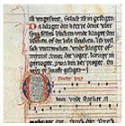CATCH



CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS



Kickoff Workspaces

Hennie Brugman

Technisch coordinator CATCHPlus

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Agenda

10.00 – 12.00

1. *Inleiding (Hennie)*
2. *Voorstelronde*
3. *Scope en doelen (Hennie)*
4. *Project-aanpak (Rolf)*
5. *Globaal ontwerp (Rolf en Minne)*

Lunch

13.00 – uiterlijk 15.00

1. *Andere cases*
2. *Afspraken*
3. *Afsluiting*

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Inleiding

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Aanleiding

- Gedeeld usage pattern
  - Ingest van resources, per gebruiker of gebruikersgroep
  - Authenticatie
  - Definiëren van data sets
  - Services (tijdelijk) toegang verlenen tot die data sets
  - ‘Bundelen’ van resources
  - In veel gevallen is ook sprake van annotatie

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



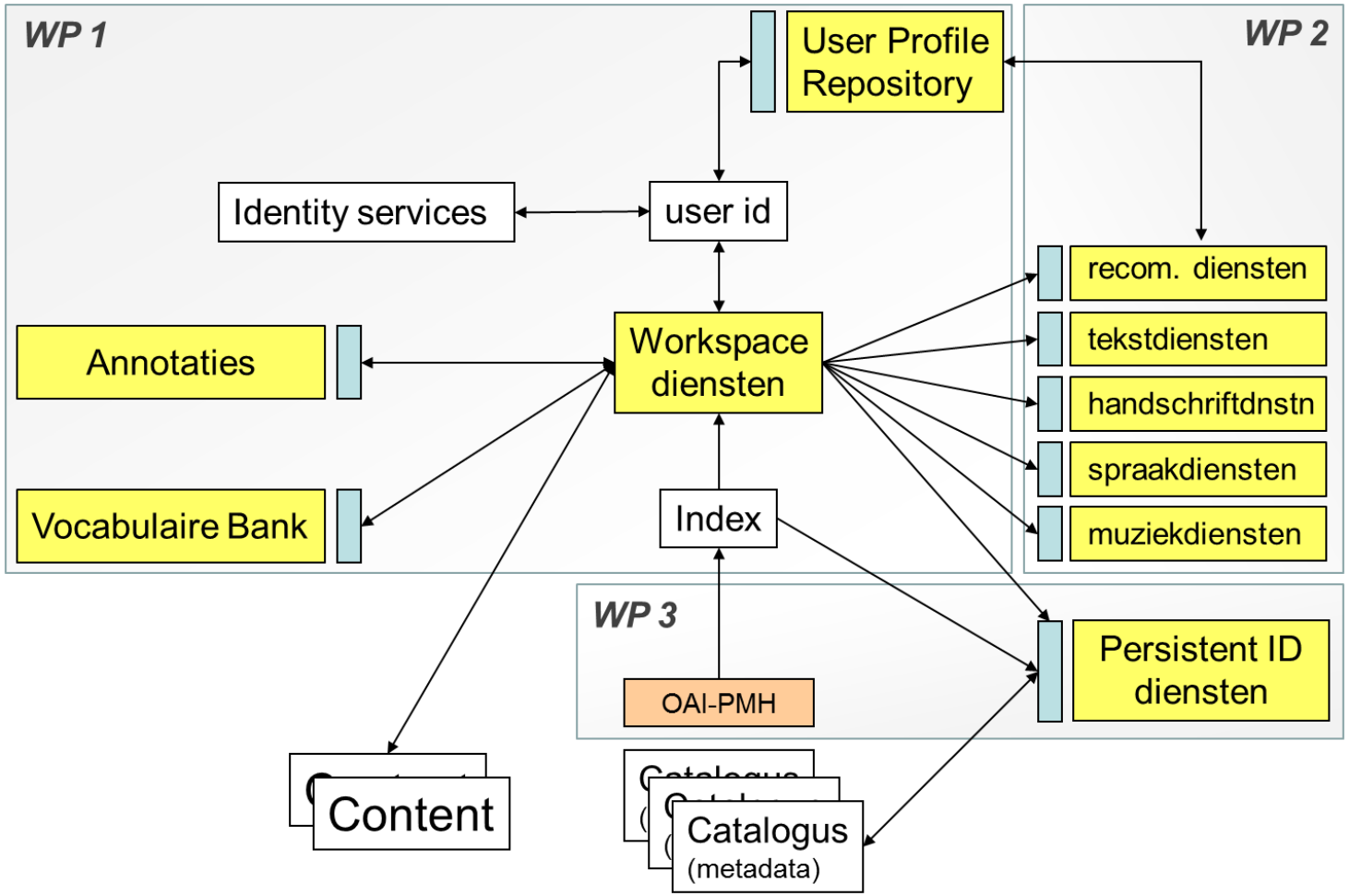
Wat willen we doen?

