarchical, the ASN.1 thus _has_ to be a sequence (of sets of unitary length) of this attribute.

The inclusion of domainComponent, in the way that it is worded today, was definitely not accidental or a refactoring anomaly. Our community worked together with several members, specifically DigiCert and Comodo (around ~2011), to align the BR text with these joint WebPKI security requirements. And ensure that domainComponent, and its standardized semantics, were included consistently in BR. The result was explicitly reconfirmed in *Ballot 102* (https://cabforum.org/2013/05/31/ballot-102-br-9-2-3-domaincomponent/).

We therefore hereby reconfirm the existing current and continuing case for domainComponent, in line with the purpose as included in BR and in Ballot 102.

The use of domainComponents, as may be observed from Censys, identifies the long-term domain name from the domain name owner, e.g. the Trans-European Research and Education Networking Association TERENA – terena.org, now GEANT; or the InCommon federation – incommon.org, providing certificates for the US research and education community (even if our issuing partners change). Intentionally, for security reasons certificate validity periods have always been rather short (the community had settled on max 400 days since its conception in 2001), but the subject names themselves are much longer lived: many of the research use cases is easily 30+ years. DomainComponents are the most appropriate attribute to carry the uniqueness information.

We note by the way that the rendering by Censys.io is _not_ what is actually in the certificate subject names, since there is a bug in Censys that duplicates the sequence of domainComponents in the string rendering. It is in fact of course "[…non-DC-attributes…], dc=tcs, dc=terena, dc=org" (in RFC2253 ordering). Crt.sh shows it correctly (e.g. https://crt.sh/?id=5460913588).

Unique certificate subject naming, when used in combination with relying-party defined namespace constraints, also permits mutual authentication between endpoints. Admittedly, this is beyond the browser use case today, but the concept provides some supplementary protections by moving the source of authority for NameConstraints to the relying party, as discussed in GFD-I.189 (https://www.ogf.org/documents/GFD.189.pdf). Even in browsing scenarios this adds some security (admittedly only slightly) for text-browser-based access by end-users to the web services, since there global name uniqueness can be verified in appropriate clients.

Without the availability of (RFC standards-compliant) domainComponents, subject distinguished names will no longer be distinguishable …