- "Workspaces" realiseren adhv een eerste case (Scratch4All)
  - Persoonlijke, tijdelijke opslag van data sets. Tijdelijk verlenen van toegangsrechten tot die data sets
  - Web services op basis van REST
- Actief betrekken mogelijke andere cases bij het ontwikkelproces
- Daardoor (hopelijk) doorgroeien naar breed inzetbare dienst

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Het “landschap”



CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Doelen voor deze kickoff

- Project toelichten (inhoud en proces)
- Status update geven
- Toetsen van onze eerste aanzet
- Gelegenheid tot inbrengen van eigen wensen en inzichten
- Desgewenst concrete vervolgafspraken en -akties

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Voorstelronde

CATCH



CONTINUOUS
ACCESS
TO
CULTURAL
HERITAGE
PLUS



Scope en doelen

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

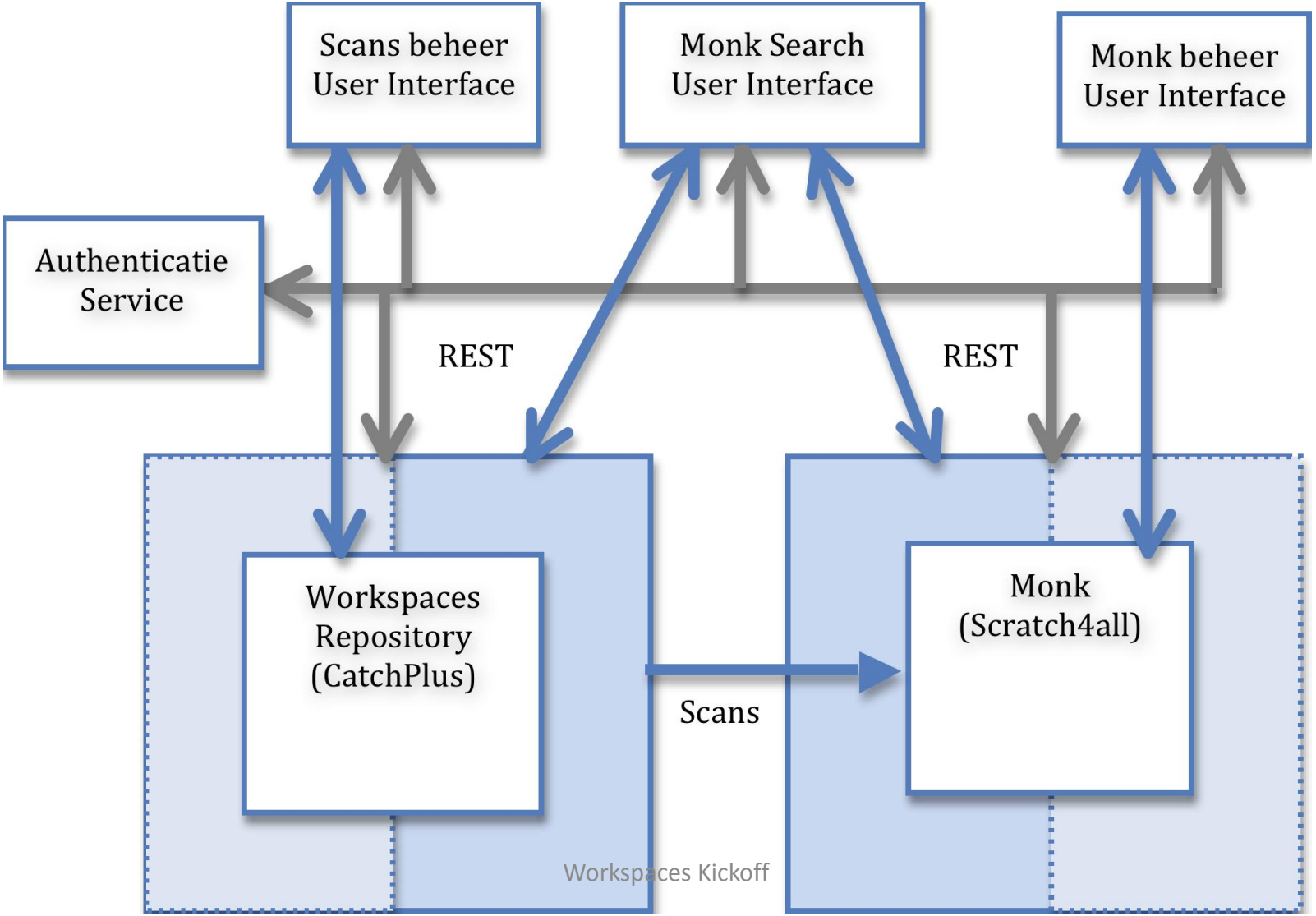
Workspace toepassingen

- Primair toegepast tbv Scratch4All (images)
- In tweede instantie toegepast op één of meer van de andere CATCHPlus cases (spraak, sets van zoekresultaten)

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Implementatie ikhv Scratch4All



CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Mogelijke andere cases

- Andere Scratch4All collecties
- Spraakherkenning/CHORALPlus
- CODA (CATCHPlus Open Document Annotation)
- (CLARIN web service workflows)
- (Multiply/DSS)
- (MuseumPlus)

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Wat bouwen we in het lopende project?

- Workspace repository (Target Holding)
- RESTful API (Target Holding)
- Webapps voor workspace beheer en workspace repository beheer (Target Holding)
- Basale authenticatie-oplossing
- Tijdelijk delegeren van toegangsrechten (op basis OAuth?)

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Wat bouwen we (nog) niet?

- Volwaardige authenticatie en authorisatie oplossing
- 2-4 toepassingen binnen CATCHPlus
- Toepassing tbv web service workflows
- Annotatie diensten
- Ondersteuning voor andere objecten dan gescande documenten en toebehoor

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Status

- Implementatie Workspaces en aansluiting op Scratch4All is gestart
  - Samenwerking CATCHPlus en Target Holding
  - Met instemming Nationaal Archief
  - Afronding: december 2011
- Gesprekken over Authenticatie en Authorisatie module zijn nog gaande

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Mogelijke voordelen

- Voor aangesloten erfgoed-instellingen of –gebruikers
  - Flexibele, tijdelijke hosting on-demand
  - Inbrengen data tbv externe samenwerking
  - Veilige toegang tot data (passwords uitdelen of intrekken(!) niet nodig)
  - Online plek voor collecties ‘in ontwikkeling’ (voorafgaand aan archivering of publicatie)
- Voor diensten-aanbieders
  - Vaste toegangsmethode tot data sets
  - Hoeven niet te beschikken over actuele user credentials (“passwords”)

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Mogelijke voordelen

- Voor aanbieders van workspace diensten
  - Cloud hosting service op maat, toegespitst op doelgroep en bruikbaar in service oriented architecture
- Voor ontwikkelaars
  - Hergebruik van modules (inloggen, authorisatie, ingest) en operationele services
  - Heldere interfaces
- Voor erfgoedinfrastructuur
  - Standaardisatie
  - Samenhangend framework waarop kan worden ‘ingeplugd’
  - Integrerend, basis voor VRE’s, lanceerplatform voor diensten

Potentieel

- Goede ‘home base’ voor geïntegreerde applicatie-omgevingen (b.v. Virtual Research Environments)
- Cloud hosting aspect: tijdelijke, persoonlijke, veilige opslag op het web, met flexibele toegankelijkheid
- Voorportaal voor digitale archiving
  - data curatie
  - toekenning persistente identifiers

Suggesties mbt mogelijke rollen

- Target – ontwikkelaar, diensten-aanbieder, aanbieder workspace services
- NA – gebruiker, aanbieder data collecties
- RuG– dienstenaanbieder, belang bij data
- CLARIN – gebruiker, diensten-aanbieder
- Spraakherkenning – gebruiker, diensten-aanbieder
- Multiply – gebruiker, diensten-aanbieder
- SARA (BigGrid) – aanbieder workspace services, ontwikkelaar, rol mbt AAI?

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Andere cases

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



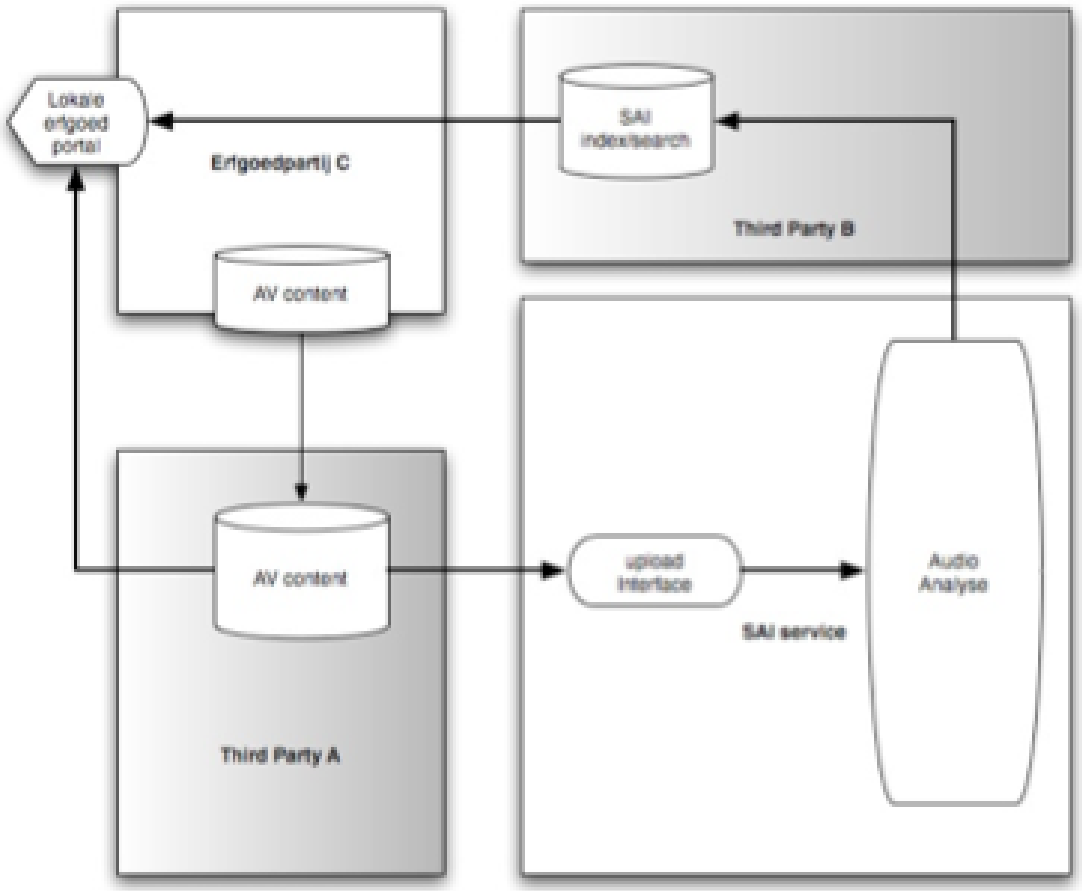
Cases

1. Spraakherkenning (GAR)
2. DSS/Multiply (BenG)
3. MuseumPlus (gM)
4. CLARIN workflows
5. CODA

CATCH

CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS

Spraakherkenning



Spraakherkenning

- Type data
  - Audio files, eventueel gekoppeld aan tekstbestanden
- Welke web services?
  - SAI audio services
  - Search engine op basis van audio index
- Welke web applicaties?
  - Spraak-specifieke upload interface
  - zoekinterface

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



DSS/Multiply

- *Geïntegreerde omgeving, waarin aan documentalisten gerangschikte kandidaat-trefwoorden worden aangeboden, die automatisch worden gegenereerd op basis van tekstdocumenten (mbt radio- en televisie-programma's)*
- Type data
  - Referenties naar metadata
  - Referenties naar AV documenten
  - Tekstdocumenten
- Welke web services?
  - Semantische annotatie-service voor tekst
  - Ranking service (op basis annotaties)
- Welke web applicaties?
  - Geïntegreerde omgeving

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



MuseumPlus

- *Op basis van slim zoeken in metadata over meerdere collecties (wie-wat-waar-wanneer) worden resultaatsets als virtuele collecties beschikbaar gesteld. Erfgoedprofessionals en het publiek kunnen deze collecties en hun objecten in onderlinge interactie verrijken.*
- Type data
  - (links naar) metadata beschrijvingen
  - Links naar collectie objecten
  - zoekresultatenlijsten
- Welke web services?
  - MuseumPlus search engine
- Welke web applicaties?
  - Annotatie/collaboratie omgeving

CLARIN workflows

- *Teksten kunnen worden onderworpen aan een keten van opeenvolgende web services, die allemaal iets aan een gewenste linguistische analyse toevoegen. Die opeenvolgingen en hoe ze samenwerken wordt gespecificeerd in z.g. workflows. Tussenresultaten en z.g. 'provenance' informatie worden opgeslagen in en uitgewisseld via workspaces.*
- Type data
  - Annotaties, gekoppeld aan tekst-segmenten
- Welke web services?
  - Diverse linguistische services

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



CODA

- CATCHPlus Open Document Annotation
  - Named entities in transcripties van documenten
  - Automatische line strips tbv handmatige annotatie
- Type data
  - (links naar) scans
  - Annotaties bij ‘line strips’ of ‘wordzones’
- Welke web services?
  - Line strip detectie service
  - Named Entity Recognition service
- Welke web applicaties?
  - Document annotatie omgeving
  - Document upload tools
  - Annotatie workflow management tools

CATCH



CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



Mbt de resultaten

- Broncode komt onder Open source licentie en online beschikbaar (b.v. via GitHub, geldt niet voor Monk service)
- Vrij gebruik voor Monk gebruikers (?)
- Kan vrij worden gedeployed door wie dat wil
- Kan als basis dienen voor doorontwikkeling of verbetering ihkv andere projecten
- Zoals voor alle CATCHPlus diensten, moet er een businessmodel voor duurzame exploitatie worden ontwikkeld.